PhD Thesis in Sciences

Faculty of Sciences
University of Liege, Belgium

Progress in hybrid diffractive/refractive lens solutions for compact space IR imager

---

- Composition of the Jury -

Prof Rudolf Saathof - Delft University of Technology
Prof Sylvain Lecler - INSA Strasbourg
Prof Gaëtan Kerschen - University of Liege
Prof Bertrand Bonfond - University of Liege

- Promotor-

Prof Serge Habraken - University of Liege

- Co-promotor -

Prof Jérôme Loicq - Delft University of Technology
- University of Liege

A thesis presented by
Victor Laborde
for the degree of Doctor in Sciences

October 20 2022

Remerciements

Je tiens à remercier le Professeur Serge Habraken qui a supervisé ma thèse durant ces cinq années au sein du centre spatial de Liège. Il a été disponible et à l'écoute tout en me laissant un grand espace de liberté et de créativité. Serge, je te suis reconnaissant de la confiance que tu m'as témoignée en tant que doctorant et assistant. J'exprime également ma gratitude à mon co-promoteur, le Professeur Jérôme Loicq, pour son soutien indéfectible et son expertise qui a considérablement enrichi ma thèse. Quel soulagement d'avoir pu, lors de nos réunions du vendredi matin, t'exposer mes interrogations et recevoir des pistes de réflexion. Nous avons même discuté optique sur le GR de la côte d'Opale! Merci Jérôme d'avoir posé un regard critique sur mon travail et de m'avoir permis de l'approfondir.

Je dis merci à mes camarades doctorants d'avoir embelli cette expérience. Christian tu m'as prodigué des conseils avisés et ton expertise en optique a été, Ô combien! précieuse. Bises à Lola! Merci à Théo, camarade de galère, d'avoir montré la voie. Finalement, mention spéciale à mes trois acolytes Colin, Jesus et Bogdan pour nos discussions et cette ambiance de bureau amicale!

Je remercie chaleureusement mes collègues du CSL, Camille Galy et Alexandra Mazzoli, grâce à qui beaucoup d'obstacles techniques ont pu être levés. Je suis reconnaissant envers Fabian, Yu Chen, Marc, Yvan, Pascal, Aline, Agnula pour nos échanges et leur accueil au sein du CSL. Enfin, je tiens à mettre en valeur Juriy Hastanin et Lionel Clermont pour m'avoir accompagné durant toutes ces années. Juriy, ce fut un plaisir de publier des articles avec toi, un plaisir de suivre le fil de tes idées innovantes et un plasir de partager ta bonne humeur! Un merci sincère à Lionel, camarade de grimpe et fournisseur officiel de "Science" pour m'avoir épaulé et aiguillé tout au long de ma thèse. Bises à tes trois petits monstres et à Nancy!

Mes amis John, Macélia et Tatiana, vous avez égayé mon quotidien au kot. Nos soirées jeux sont inoubliables (Tatiana, je sais que tu triches)! Merci à Adrienne et Emmanuel pour ces restos sushi-conseils. Bisous au petit ange, scientifique en devenir!

Merci de tout coeur à mes meilleurs amis de l'ENSICA: Romain, Louise, Hugo et Morgan. Quel bonheur de se retrouver chaque année au ski, dans le sud ou à Paris. Louise et Romain, merci de m'avoir hébergé durant mon télétravail! Lieutenant Morgan Morel, merci de m'avoir transmis rigueur et discipline! Je remercie tendrement Mélissa, Rémi, Bastien et Bruno, mes chers amis de Toulouse.

Viennent ensuite Samy, Lucas et Corentin, mes amis d'enfance. Corentin, tu es un ami en or, notre amitié n'est pas quantifiable!

Maman, si tu ne m'avais pas expliqué la division au collège, je n'en serais pas là! Papa, par ton amour de la science fiction, tu m'as transmis la curiosité et l'amour des sciences. Je vous dois aussi l'envie d'un travail bien fait. Vous m'avez accompagné durant cinq ans sans douter. Revenir à la maison est toujours source d'apaisement. Un grand merci à toi Juju, ma soeurette, pour ton écoute et ta présence rassurante. Clin d'oeil à Agnès ma marraine sur qui je peux toujours compter en cas de pépins!

André et Hélène, vous m'avez accueilli, coaché et porté durant cette période de ma vie et vous m'accompagnez en douceur vers la suivante. Votre générosité et votre envie de partager sont une grande source de bonheur et d'apprentissage pour moi. Léa et Natan, aventuriers du rail, rendez-vous chez Harold, Meert ou en régoin champignoneuse!

Et puis, il y a ma douce Mathilde. Quelle joie de partager ma vie avec toi. Sans ta bienveillance, cette thèse n'aurait jamais vu le jour. Je t'aime Mathilde.

Résumé

Dans cette thèse de doctorat, nous analysons le potentiel d'un élément optique innovant, nommé *élément optique diffractif multicouches* (multilayer diffractive optical element, MLDOE). Cette recherche est menée dans le cadre de l'observation de la Terre dans l'infrarouge thermique. Certaines applications infrarouges majeures ainsi que certaines technologies de détection infrarouge forment le contexte de cette étude et amènent les principales contraintes optiques. Un système optique est également soumis à certaines aberrations intrinsèques: thermique et chromatique. Les différents leviers permettant la conception optique infrarouge de haute résolution sont analysés par le budget radiométrique, qui estime le signal lumineux reçu. L'élément diffractif multicouches innovant possède idéalement des performances optiques proches de la limite de diffraction, et ce, sur une large gamme de longueurs d'onde. Ce manuscrit explore différentes modélisations de cet élément, du tracé de rayons en passant par la discrétisation des équations de Maxwell. Les performances optiques du MLDOE sont majoritairement le fruit des matériaux qui le composent ainsi que de la forme que prennent ces couches diffractives. Nous montrons que ces performances sont en réalité loin d'être idéales pour un grand nombre de MLDOEs. Suite à de multiples optimisations, un arrangement de couches en particulier atteint de hautes performances.

Mots clés: Infrarouge thermique, optique de Fourier, optique diffractive, MLDOE

Abstract

This research contribution analyses the potential of an innovative optical component called *multilayer diffractive optical element* (MLDOE) for infrared dual-band systems. The present manuscript intends to develop a comprehensive background centred around Earth infrared remote sensing. Infrared applications and detectors constrain the optical designer, who must also account for the intrinsic thermal and chromatic optical aberrations. The radiometric budget allows for understanding the available levers to build a high-resolution infrared dual-band imager and to estimate the resulting SNR. The innovative MLDOE ideally has a near-diffraction performance in the whole thermal infrared waveband, with a suitable chromatic aberration. This manuscript deeply explores MLDOE modelling using ray-tracing and an exact electromagnetic approach. An MLDOE optical performance is impacted by the material choice and the layer design, and is largely overestimated in the literature. One MLDOE design arose, resulting from various optimisation studies.

Keywords: Dual-band infrared, Fourier optics, diffractive optics, MLDOE

Contents

Introduction

This thesis work occurs within the context of infrared Earth remote sensing from space. Following the trail led by the Arkyd-6 spacecraft, we aim to develop a compact multi-band infrared imager that can embark on a CubeSat. Arkyd-6 is a 6U CubeSat technology demonstration mission of Planetary Resources Inc. with the goal to explore potentially useful resources on asteroids and develop technologies for asteroid mining. Arkyd-6 embarks a 26 m resolution mid-wave infrared camera that has proven valuable to detect refinery flame towers hot spots in Algeria for demonstration purposes. To achieve the feasibility study of a compact multi-band infrared camera and its design, we have divided this manuscript into three parts, namely:

- Part I: Radiometric study
- Part II: Modelling tools
- Part III: Designing innovative dual-band camera

In Part I we propose to analyse the broad thesis context. Among the many infrared applications for Earth remote sensing, we detail the monitoring of crop water stress, wildfires and volcanoes. These crucial applications come with their physics and requirements: What is the detectable temperature variation? How many observations must be done per day, per week? What is the best wavelength (or set of wavelengths) to map/detect the studied phenomena? What is the most optimal Earth feature size to get a valuable image? As a result, a 10 m ground sampling, imaged daily in dual-band MWIR/LWIR infrared at a 1 °K temperature resolution, enables all these applications simultaneously. We propose to use a constellation of CubeSats to ensure the essential daily coverage over some of the largest crop fields. Any hot event occurring in the spacecraft's field of view, such as a wildfire or an eruption, can be accurately imaged since a single optical system fulfils all the abovementioned requirements.

These requirements raise the main research challenge to overcome in this thesis: How to build a dual-band infrared system compact enough to fit inside a CubeSat? To answer this issue, we propose to study only cameras since they are generally more compact than telescopes. Diffractive optics, whose innovations are at the heart of this thesis, are closely tied to refractive systems as well. In General, the words "compact", "high resolution", and "infrared" are antonymic, especially when multiple bandwidths are considered. Therefore, we identify the various challenges of this thesis: Is it possible to build a dual-band camera with one compact optical system and one detector? Is it possible to achieve sufficient resolution? How to manage the inevitable infrared thermal and chromatic defocuses?

After developing a comprehensive radiometric study, we demonstrate the theoretical feasibility of such a high-resolution camera, resulting in tied optical constraints. We are able to solve the infrared intrinsic aberrations issue using dedicated optical design software and an optimisation algorithm. This algorithm, resulting from a bibliographical synthesis, is implemented and applied to reduce all inherent aberrations. To do so, the camera extensively uses diffractive lenses, whose optical properties and compactness are the bread-and-butter of this design. At this stage, however, the high-resolution camera can only operate in MWIR due to the lack of dual-band diffractive lenses.

Consequently, the main challenge tackled in this thesis becomes: Is it possible to extend the behaviour of diffractive lenses in dual-wavebands?
The element answering this question is called a multilayer diffractive optical element (MLDOE) and constitutes the main innovative topic of this thesis. Unfortunately, the analytical MLDOE modelling method provided in the literature is inaccurate due to the relatively high MLDOE thickness. Two major questions arise: How can we model the MLDOE internal behaviour? How can we retrieve its fundamental optical figures of merit?

To prepare for the study of this crucial question, we introduce multiple modelling tools in Part II. Since our system relies on a complex, innovative micro-optical element, we must combine various models to assess its performance accurately.

Responding to the second question, we develop a free-space wave propagation tool using the angular spectrum method (Fourier optics). This computer-friendly tool is very fast and accurate, calculating a wavefront phase and amplitude at any position along the propagation direction. Consequently, it constitutes our go-to method for retrieving the MLDOE focal length and its various figures of merit, such as the Strehl ratio. Since this method only works in free space, it cannot describe how an MLDOE deforms a plane wavefront.

Therefore, we propose using a combination of ray-tracing and Fourier optics, called the ray model, to approximate the phase delay introduced by an MLDOE. This model is a major innovation of this thesis and answers the question: How to estimate the performance/behaviour of an MLDOE using a fast and parameterisable approach? This model is crucial since it allows for swift, yet reliable, MLDOE shape and material optimisation. For speed and parametrisation reasons, we have developed the ray-tracing calculator of this model on Matlab. It is specific to MLDOEs, analytical, extremely rapid and as accurate as professional three-dimensional ray-tracing tools (ASAP).

However, when high accuracy comes in to play, the Finite Differences Time Domain (FDTD) method remains unmatched. This numerical method offers precise near-field MLDOE modelling at the cost of a high computational effort. This method's numerical results are considered a reference in this thesis since our fine sampling ensures the validity of its predictions. To our knowledge, only direct experimentation on an MLDOE prototype can provide more trustable results.

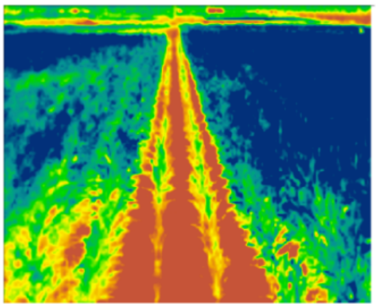
These three approaches combined in Part III provide a valuable and innovative MLDOE description. As already stated, using the ray model (or FDTD) with Fourier optics enables a powerful and fast (or accurate) MLDOE modelling tool. We publish the result of this approach with many MLDOE designs, F-numbers, and wavelength, to fully assess their potential. In doing so, we demonstrate that the analytical designing equations cannot produce performing MLDOEs, except on rare occasions. Having answered the modelling question, we raise a final challenge: Is it possible to optimise an MLDOE to provide an efficient dual-band component that could fit inside the designed camera?

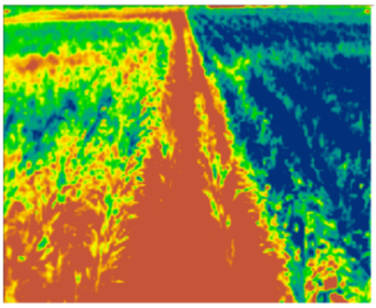
First, we propose to optimise the choice of the material composing the MLDOE. We look at three intuitive optical metrics: Total internal reflection, transmission, and total thickness. They provide a simple, yet valuable, material combinations filtering tool. We demonstrate that MLDOEs have higher performance when their gap is filled with a third material instead of air, potentially increasing the difficulty of manufacturing them. We define the most suitable materials to use, sorted between easy-to-manufacture and high-performance materials.

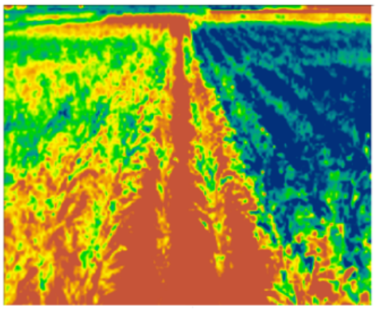
The ray model is once more a powerful and innovative optimisation tool that can quickly estimate the best MLDOE shape, given a reference equation (extended scalar theory). We also demonstrate the limits of our model compared to FDTD since it provides reliable but non-optimal solutions with variable accuracy. Nonetheless, thanks to the ray model, we still manage to design an efficient dual-band MLDOE using mouldable materials. Since we have finally found a suitable dual-band diffractive element but lack the proper software to integrate it into the MWIR camera, we focus on studying its manufacturing/assembling tolerances. We show, using FDTD, that the groove thicknesses must be manufactured under very tied tolerances of a few micrometres.

Part I

Part 1: Radiometric study







Chapter 1

Spaceborne thermal infrared remote sensing applications

Among the many infrared applications for Earth remote sensing, we detail the monitoring of crop water stress, wildfires and volcanoes.Infrared monitoring has enabled new irrigation strategies and innovative water management. Wildfires and volcano eruptions are hazardous local events that threaten human lives, health and ecosystems on a large scale. Their detection, monitoring and even prevention (in the case of wildfires) are make-or-break issues of this century. This chapter answers various earth remote sensing questions concerning the required spatial, spectral, thermal, and temporal resolution of the imaging system. As a result, a 10 m ground sampling, imaged daily in dual-band MWIR/LWIR infrared at a 1 °K temperature resolution, enables all these applications simultaneously. We propose to use a constellation of small satellites to ensure the essential daily coverage.

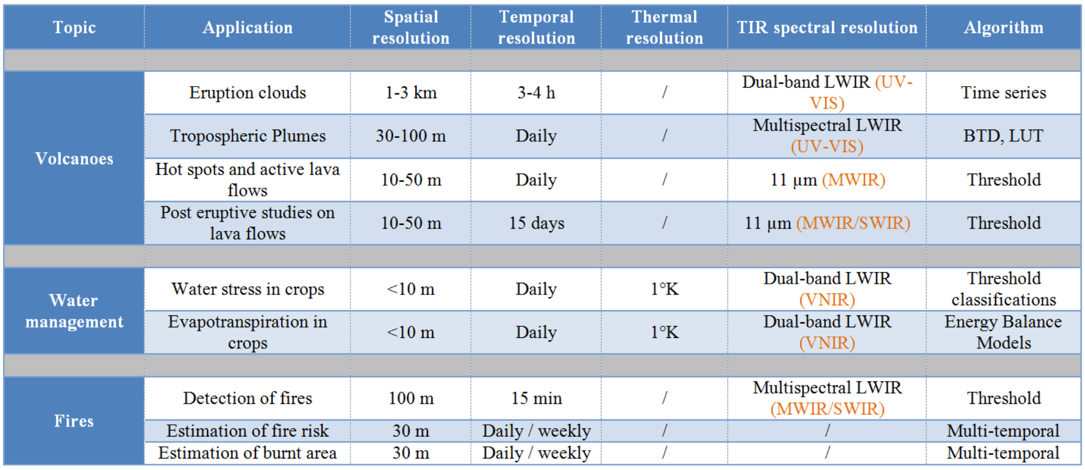
Contents

1.1 Thermal infrared applications and requirements

This section is based on the Fuegosat synthesis study [1], funded by the European Space Agency (ESA). This introduction aims to classify many infrared (IR) remote sensing applications, providing a complete set of requirements (spatial, temporal, thermal, spectral resolutions, algorithms). Thermal remote sensing applications are classified into three categories: land & solid Earth, health & hazard and security & surveillance. A non-exhaustive list of those applications is given [1]:

- **Land and solid Earth**:
  - **Water management** (Section 1.2): Detection of water stress and evapotranspiration retrieval are key applications for water management purposes. Water stress is a major environmental factor limiting plant productivity worldwide. It occurs when evaporative losses cannot be sustained by the roots' extraction of water from the soil. Evapotranspiration describes the loss of water from the Earth's surface to the atmosphere (evaporation from the surface combined with transpiration from the vegetation)
  - **Fire monitoring** (Section 1.3): Wildfires are a major security hazard in numerous countries around the world. They are responsible for vegetation resource and property losses, global warming and air pollution.
  - **Volcano monitoring** (Section 1.4): Volcanic eruptions pose serious hazards to vegetation ecosystems, urban and communication networks, and threaten populated regions.
  - Earthquakes prevention
  - Coal mine fires
- **Health and hazard**:
  - Urban heat island
  - Epidemiology
  - Coastal inundation
  - Industrial risks
- **Security and surveillance**:
  - Detection of minefields
  - Ship and port monitoring
  - military

In this thesis, we will focus on three "land and solid Earth" topics that have high "priority" [1]: water management, fire and volcano monitoring. Their requirements are detailed in Table 1.1:



Topic	Application	Spatial resolution	Temporal resolution	Thermal resolution	TIR spectral resolution	Algorithm
Volcanoes	Eruption clouds	1-3 km	3-4 h	/	Dual-band LWIR (UV-VIS)	Time series
	Tropospheric Plumes	30-100 m	Daily	/	Multispectral LWIR (UV-VIS)	BTD, LUT
	Hot spots and active lava flows	10-50 m	Daily	/	11 µm (MWIR)	Threshold
	Post eruptive studies on lava flows	10-50 m	15 days	/	11 µm (MWIR/SWIR)	Threshold
Water management	Water stress in crops	<10 m	Daily	1°K	Dual-band LWIR (VNIR)	Threshold classifications
	Evapotranspiration in crops	<10 m	Daily	1°K	Dual-band LWIR (VNIR)	Energy Balance Models
Fires	Detection of fires	100 m	15 min	/	Multispectral LWIR (MWIR/SWIR)	Threshold
	Estimation of fire risk	30 m	Daily / weekly	/	/	Multi-temporal
	Estimation of burnt area	30 m	Daily / weekly	/	/	Multi-temporal

Table 1.1: List of "land and solid Earth" topics tackled in this thesis [1], with their respective requirements. The spectral bands written in orange are complementary wavebands that improve IR detection.

In the following, we provide a detailed explanation of the topics mentioned above (water, fire and volcano management).

1.2 Crop management

1.2.1 Vegetation remote sensing

Water is essential for plants' growth, metabolic processes, and general health. Water stress is one of the most critical limitations for plant growth, crop yield and food production quality. Currently, about 80% to 90% of Earth's global freshwater is used by humans, and about two-thirds is consumed by agriculture for crop irrigation [2]. Due to global warming and the increasing world population, agricultural water demand will further increase. Therefore, a challenging task of this generation is to reduce the amount of water used per unit yield, ensuring global food security. Precision agriculture offers new farming methods to improve the sustainability of crop production.

Remote sensing offers a non-destructive method for measuring numerous physiological, biochemical and structural crop characteristics (drought, disease, freeze) at different scales. Thus, it is one of the key technologies used in precision agriculture [2]. The detection of plant response to water stress involves multispectral remote sensing imaging: Long wave infrared (LWIR; 8 – 14 µm), visible (VIS), near and shortwave infrared reflectance (VNIR/SWIR; 0.4 – 2.5 µm), and sun-induced fluorescence (SIF; 0.69 and 0.76 µm). The entire infrared spectrum is depicted in Figure 1.1.

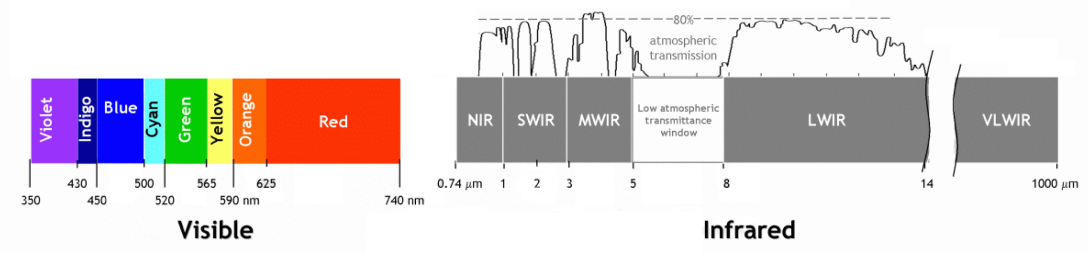


Figure 1.1: Infrared spectrum.

Crop fields display two vegetation types: the growing plant's leaves and the soil. Green plant leaves have very low reflectance and transmittance in the visible (VIS) due to strong absorbance by pigments (chlorophyll) [3], which is not the case in the near-infrared (NIR) waveband. Interestingly, plant stress results in lower chlorophyll concentration, broadening the VIS reflectance spectrum and decreasing the NIR reflectance. Compared with growing plants, most agricultural soils display a monotonic reflectance increase throughout VIS and NIR [3]. Highly watered soils have a decreased reflectance, while dry soil appears brighter. Therefore, many studies and vegetation index definitions use the VIS and NIR wavebands. Vegetation indices provide quantitative metrics describing a plant's status and are extremely valuable for extracting plant signals from complex canopy spectra. They are often based on differences, ratios or linear combinations of reflected light in VIS and NIR wavebands [3]. For instance, a well-known vegetation index called the "normalised difference vegetation index" (NDVI) is expressed as:

$$NDVI \equiv \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \tag{1.1}$$

where R_{NIR} and R_{RED} are NIR and red reflectances. The NDVI performs exceptionally well as a quantitative metric for tracking green biomass or leaf area index through the season or for detecting uneven patterns of growth within a field [3]. The more specific the index (narrow bandwidth), the more correlated to a particular physiological plant response (and diagnostic) it will be.

In addition to these wavebands, thermal infrared spectra are of great use in identifying temperature-related phenomena such as evapotranspiration or water stress. Plant leaves and soils, having an average temperature of 300 °K, emit most of their infrared radiation in the long wave infrared (LWIR) due to Planck's law (see Chapter 3). Therefore, many thermal indices, such as the "crop water stress index" (CWSI), the "stress-degree-day" (SDD) or the "water deficit index" (WDI), have been created.

Finally, it is worth mentioning that plant response to environmental stress is not straightforward, as it is heavily impacted by illumination, topography, and meteorological conditions [3].

1.2.2 Water stress

Water management refers to determining irrigation schedules and good practices. It is based on plant status monitoring, evapotranspiration rates and crop coefficients. The closure of a plant's stomata is triggered as it depletes the soil water reserves (or if the atmosphere is too dry). As a result, its transpiration cooling reduces, increasing its temperature (relative to ambient air temperature / well-watered crop) [2, 3]. The temperature variation range depends on the atmosphere evaporative demand and on the transpiration characteristics of the plant species. Water stress generally causes physiological and biochemical changes to plants, depending on the severity and duration of the water deficit. Prolonged or severe water stress leads to irreversible damage to the photosynthetic machinery [2]. Figure 1.2 provides a chart linking the various plant stresses to the most appropriate detection wavelengths:

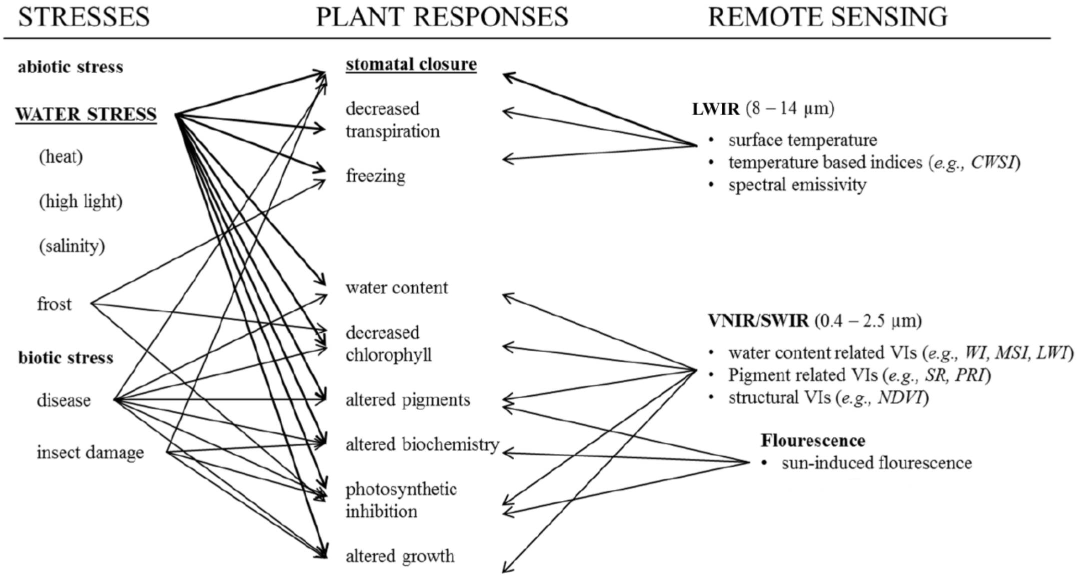


Figure 1.2: Most important relationships between primary plant stresses, the induced plant responses, and the multispectral remote sensing techniques. The acronym "VIs" refers to "vegetation index". Original source: [2].

As previously stated, leaf transpiration reduces when plant water status decreases due to active stomatal regulation. Consequently, the inhibited evaporative cooling effect leads to higher leaf and canopy temperatures than in a well-watered plant. Usually, the leaf temperature of a fully transpiring plant is about 2 – 5 °K below the ambient air temperature [2]. Thermal remote sensing of leaf and canopy temperature is an established technique to detect pre-visual water stress effects. However, temperature-based approaches suffer from numerous noises: air temperature, humidity, vapour pressure deficit (VPD), wind speed and incident radiation. Many water stress indices allow for quantitative assessment of plant water stress by normalising measured leaf temperatures to environmental conditions [2].

Stress Degree Day (SDD)

The "Stress Degree Day" (SDD), measured 1h30 after solar noon, is based on the difference between canopy T_c and air T_a temperatures and assumes that plant water stress appears whenever $T_c - T_a > 0°K$. The air temperature must be measured in-situ by a probe.

Crop Water Stress Index CWSI

The most commonly used improvement of the SDD is called the "Crop Water Stress Index" (CWSI) [4]:

$$CWSI \equiv \frac{(T_c - T_a)_M - (T_c - T_a)_{LL}}{(T_c - T_a)_{UL} - (T_c - T_a)_{LL}} \tag{1.2}$$

where the subscripts M, LL and UL respectively denote measured, lower limit (no stress, maximum potential transpiration rate), and upper limit (complete stress, no transpiration / closed stomata) [2, 4]. The CWSI does not only consider air temperature but also other environmental factors (wind, radiation, VPD) and can theoretically be adapted to any crop under any meteorological conditions thanks to the boundary definitions. Wind speed is the primary CWSI erroneous factor since it increases evapotranspiration, decreasing the temperature differences between a canopy and ambient air [5].

For instance, the CWSI is applied to determine the best irrigation approach in a maize field in [6]. New "deficit irrigation" strategies have been proposed for sustainable water management, especially in arid/semi-arid regions. The crops are being allowed to experience some degree of water stress. They must be accurately monitored to avoid their water stress from intensifying higher than a safe level, jeopardising the crop yield. In the proposed example [6], HFDI and LFDI irrigation approaches are studied, standing respectively for high and low-frequency deficit irrigation. The former consists of seven low-depth irrigation events, while the latter consists of three high-depth irrigation events. The CWSI is computed between July and September 2012 for both irrigation approaches and presented in Figure 1.3:

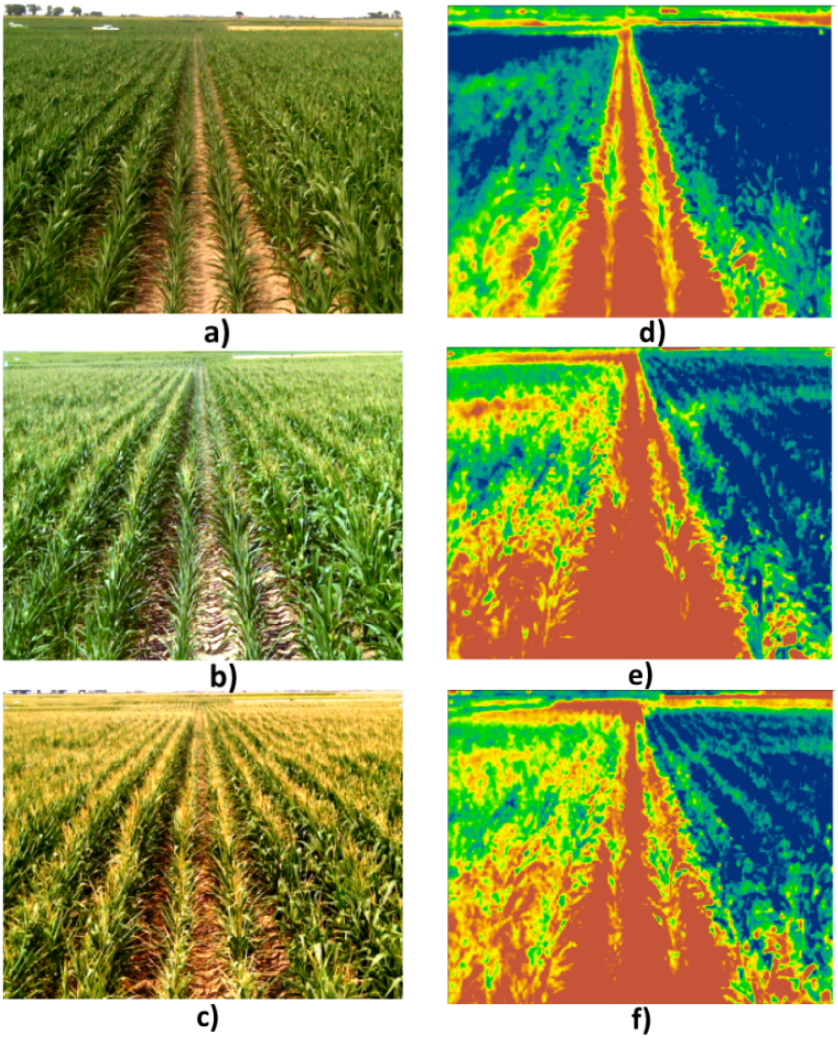


Figure 1.3: Maise test field imaged on three dates (from top to bottom): July 19, July 26 and August 9/2012. Images (a-c) are taken in the visible, while images (d-f) correspond to the CWSI computed from thermal images. The maize field was separated into HFDI (right) and LFDI (left). The centre rows delimit the two irrigation approaches and have not been irrigated, displaying high CWSI. Visible images do not provide differences between the two deficit irrigation methods, contrary to the CWSI. Original source: [6].

Beyond determining the best irrigation strategy, Figure 1.3 highlights the advantages of using infrared imaging over visible imaging when studying crop fields.

Thermal infrared wavebands and the CWSI are especially useful for detecting acute water stress. The CWSI provides adequate lead time for irrigation scheduling in hot and dry climates with high VPD, low wind speed variability, and low cloud cover [2, 4]. Consequently, the CWSI is less responsive to plant and soil water conditions in humid locations and is most responsive during clear skies and within a few hours of solar noon.

Finally, it is prevalent to have partial canopy cover, especially at the beginning of the irrigation season. The temperature of dry, sunlit soil can be 30 °C greater than green, transpiring vegetation, posing an overestimation bias to the CWSI. The temperature of shaded soil is also usually different from vegetation, which may also introduce errors in CWSI calculations [4]. The relationship between canopy and soil temperatures is given by [5]:

$$\begin{cases} T_b^4 = f_c T_{b,c}^4 + (1 - f_c T_{b,s}^4) \\ T_c = T_{b,c} . \epsilon^{-1/4} \end{cases} \tag{1.3}$$

where T_b is the brightness composite temperature, $T_{b,c}$ is the brightness canopy temperature, $T_{b,s}$ is the brightness soil temperature, f_c is the proportion of canopy cover and T_c is the canopy temperature, deduced from $T_{b,c}$ using the canopy emissivity $\epsilon \simeq 0.96$.

Water Deficit Index (WDI)

The water Deficit Index (WDI) is an extension of the CWSI that accounts for the soil temperature and the partial canopy cover. The WDI is expressed similarly to Equation 1.2, replacing the canopy temperature T_c by the surface temperature T_s, referring to a composite surface that may include both canopy and soil. Figure 1.4 graphically represents this index as a trapezoid [4].

The WDI is more complicated to obtain than the CWSI since it requires reflectances measurements (NDVI, red and NIR wavebands) to estimate the fraction of vegetation f_{cr}.

Time-temperature threshold (TTT)

The time-temperature threshold (TTT) method is based on the fact that plant enzymes are most productive under a relatively narrow range of temperatures, called the thermal kinetic window [4]. This method is beneficial for irrigation scheduling and only requires simple temperature measurements. In the TTT approach, the accumulated time for which the canopy temperature exceeds a threshold temperature is used as a criterion to start an irrigation event. This time-integrated method is more robust than the CWSI and the WDI to meteorological variations.

A significant drawback of the TTT method is that it requires continuous canopy temperature measurements throughout the daytime. It can be achieved at a fixed location but is complicated and costly to provide at a remote location, when crop fields are extended. However, the diurnal canopy temperature for remote locations can be predicted using a scaling procedure based on the continuous temperature measurement at a fixed location [4]:

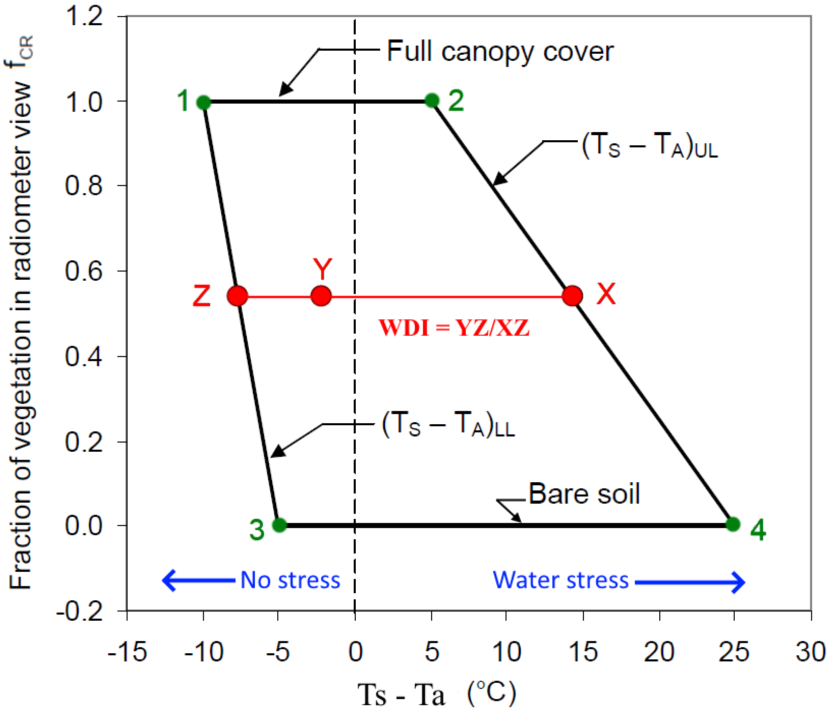


Figure 1.4: Water Deficit Index (WDI), defined as $WDI = YZ/XZ$. Point 1 represents a non-stressed fully covered canopy, point 2 represents a water-stressed fully covered canopy, point 3 corresponds to wet bare soil and point 4 corresponds to dry bare soil. The lower and upper temperature limits are respectively $(T_s - T_a)_{LL}$ and $(T_s - T_a)_{UL}$. The measured composite surface temperature (T_s) is displayed at point Y. Adapted from [4].

$$T_c(t) = T_E + \frac{(T_c(t_0) - T_E)(T_{REF}(t) - T_E)}{T_{REF}(t_0) - T_E} \tag{1.4}$$

where $T_c(t)$ is the remote field location predicted canopy temperature at any time of day. T_E is the predawn canopy temperature, $T_c(t_0)$ is a one-time-of-day measurement taken at a remote field location and $T_{REF}(t)$ is the stationary location reference temperature measurement throughout the day. The scaling method enables calculating water stress indices over a longer portion of the day (time-integrated water stress index) and automatise irrigation systems [4]. The TTT method improves crops and water management, providing the timing for irrigation events. However, the exact amount of irrigation needed by the soil remains unknown and depends on the soil water depletion in the root zone.

1.2.3 Evapotranspiration (ET) - Soil Water Balance

Soil water depletion is a crucial parameter that provides the necessary irrigation at a particular location. According to [4], its measurement is complex and requires expensive and radioactive in-situ probes (neutron probe), limiting the measurement frequency and coverage. Therefore, soil water depletion is usually calculated between measurement times using a soil water balance, where evapotranspiration (ET) is the primary sink [7]. Evapotranspiration combines evaporation and transpiration (vaporisation of liquid water contained in plant tissues). When the crop is young, water is predominately lost by soil evaporation, but once the crop is well developed and completely covers the soil, transpiration becomes the main process [7]. Soil evaporation and crop transpiration deplete water from the root zone. If all fluxes other than evapotranspiration (ET) are known, the evapotranspiration can be deduced from the change in soil water content (ΔSW):

$$ET = I + P - RO - DP + CR \pm \Delta SF \pm \Delta SW \tag{1.5}$$

Irrigation (I) and rainfall (P) add water to the root zone. Part of I and P might be lost by surface runoff (RO) and by deep percolation (DP). Water might also be transported upward by the capillary rise (CR) towards the root zone [7]. Figure 1.5 depicts the soil water balance:

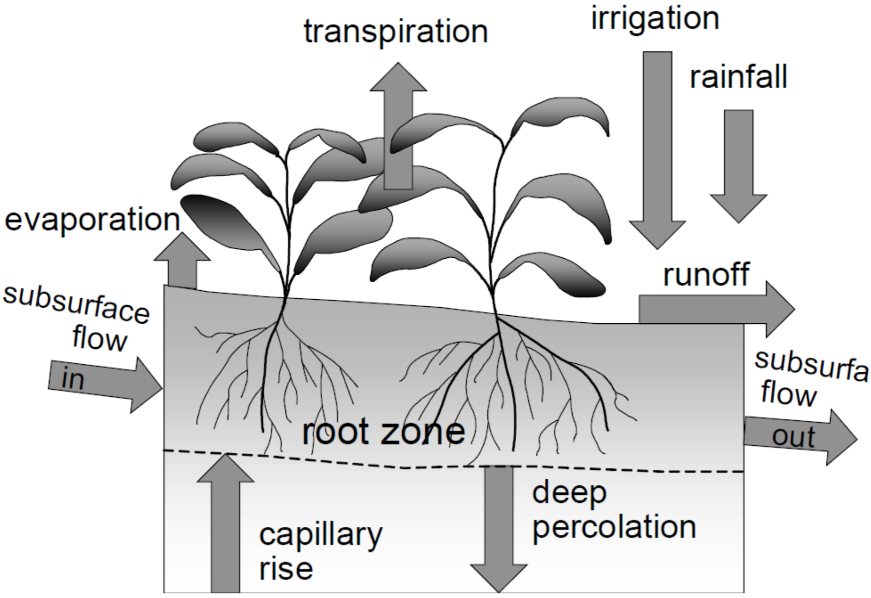


Figure 1.5: Soil water balance method for estimating evapotranspiration. Original source: [7].

The soil water balance method can usually only give ET estimates over long periods (ten days). ET can also be calculated using the canopy temperature in an energy balance model. In this approach, canopy temperature measurements provide the real-time feedback aspect. Since water stress indices are also derived from energy balance considerations, they are related to ET in the following general form [4]:

$$ET = ET_p(1 - WSI) \tag{1.6}$$

ET_p is the potential ET when water is non-limiting (i.e. $WSI = 0$). Therefore, when conditions are favourable, thermal remote sensing can compute the evapotranspiration and inject it into a soil balance model to retrieve the soil water content (ΔSW). It enables management strategies to reduce evapotranspiration losses and increase water use efficiency.

In conclusion, to be useful for day-to-day, site-specific irrigation management, canopy temperature data must have a spatial resolution of a few meters, a measurement frequency of at most a week, and a turnaround time (i.e. the time from measurement to useful information product) of a few minutes [4]. In addition, the field coverage must have an adequate number of samples with enough spatial distribution. The thermal resolution must be approxiamtely 1 °K to detect subtle temperature change and provide accurate water stress map (Table 1.1).

1.3 Wildfire remote sensing

Wildfires significantly impact environments and communities by changing the vegetation composition, affecting the soil characteristics, the hydrologic regimes, causing property losses and human casualties. The global climate is also affected by fire regimes, increasing the vegetation's susceptibility to wildfires and the amount of CO_2 and aerosol releases [8]. Between 1997 and 2008, global vegetation fires cumulatively burned 44.5 million km^2, equivalent to the combined area of North and South America (40% of the Earth's vegetated area) [9]. The origin of this biomass burning is multiple: planned or left to spread, under human control or accidental/caused by natural phenomena. On average, 3.4% of the Earth's vegetated area burns annually, resulting in large-scale effects on land surface properties, atmospheric composition, and human health and security. Thereby, wildfires release one-quarter of the global annual CO_2 industrial emissions [9], which has globally detectable effects on the atmosphere composition through the release of various chemical compounds. On a regional scale, important fire events threaten human lives and properties, considerably impacting national budgets. Air quality is degraded due to significant smoke and haze events, affecting human health.

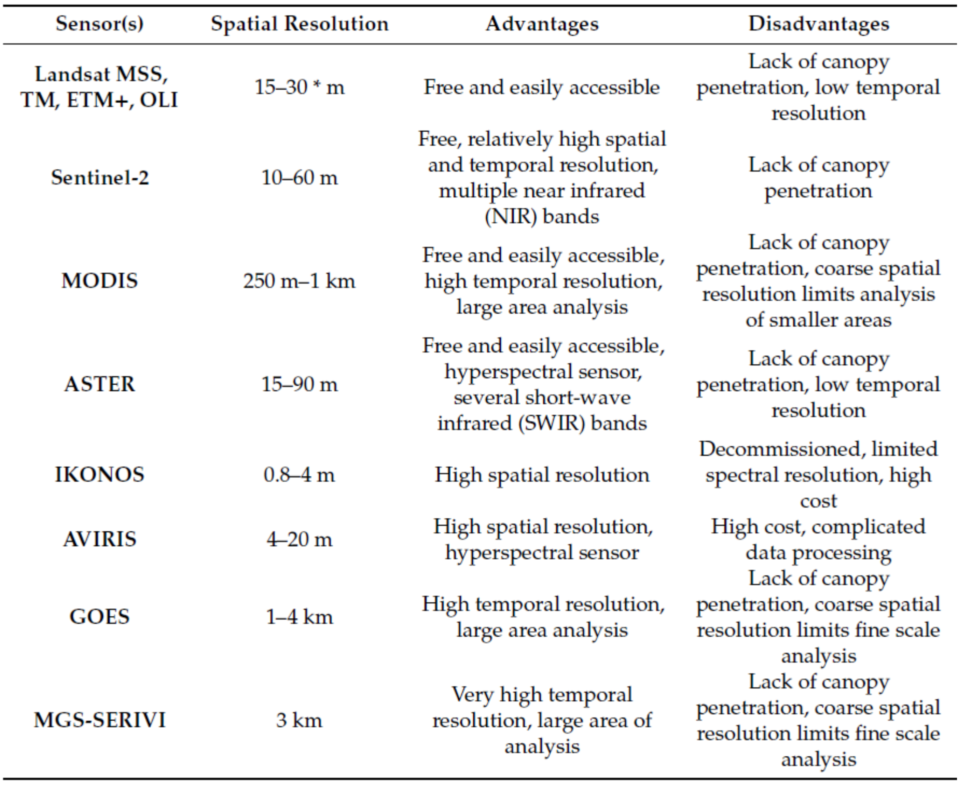
Vegetation fires occur sporadically over vast areas and are rapidly evolving in time. They have an international scope but often occur in isolated regions, where their detection and prevention are difficult. Consequently, thermal remote sensing constitutes a crucial tool in their identification and study: fire extent mapping, fire regimes determination, characterisation of fire impact and severity, estimation of fire emissions chemical composition and vegetation recovery monitoring [8, 9].

Table 1.2 provides common multispectral spaceborne instruments used for fire ecology [8].

The light detection and ranging (lidar) approach is worth mentioning for creating highly accurate digital terrain models and classifications [8]. However, lidar is out of the scope of this thesis, and this section only treats thermal and visible fire remote sensing. Similarly, the study of fire risk and fuel mapping (where and how fires occur) is not detailed in this thesis since it relies on very high-resolution multispectral images in the visible (VIS) and near-infrared (NIR) [8].

1.3.1 Physics of fire remote sensing

Vegetation combustion is a complex non-uniform process involving chemical reactions, coupled heat/mass transfers. Biomass burning generates significant energy releases, including radiant energy, detectable using the appropriate sensor and detection method. It is estimated in [9] that 10% to 20% of a fire's energy is radiated away from the combustion zone in the form of electromagnetic radiation of multiple wavelengths. Fires radiate as blackbodies, following Planck's radiation law (see Equation 3.1). Thermal remote sensing has some significant advantages compared to visible remote sensing due to the smoke screen generated by burning vegetation. Visible radiations are strongly hindered by the overlying smoke of already burned fuels, and only actively burning fires are detectable. On the contrary, Figure 1.6 shows that the smoke becomes increasingly transparent at longer IR wavelengths:



Sensor(s)	Spatial Resolution	Advantages	Disadvantages
Landsat MSS, TM, ETM+, OLI	15–30 * m	Free and easily accessible	Lack of canopy penetration, low temporal resolution
Sentinel-2	10–60 m	Free, relatively high spatial and temporal resolution, multiple near infrared (NIR) bands	Lack of canopy penetration
MODIS	250 m–1 km	Free and easily accessible, high temporal resolution, large area analysis	Lack of canopy penetration, coarse spatial resolution limits analysis of smaller areas
ASTER	15–90 m	Free and easily accessible, hyperspectral sensor, several short-wave infrared (SWIR) bands	Lack of canopy penetration, low temporal resolution
IKONOS	0.8–4 m	High spatial resolution	Decommissioned, limited spectral resolution, high cost
AVIRIS	4–20 m	High spatial resolution, hyperspectral sensor	High cost, complicated data processing
GOES	1–4 km	High temporal resolution, large area analysis	Lack of canopy penetration, coarse spatial resolution limits fine scale analysis
MGS-SERIVI	3 km	Very high temporal resolution, large area of analysis	Lack of canopy penetration, coarse spatial resolution limits fine scale analysis

Table 1.2: List of sensors commonly used in fire remote sensing ecology. While some sensors have very high spatial resolutions (low Earth polar orbit), others have high spatial resolutions, such as SEVIRI, operating from a geostationary orbit.

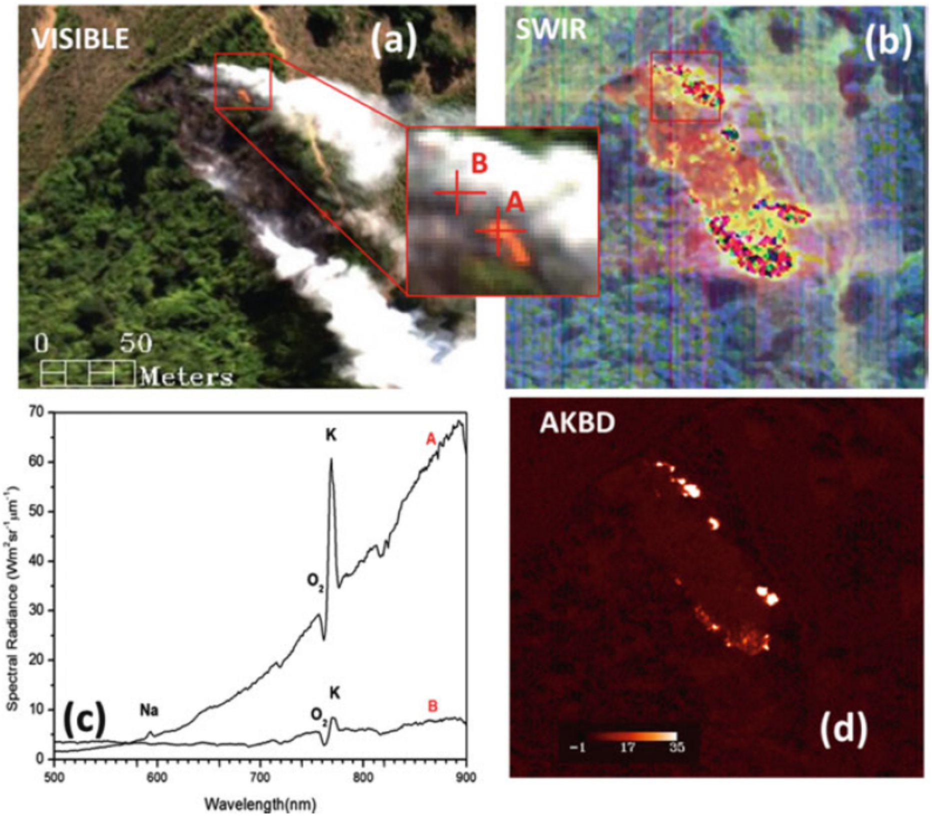


Figure 1.6: Forest fire remote sensing multispectral data from [9]. a) Composite VIS image of a flaming combustion area. b) The false colour SWIR image of the same area illustrates the ability to penetrate the smoke and discriminate between the actively burning area A and the already burned area B. c) Spectra at locations A (flaming smoke-free fire) and B (smoke-covered fire), identified in the inset of a). d) The "Advanced potassium band difference" (AKBD) metric, measured in the same fire area, allows for retrieving the burning areas underneath the smoke.

In Figure 1.6, SWIR wavebands easily identify the burning fuel area and the reflected solar SWIR radiation from areas of already burned vegetation, even through smoke. NIR line emission of burning vegetation allows for tracing chemical elements, such as potassium and sodium (Figure 1.6 d). NIR can identify specific areas of flaming activities through the smoke. However, the most reliable fire remote sensing detection wavebands,given by Plank's law, are the middle wave infrared (MWIR: 3-5 µm) and long wave infrared (LWIR 8-14 µm). Generally, the MWIR waveband is used for fire emissions that peak around $600°K$. It has a lower solar radiation contribution than the SWIR waveband. LWIR radiations enhance fire detection methods since they can go through the full depth of Earth's atmosphere and even through significant smoke, as depicted in Figure 1.7:

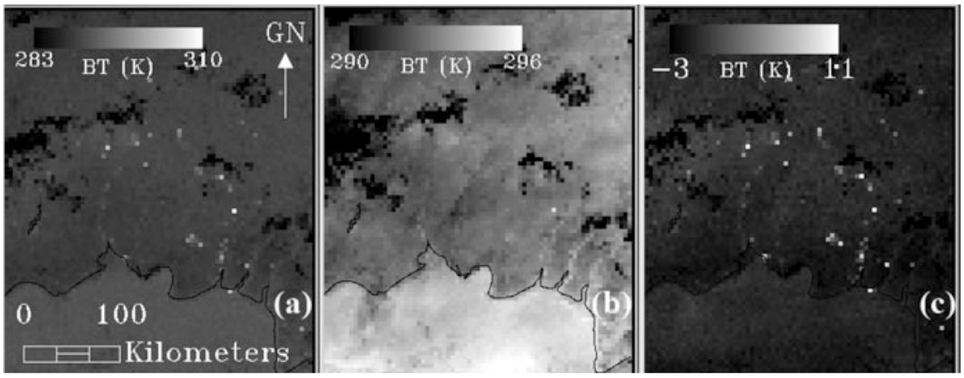


Figure 1.7: (a) MWIR and (b) LWIR forest fire remote sensing data from [9]. The nadir spatial resolution (AVHRR sensor) is 1.1 km, meaning that fires are highly sub-pixel events. Using the inverse Planck's law, the spectral radiance measurements have been converted to brightness temperatures (BT). MWIR pixels display much larger integrated BT than LWIR pixels. c) The image difference between MWIR and LWIR BTs images allows for recovery of fire-affected pixels particularly well.

Measuring the thermal radiation during combustion is called "active fire" remote sensing, contrary to the passive solar reflectance observations used to detect burned scars. Taking $600°K$ and $1000°K$ as representative temperatures for smouldering and flaming combustion, a fire has much higher MWIR thermal emission than its background ($T \simeq 300°K$). Consequently, identifying actively burning fires only requires relatively simple detection algorithms. It also means that hotspots count and fire location maps can be rapidly delivered to the users. High spatial resolution is not needed since the MWIR pixel-integrated signal is significantly high. Though very detailed fire observations can be made through smoke, meteorological cloud cover remains a problem. Fortunately, most fire-affected regions have a fire season that coincides with a dryer period and a lower cloud cover.

1.3.2 Active fires spaceborne detection methods

As illustrated in Figure 1.6, flaming fires emit very significantly in the SWIR waveband. However, the strong daytime solar reflections and the fact that many fires occur during the day make the SWIR waveband less attractive than the MWIR waveband for fire monitoring. MWIR imagery is sufficiently sensitive to fire thermal emissions ($T \simeq 1000°K$) to detect even sub-pixel fires (Figure 1.7). In addition, cooler smouldering fires that might be very difficult to detect in SWIR appear in MWIR. A simple active fire detection algorithm would consist in applying a single threshold in the MWIR channel. However, due to solar heating of bare ground or specular sunlight reflections, the MWIR signal can reach the threshold even for non-fire pixels, resulting in "false alarms" [9, 10]. The discrimination is often made using additional spectral and spatial tests.

Instead of using the pixel-integrated spectral radiance metric, most active fire algorithms work with the brightness temperature metric (BT), calculated using the inverse Planck's function. Thereby, areas of solar-heated ground (vegetation, bare soil, rocks) exhibit similar BT in MWIR and LWIR. On the contrary, as previously explained, actively burning sub-pixel fires exhibit a higher MWIR BT. Using dual MWIR/LWIR channels is a common feature of active fire remote sensing algorithms. Other channels, such as the visible (VIS), enable sunglints discrimination from clouds or water bodies. An example of active fire pixel thresholds is provided in Equation 1.7 and has been used for AVHRR pixels [9]:

$$\begin{cases} BT_{MWIR} > 314^\circ K \\ BT_{MWIR} - BT_{LWIR} > 10^\circ K \\ BT_{LWIR}(\text{clear sky}) - BT_{LWIR} < 6^\circ K \\ BT_{LWIR} < 310^\circ K \end{cases} \tag{1.7}$$

Note that fixed threshold approaches are simple and work very well for individual scenes but provide poor performance during multi-regional/multi-seasonal/multi-sensor analyses. Hence, "contextual" active fire approaches are used in these scenarios, implying multiple algorithm steps. A pixel thresholding method determines potential fire pixels (PFPs). It is followed by statistical tests operated on PFP's immediate "ambient background" surrounding to confirm whether or not PFPs are true fire pixels [10]. Meteorological clouds and thick smoke plumes are generally quite opaque even at long infrared wavelengths, and may absorb a fire pixel's IR radiation to the point where it appears as a "non fire" pixel. Therefore, satellite-based active fire sensors also include cloudiness metrics for each pixel, in conjunction with land/water classification and a mask of the identified sunglint areas as depicted in Figure 1.8:

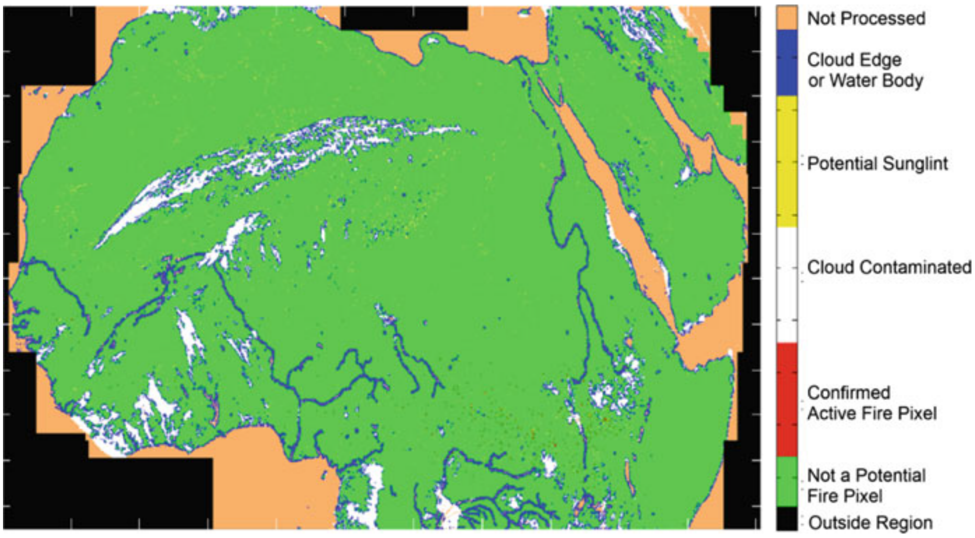


Figure 1.8: Fire masks in North Africa, created from Meteosat SEVIRI imagery [9]. The mask delineates the pixels processed by the active fire detection algorithm. Other information such as sunglint contamination is displayed.

In addition to polar orbiting instruments with a high spatial resolution (ASTER, 30 m), geostationary sensors can detect fires with improved temporal resolution but a much lower spatial resolution (Meteosat SEVIRI instrument has a 3 km spatial resolution and a 15 min temporal resolution). The maximum temperature difference between two consecutive scenes is $\pm 15^{\circ} K$ for a typical summer day in middle latitudes [10]. It becomes appropriate to consider System 1.7 with a fire detection threshold increased by 5 °K in MWIR to avoid any false alarm.

1.3.3 Burned area mapping

The relevance of spaceborne thermal remote sensing goes beyond the field of active fire detection. Such measurements can be of precious value in identifying and mapping burned areas. Burned area estimates provide accurate spatial representations of fire extents and perimeters. It is crucial to map wildfire-affected regions to draw rehabilitation planning, calculate the environmental and economical cost of fires and estimate gas and particulate emissions on a global/regional scale [8]. The NIR and SWIR wavelengths are commonly used for burned area mapping. Vegetation cover changes and bare soil / charred surface proportions significantly impact reflectances. A low albedo characterises burned areas, with exposed soil cover and presence of still smouldering combustion [9]. Burned areas (BA) are commonly estimated using active fire images, counting the number of fire pixels [8]:

$$BA(i,t) = a_{px} N_f(i,t) \tag{1.8}$$

where i is the grid cell, t is the time period, N_f the number of detected fire pixels and a_{px} the pixel area. However, Equation 1.8 can be further enhanced using variable pixel areas, distinguishing between herbaceous or tree cover, for instance. A temporal index based on NIR data can also be used. It compares NIR pixel values at a given time ($S1_{NIR}$) to the average NIR reflectance (IC_{NIR}) for all observations prior to the considered time [8]:

$$BA = \frac{S1_{NIR} - IC_{NIR}}{S1_{NIR} + IC_{NIR}} \tag{1.9}$$

Appropriate sensors for burned area determination depend on the considered scale. At the local scale, the same sensors as for burn severity mapping (Section 1.3.4) are used ($<$ 100 m spatial resolution). At regional and global scales, coarse spatial resolution is appropriate ($>$ 100 m) since the mapping of large areas over a short time ($<$ 2 days) is required. High temporal resolution is unnecessary, meaning that geostationary sensors are not appropriate [8].

1.3.4 Burn severity assessment

The *burn severity* metric measures the wildfire impact on the landscape: vegetation mortality, soil nutrient composition, and decreased infiltration due to soil hydrophobicity. The normalised burn ratio (NBR) is a spectral index calculated using SWIR and NIR satellite imagery data. The NIR waveband is sensitive to leaf structures of living vegetation, while the SWIR waveband is sensitive to moisture content and soil conditions. Fire-affected vegetation exhibits a decreased NIR reflectance and an increased SWIR reflectance [8]. The NBR is calculated pre and post-fire, and the difference NBR (dNBR) enables the identification of areas of significant change. Figure 1.9 illustrates the dNBR ability to determine various degrees of burn severity.

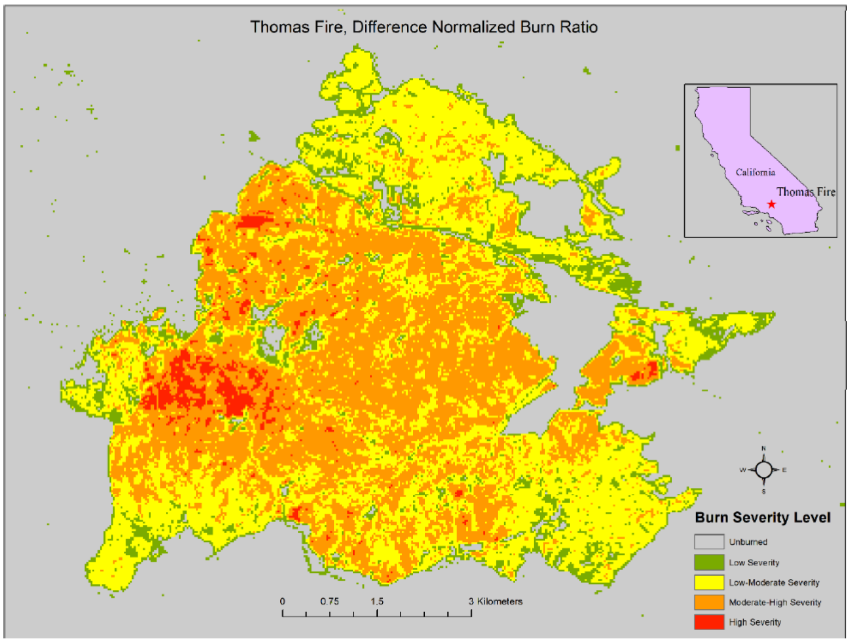


Figure 1.9: The difference normalised burn ratio (dNBR) shows the burn severity levels for the Thomas Fire, CA 2008. Using changes in NIR and SWIR reflectances, burn severity levels can rapidly be identified [8] .

Burn severity assessment requires imagery with a high to moderate spatial resolution (< 100 m), as coarser resolutions cannot detect burn severity patterns [8].

1.3.5 Fire characterization

The concept of fire monitoring involves the knowledge of fire temperature, active area, intensity, and front. This task requires high temporal resolution sensors, typically geostationary satellites, to follow the fire evolution throughout its development. The complete combustion of a fixed amount of biomass releases a relatively fixed amount of thermal energy: the *fuel heat yield* [9]. The radiated fraction of the fuel heat yield is called fire radiative power (FRP [MW]), and varies between 10% and 20%. Consequently, measuring a fire's IR radiation provides a way to estimate the amount of burned fuel. Approximating Planck's function with a power law, we get:

$$FRP = GSD(L_f - L_{bg})\sigma/a \tag{1.10}$$

L_f and L_{bg} denotes the fire and background emitted MWIR radiance [W.m^{-2}.sr^{-1}.µm^{-1}]. $\sigma = 5.67.10^{-8}$ [J.s^{-1}.m^{-2}.K^{-4}] is the Stefan-Boltzmann constant, GSD is the ground sampling distance (km^2) and a [W.m^{-2}.sr^{-1}µm^{-1}.K^{-4}] depends on the sensor spectral response. Equation 1.10 models a fire as a greybody with constant spectral emissivity. The FRP allows estimating the fire's destructive power and atmospheric emissions. For high spatial resolution sensors, the sub-pixel analysis can identify the fire direction and the flaming front, as displayed in Figure 1.10:

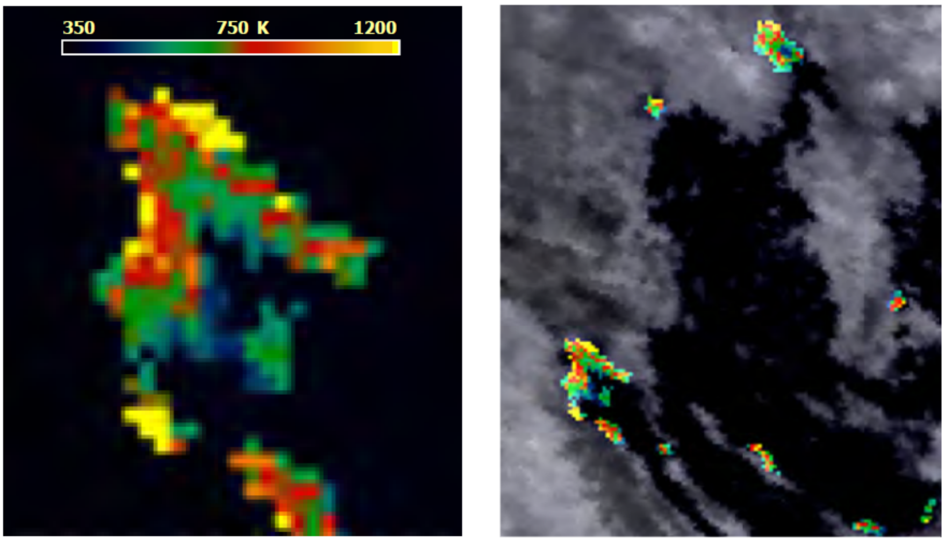


Figure 1.10: Fire temperature, obtained using Equation 1.10 with the high resolution sensor MODIS (left picture). With enough resolution, the flaming front and fire direction appear [10] (right picture).

The real interest of IR remote sensing in early fire detection arises when the sensor time resolution is $\leqslant$ 15 min, which is currently only possible with geostationary satellites (SEVIRI). The amount of burned biomass is directly linked to the fire radiative energy (FRE, [MJ]) [10], expressed as:

$$FRE = \int FRP \, \mathrm{d}t \tag{1.11}$$

Geostationary satellites fail at detecting lower FRP fires, which are the most frequent. An efficient solution is to blend FRP data with burned areas maps (Section 1.3.3).

1.4 Volcano monitoring

Volcano remote sensing is a primary scientific focus, reflecting that 500 million people live under the direct or indirect threat of volcanic eruptions. Warning signs of eruptions are inconsistent, and prediction time windows are inaccurate [11]. The study of heat fluxes, temperature variations and cooling rates in active volcanic areas provides a fundamental contribution to understanding volcanic processes and hazards. Thermal data is valuable to detect precursors of eruptions onsets, combined with geophysical and geochemical phenomena [11, 12]. For instance, temperature variations of fumaroles, open-conduit vents, fractures, groundwater, and lava domes are often followed by eruptions [12]. Effusive eruptions, resulting in lava flows, are well characterised by the lava temperature, among other magma physical properties. These properties play a crucial role in determining lava emplacement and flow length. For instance, the presence of lava tubes provides magma insulation enabling much longer lava flows [12], accentuating the risks for human lives and properties.

Spaceborne remote sensing provides tracking and mapping of thermal anomalies in inaccessible active volcano areas. In particular, thermal cameras represent powerful tools to estimate surface temperature, ejecta velocity, cooling and effusion rates, and heat/mass fluxes. The MWIR waveband is particularly useful for volcanological applications since it is susceptible to high temperatures (magma, fumaroles) and is less affected by atmospheric attenuation (volcanic gases). In addition, the LWIR waveband is beneficial for removing solar reflection, aerosol and ash optical effects. Daily records of thermal data have enabled us to fix temperature thresholds for surveillance purposes and to gain more profound knowledge about many volcanic phenomena: lava lake convection and degassing, explosive activity, lava flowing within tubes, etc. [12].

Numerous factors affect the accuracy of apparent temperature measurements using thermal cameras [12]:

- Emissivity, viewing angle and target roughness.
- Solar reflection and solar heating: The MWIR waveband is more polluted by solar reflection than the LWIR. Solar heating reduces the contrast between volcanic anomalies and their background, rendering them indistinguishable, especially for relatively cool targets.
- Atmospheric effects, volcanic gas, aerosols and airborne ash: These particules absorb infrared radiations, particularly in LWIR, suggesting the suitability of the MWIR bandwidth for volcanological applications.
- Mixed pixel problem: The targeted anomaly temperature is diluted with a different background temperature due to thermal heterogeneity at sub-pixel level.

A standard approach assumes that mixed pixels include only two thermal components: a hot target and a cold background. The dual-band MWIR/LWIR method is used to retrieve sub-pixel radiant temperatures and the portion of the pixel occupied by each of the two thermal components. The dual-band method is beneficial for estimating lava surface temperature (composed of hot glowing cracks and cool homogeneous crust), analysing fumarolic activity, studying lava domes growth, and modelling lava flow core temperature [12]).

The review [12] describes many volcano monitoring applications, falling into three categories:

- Monitoring and analysing non-eruptive/resting volcanoes to look for anomalous behaviours
- Searching for thermal anomalies that might represent eruption precursors
- Monitoring of ongoing eruptions

This section depicts two particular applications of thermal remote sensing for volcano monitoring.

1.4.1 Lava bodies

Detecting hot volcanic features such as lava flows, domes, and lakes using visible imaging is often impractical and unsafe. Sometimes, the lava bodies are masked by thick curtains of volcanic gas and airborne ash, hampering observation. In such cases, the use of thermal cameras can be valuable for eruption surveillance, for locating and distinguishing active lava, as depicted in Figure 1.11:

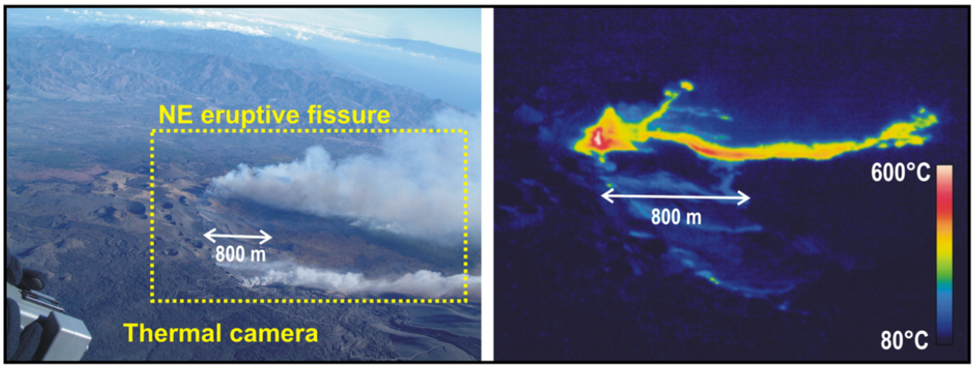


Figure 1.11: Activity of Mount Etna's northeastern eruptive fissure on 28 October 2002. The yellow-dashed rectangle in the visible photograph shows the area captured by the thermal image. In the photograph, ash and gas emissions hide much of the scene. The thermal image identifies two vents feeding strombolian activity and lava flows along the lowest segment of the eruptive fissure. Original source [12]

The daily data produced by in-situ or orbiting thermal imagers are essential aids for civil protection in the field of lava flow monitoring and vent maps. Thermal imagery enables the identification and mapping of a wide range of lava flow feature [12]: eruptive fissures, eruptive vents, active lava flows at different stages of an eruptive episode, morphology of lava surface and lava channels, lava tube. Many lava lake studies use thermal imaging and spectro-radiometry, revealing magma characteristics such as convection and heat radiation processes, styles of magma activity, magma motion, crust evolution and ageing.

1.4.2 Explosive activity

Thermal investigation of explosive activity allows one to study eruptive dynamics such as photoballistics and plume expansion. Observing the crater morphology is valuable for predicting the explosion type, magnitude, and kind of expelled materials [12]. In pre-eruptive phases, infrared images allow us to visualise the development of thermal anomalies such as crater fractures, temperature and intensity increase within open-conduit systems. Thermal images have proven useful to detect and discriminate different eruptive styles and to track the explosion timing [12, 13]. The main eruption types, as well as the particle ballistics, can be deduced from thermal images, as depicted in Figure 1.12 for the Stromboli volcano (Italy).

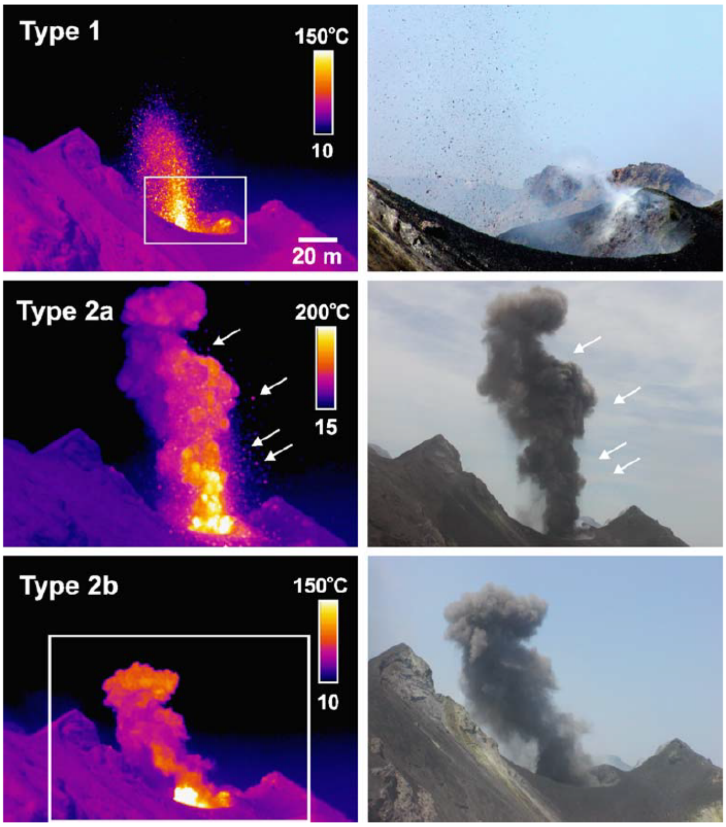
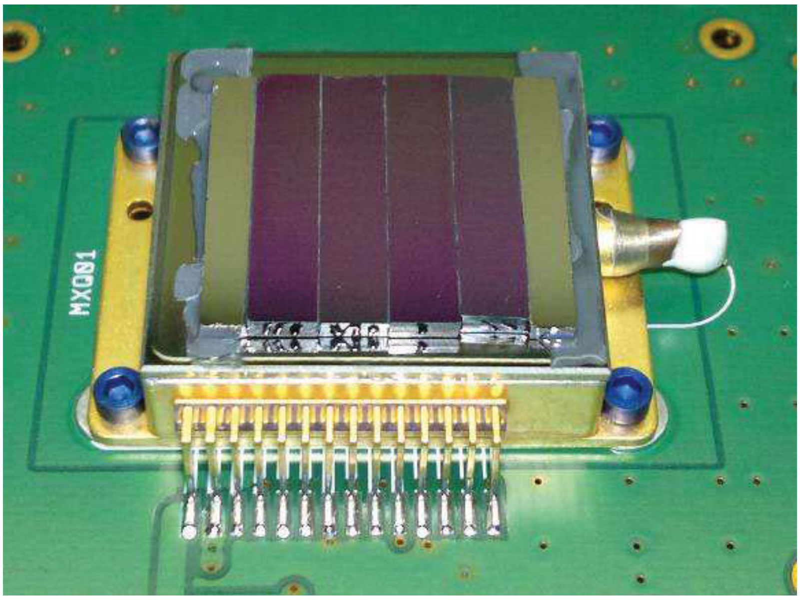


Figure 1.12: The main types of eruptions at Stromboli. Type 1 eruptions are ballistic-dominated, with little to no visible plume due to the absence of ash-sized particles. Type 2 eruptions involve the emission of an ash plume, with Type 2a containing significant ballistic particles and Type 2b having few ballistics. Arrows mark several large airborne particles. Original source: [13]

The qualitative explosive eruption classification provided in [13] for the Stromboli volcano can be further refined using thermal imagers, providing the plume height, velocity, collimation and particle type. Quantitative measurement can retrieve the maximum ballistic height, the eruption duration, the plume velocity and acceleration trends.

1.5 Conclusion

Based on three high-priority thermal infrared applications (crop management, wildfire detection and volcano monitoring), the main requirements, physical models and parameters, and detection methods have been described in the previous sections. Our aim is not to provide exhaustive information about these topics but to understand the requirements for spaceborne infrared instruments. Table 1.1 describes the main resolution and related applications of the three topics mentioned above. This study shows that a dual-band MWIR-LWIR instrument with daily coverage, 1 °K thermal resolution and at least 10 m spatial resolution is sufficient for most detection and monitoring applications. Wildfire, hot spots and lava flow require coraser thermal and spatial resolution, from 50 m to 100 m. Monitoring actively burning fires requires at least 15 min temporal resolution.


MX001

Chapter 2

Infrared detectors

We have defined a set of general requirements for three infrared Earth remote sensing applications in Chapter 1. Taking infrared images requires an optical assembly, which is the main topic of this thesis, and a detector. The latter constrains the fundamental optical parameters since its detectivity, minimum detectable temperature difference (noise equivalent difference temperature NEDT), and signal-to-noise ratio (SNR) are closely related to the selected applications. Thereby, this chapter presents infrared (IR) detector types, characteristics, noises and figures of merit. This chapter is not a comprehensive review of infrared technologies but introduces many necessary equations used in Chapter 3 radiometric budget. Finally, we propose a short state-of-the-art review of dual-band IR detection technologies.

Contents

This chapter introduces infrared (IR) detection, the main IR detectors' noises, and their performance metrics. IR detectors are transducers that sample the incident IR radiation and produce an electrical signal proportional to the total flux incident on the detector surface. An IR detector comprises two main parts: the sensitive pixel matrix (FPA) and the read-out integrated circuit (ROIC), attached to the FPA. The ROIC is an electronic device that operates the photon-to-electrons conversion and records the electronic signal for each pixel. Therefore, it is possible to define an IR detector using two concepts: detection and read-out. Various FPA technologies and ROIC architectures are presented in the following. Detectors are classified according to:

- The spectral bandpass
- The FPA technology and material
- The ROIC architecture
- The performance and noises
- The operating temperature, the weight and the size

2.1 Infrared detectors overview

There are two IR detectors families: Quantum and thermal detectors. Both types are sensitive to photons but with different response mechanisms, leading to variations in speed, spectral responsivity, and sensitivity (defined in Section 2.3). Thermal detectors rely on a change in the sensing material's electrical or mechanical properties (e.g., resistance, capacitance, voltage, mechanical displacement). This change is caused by the heating effect of the incident IR radiation. Quantum detectors exploit the photoelectric effect: Incident photons are converted into conducting electrons within the material. An absorbed photon instantaneously excites an electron from the nonconducting state into a conducting state, causing a change in the electrical properties of the semiconductor material that an external circuit can measure [14].

2.1.1 Thermal detectors

The sensitive part (mainly Vanadium Oxide VO_x, amorphous Silicon αSi, Silicon Nitride Si_3N_4) is suspended on lags connected to a Silicon heat sink. An incoming IR radiation is absorbed, heating the FPA and affecting a thermosensitive parameter. This variation is converted to an electrical signal by a transductor and then processed by the ROIC [15]. Therefore, thermal detectors must be thermally isolated from their surroundings and have low thermal capacities to provide fast response [16]. Thermal detectors are classified according to the nature of their varying thermosensitive parameter:

- Resistance variation → microbolometers
- Voltage variation → thermopiles
- Capacitance variation → pyroelectric detectors

For instance, Figure 2.1 illustrates a microbolometer.

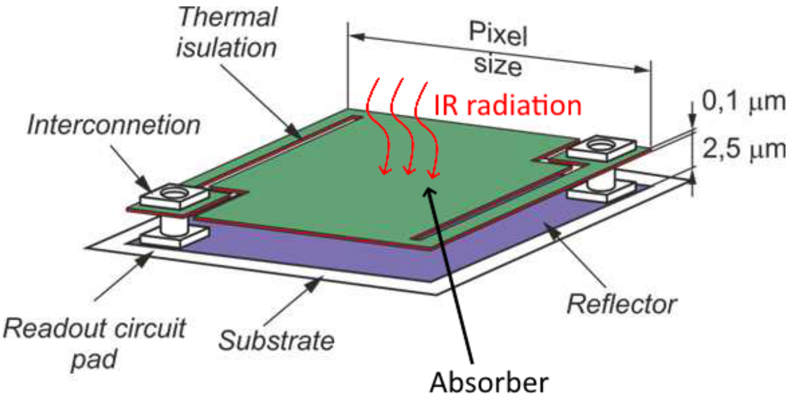


Figure 2.1: Scheme of a microbolometer. The incident IR radiation heats the absorber surface, producing a variation of electric resistance recorded and processed by the read-out circuit. The absorber is thermally isolated from its surroundings.

Thermal detectors are not sensitive to the incoming radiation's wavelength but only to the heat it produces. Therefore, they are not intrinsically able to select the desired bandwidth and are limited only by the optical properties of the FPA material. Their spectral characteristics can be modified using a bandpass filter or adequately choosing the FPA material [17].

Thermal detectors rarely require cooling and can operate at room temperature, making their price, consumption and size very affordable. Their integration with ROIC is excellent since thermal detectors are usually fabricated using Si-compatible technologies [17].

Thermal detectors are characterized by their relatively low time response due to the heating process. This time response is described in the following.

2.1.2 Quantum / photon detectors

Quantum (or photon) detectors are mainly semiconductors whose photonic characteristics are selectively modified by suitable doping to achieve specific performance. Photon detectors are split into photovoltaic detectors (photodiodes) and photoconductors. In the first case, incident photons generate a voltage across a p–n junction in the semiconductor. In the second case, they change the semiconductor's electrical conductivity (resistivity) by creating free charge carriers. In short, an absorbed photon instantaneously excites an electron from a nonconducting state into a conducting state, causing a change in the electrical properties of the semiconductor material that can be measured by the ROIC [18]. The response of a photon detector depends on the number of incoming photons (photon flux) that have an energy $E_\phi > \hbar\nu$, where ν is the bandgap optical frequency. It means that, unlike thermal detectors, a quantum detector can be designed to be sensitive to a specific waveband (i.e. ν depends on the choice of semiconductor material, doping, bandgap engineering).

Photon detectors have extremely fast response times (100 ns to 1 µms) due to their quantum nature [15]. However, they must operate at cryogenic temperatures to reduce the quantum noises making their cost, consumption and size much higher than thermal detectors. Figure 2.2 displays a photon detector photography and scheme:

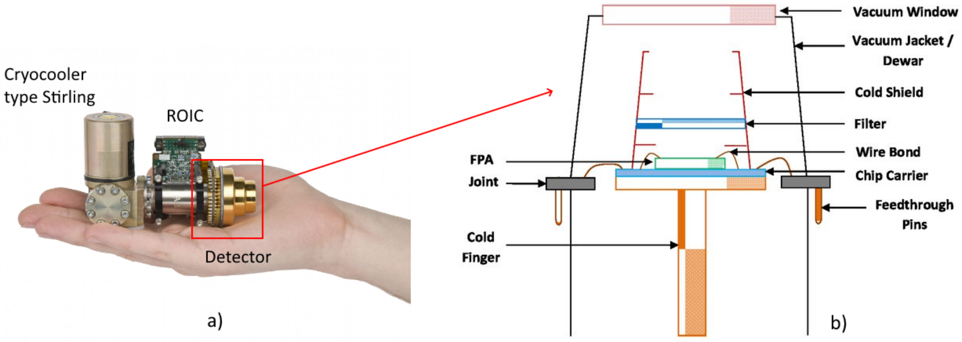


Figure 2.2: a) Picture and b) FPA scheme of a photon detector. The picture shows the InSb 'Pelican D 640' photodetector from SCD [19], and the scheme is taken from [20]. As shown in a), a cryocooler (Stirling type) and a read-out circuit (ROIC) are integrated into the detector. The FPA is thermally isolated from its surrounding by the dewar (vacuum) and only accepts light radiation in the selected waveband, defined by the filter. The cold shield decreases the amount of parasitic light.

2.1.3 Read-out circuit (ROIC) architectures

There are two types of Silicon ROIC: Charged coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS). In both architectures, the photon-electron conversion is performed at each pixel and electrons are stored in potential wells during the exposure time. A pixel is saturated or overexposed when the potential well capacity reaches its maximum. The optimal exposure time, maximising the signal-to-noise ratio, provides near-saturated pixels.

In the CCD architecture, the charges inside each pixel's well move downwards, row by row, towards a final row, as depicted in Figure 2.3:

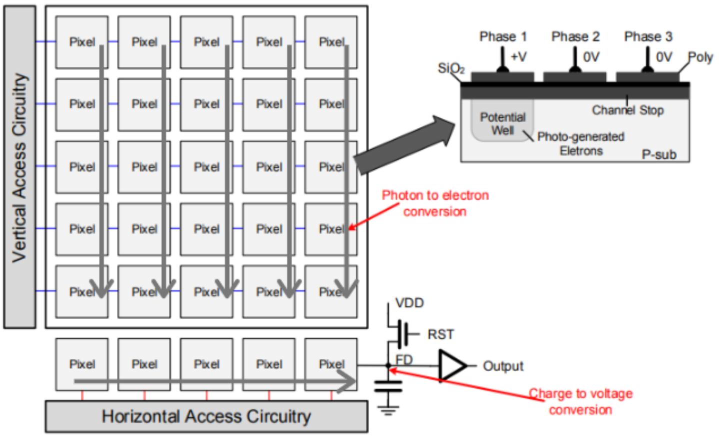


Figure 2.3: Operational principle of the CCD ROIC architecture. The charges inside each pixel's well move downwards, row by row, towards a final row known as the read-out register. Adapted from [21].

The charges in the final row (read-out register) move horizontally to a capacitor that integrates them and provides an output voltage or current corresponding to each pixel.

In the CMOS architecture, the charges-to-current (or voltage) conversion is done individually for each pixel. Each pixel possesses an "in-pixel" capacitor that integrates the charges contained in the potential well. In addition, a reset transistor cleans each pixel after integration. The pixels are all connected to a read-out circuit that processes the image.

As a summary, Table 2.1 provides a qualitative comparison of CMOS and CCD architectures:

Parameter	CCD	CMOS
Responsivity	-	+
Uniformity of the response	+	-
Dynamic range	+	-
Fill factor	+	-
Read-out noise	+	-
Speed (read-out)	-	+
Complexity of construction	+	-
Power consumption	-	+
Radiation hardness	-	+

Table 2.1: Comparison of the main properties of CCD and CMOS ROIC architectures. Adapted from [22].

CMOS detectors are more robust than CCDs, with higher responsivity and speed due to their in-pixel integration, but at the cost of increased complexity and read-out noises. Over the last two decades, CMOS ROICs have been considerably improved and used for space applications due to their low consumption and size [22].

2.1.4 Response time

The response time is defined as the time needed to reach 63% of the maximum response value after illumination by a Heaviside IR radiation step. Inversely, the response time also characterizes the time needed to reach 37% of the maximal response value after switching off the illumination [17]. For photodetectors, the response time is expressed in [µs] and represented as a step function.

For thermal detectors, the response time is known as the thermal time constant τ_{th} [s], expressed as [15, 16]:

$$\tau_{th} = \frac{C_{th}}{g_{th}} \tag{2.1}$$

where C_{th} is the thermal capacity of the detector [J/°K] and g_{th} is its thermal conductance [W/°K]. Assuming a typical capacity of 250-300 nJ/K and a thermal conductance of 25 nW/K, $\tau_{th} = 10$ ms. This typical response time is ten to a hundred longer than for photodetectors and comparable to the spacecraft integration time (see Equation 3.12), which can lead to a blurred image. Thereby, active imaging strategies such as the time delay and integration (TDI) technique [21] are often required for thermal detectors. Note that 5 τ_{th} are necessary to reach 99% of the output signal. Finally, the thermal time constant is a fixed value and cannot be reduced even by cooling down the detector.

The read-out time includes:

- the interval in which the charge is trapped into the potential wells

- the time needed to transfer the charge towards the capacitor (CCD)
- the time needed to convert the charge into a voltage/current and to register it

The read-out time can be longer than the exposure time since it starts as soon as photons are collected. Its duration depends on the range of frequency selected for the ROIC operation and the level of noise accepted. In particular, the reading-out of the charge is a critical process in terms of noise generation. ITR (integrate then read) mode is "slow" but decreases the noise level compared to the IWR (integrate while reading) mode. Each read-out is followed by a time interval that cleans the remaining charges. This process is known as idling or clearing and is carried on by the reset transistor.

Figure 2.4 summarize the various steps and time interval required to take an image, considering the ITR mode:

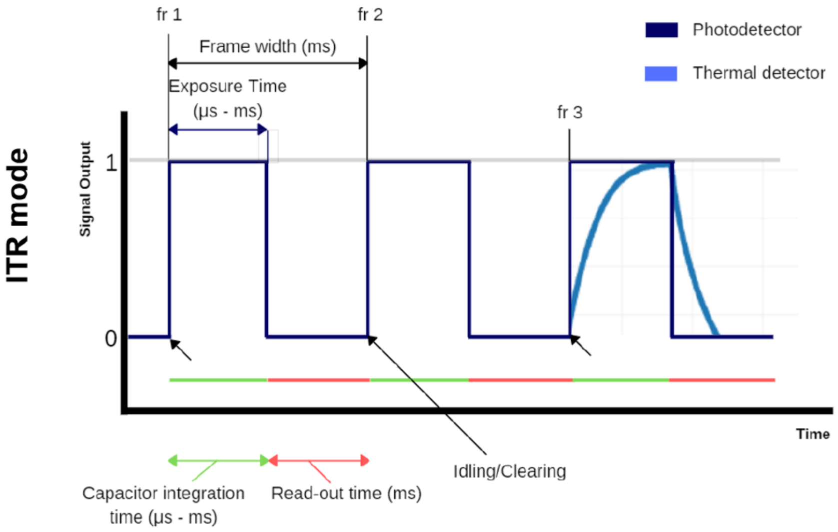


Figure 2.4: Characteristic time constants for the FPA and read-out circuits of IR detectors. The read-out mode is ITR (integrate then read). Original source [22]

The detector's frame rate (Hz) is defined as the maximal number of observations possibly taken in one second. Slower and faster frame rate modes are often available, resulting in a trade-off between the exposure time and the response time [22]. For instance, a thermal detector with an 8 ms response time can operate at full resolution with a 30 Hz frame rate ($5 * 8 = 24$ ms $< 1/30s$). In the 60 Hz mode, the exposure time finishes before the detector output signal has reached 95% of its final value [22].

2.2 Principal noises

2.2.1 Shot noise

The shot noise only occurs for photodetectors since it is related to photon counting. It defines the noise coming from the random fluctuations of a stream of electrons as a result of the discrete nature of the current carriers. Thereby, a continuous current is constituted by short pulses and the mean-square current fluctuation averaged over an integration time is called shot noise $I_{n,shot}$ [A]. It is generally expressed as [23]:

$$I^2_{n,shot} = 2\Gamma^2 q^2 (G_{opt} + G_{th} + R)\Delta f \tag{2.2}$$

where q is the elementary electron charge [C], Δf is the frequency bandwidth at which the detector circuits operate [Hz], and Γ is the photoelectric gain (the ratio between the number of electrons flowing through the electronic circuit and the number of absorbed photons). The mean current noise is split into an optical signal generation rate G_{opt} caused by the incident flux fluctuations, a thermal generation rate G_{th} (dark current noise) and a recombination rate R. For cooled photodetectors, it is assumed that the shot noise is dominated by G_{opt}: $G_{th} = 0$ and $R = 0$. The final expression of the shot noise is, therefore [23]:

$$I^2_{n,shot} = 2\Phi A \eta q^2 \Gamma^2 \Delta f \tag{2.3}$$

η is the quantum efficiency, defined between 0 and 1 as the number of electron-hole pairs generated per incident photon. A is the pixel size and Φ [photon/s] is the considered photon flux [23]:

- $\Phi = \Phi_S$ is the signal photon flux. It is the ideal case; the signal photon noise entirely determines the detector's noise.
- $\Phi = \Phi_B$ is the background photon flux. It is the most frequent case; the background signal dominates the detector's noise.

2.2.2 Johnson noise

The second prominent photodetector noise, the Johnson noise, results from the stochastic motion of charge carriers (to maintain thermal equilibrium) within a material of finite resistance. For uncooled thermal bolometers, this noise arises from the temperature fluctuations of the resistances receiving an incoming flux of photons [17]. It is expressed as a current [A] for photodetectors and as a voltage [V] for thermal detectors:

$$\begin{cases} I_J^2 = 4\dfrac{k_b T}{R}\Delta f \\ V_J^2 = 4 k_b T R \Delta f \end{cases} \tag{2.4}$$

where R is the semiconductor resistance [K/W], k_b is the Boltzmann constant [J/K], T is the detector temperature [°K] and Δf is the detector electronics frequency bandwidth [Hz]. The Johnson noise appears during the photon-electron conversion, and the read-out [22]. Similarly to the dark current noise, cooling the detector drastically reduces the thermal noise. The Johnson noise is a white noise dominant at high frequencies [17].

2.2.3 Flicker noise 1/f

The 1/f noise is related to imperfections in the material structure, and the electrical contacts [17]. The 1/f noise is the main limitation of the microbolometer's performance [23]. Due to its complexity, the flicker noise is commonly described with an empirical relation:

$$\begin{cases} I_{1/f}^2 = K\dfrac{i^\alpha}{f^\beta}\Delta f \\ V_{1/f}^2 = K\dfrac{v^\alpha}{f^\beta}\Delta f \end{cases} \tag{2.5}$$

where K is a proportionality factor, known as the 1/f parameter. i [A] and v [V] are respectively the average current flow and voltage through the detector. Δf is the frequency bandwidth range of the detector circuits [Hz], α is a constant of value $\simeq 2$, and β is also a constant that ranges from [0.5 to 1.5] with a general value of 1. f is the frequency of oscillation of the fluctuations producing the flicker noise [22].

2.2.4 Temperature fluctuation noise

Only present for thermal detectors, the "temperature noise" arises from thermal fluctuations in the detector caused by heat conductance variations between the detector and its surrounding substrate. This noise is expressed using the variance in temperature $\overline{\Delta \mathrm{T}}$ as [23]:

$$V_{th}^2 = \xi^2 \overline{\Delta \mathrm{T}}^2 = \xi^2 \frac{4k_b T^2 \Delta f}{1+\omega^2\tau_{th}^2} R_{th} \tag{2.6}$$

R_{th}, the semiconductor resistance [K/W], is the principal heat loss mechanism and the key design parameter that affects the temperature fluctuation noise. τ_{th} [s] is the thermal time constant (Equation 2.1), ω is the voltage responsivity [Hz] and ξ [V/K] is a coefficient that reflects "how well" the temperature variations translate into an electrical output voltage [23].

2.2.5 Background temperature noise

Only present for thermal detectors, the background noise results from radiative heat exchange between the detector at temperature T_d and the surrounding environment at temperature T_b. It is the ultimate limit of a thermal detector's performance capability and is given for a 2π field of view (FOV) [23]:

$$V_b^2 = \frac{8k_b\epsilon\sigma A(T_d^5 + T_b^5)}{1+\omega^2\tau_{th}^2}\Delta f \xi^2 R_{th}^2 \tag{2.7}$$

where σ is the Stefan–Boltzmann constant $\sigma = 5.67.10^{-8}$ [W.m^{-2}.K^{-4}], ϵ is emissivity of the detector and A [m^2] the detector sensitive area.

2.3 Figures of merit

IR detectors convert the received IR flux to an electrical signal as tension or current [15]. Their performance is measured using various figures of merit. The ideal detector would have a maximal responsivity (i.e. sensitivity), a perfect spectral responsivity for the considered waveband (photon detector), an instantaneous response time and a maximal detectivity.

2.3.1 Responsivity

An IR detector's responsivity is the ratio of the root mean square (RMS) electrical output signal (voltage V_s for a thermal detector or current I_s for a photon detector) to the RMS value of the input radiation power Φ_w [W] [23]. Voltage $\mathcal{R}_v$ and current $\mathcal{R}_I$ responsivities are respectively given in [V/W] and [A/W]:

$$\begin{cases} \text{Thermal: } \mathcal{R}_v(T,f) = \dfrac{V_s}{\Phi_w} = \dfrac{\xi \Delta T}{\Phi_w} = \xi \dfrac{\epsilon R_{th}}{\sqrt{1+\omega^2 \tau_{th}^2}} \\ \text{Photon: } \mathcal{R}_I(\lambda) = \dfrac{I_s(\lambda)}{\Phi_w} = \eta q \Gamma \dfrac{\lambda}{\hbar c} \end{cases} \tag{2.8}$$

Each parameter has been described in Section 2.2. The ratio $\lambda/\hbar c$ is used to convert the incident radiant power Φ_w [W] into an incident photon flux [photon/s], with $\hbar = 6.626.10^{-34}$ [J.s/photon] and $c = 3.10^{14}$ [µm/s].

Although responsivity is a valuable measurement to predict the signal level for a given irradiance, it does not indicate the minimum detectable radiant flux. In other words, it does not consider the amount of noise that ultimately determines the signal-to-noise ratio (SNR).

2.3.2 Noise equivalent power (NEP) and detectivity (D)

The random fluctuations in a detector's output limit its responsivity to a minimum detectable power, known as the Noise Equivalent Power (NEP), expressed in Watts. The NEP (i.e. the signal level that produces a unity SNR) measures the amount of radiative power that produces an RMS signal V_s or I_s equal to the noise V_N or I_N generated by the detector when it is shuttered from a blackbody source.

$$NEP = \frac{V_N}{\mathcal{R}_v} = \frac{I_N}{\mathcal{R}_I} \tag{2.9}$$

The NEP must be specified for a particular source temperature T, modulation frequency f, system bandwidth BW, and detector area (or pixel size) A [16]. The detectivity D is defined as:

$$D = \frac{1}{NEP} \tag{2.10}$$

2.3.3 Signal-to-noise ratio (SNR) and specific detectivity (D*)

The signal-to-noise ratio (SNR) is defined as [23]:

$$SNR = \frac{V_s}{V_N} = \frac{I_s}{I_N} = \Phi_w D \tag{2.11}$$

where Φ_w is the incident radiant power [W] and I_N / V_N are the current / voltage noises defined in Section 2.2.

Since the detectivity D greatly varies depending on the temperature, detector area, bandwidth and frequency, it is rarely used for comparisons. The detectivity is often normalized to a bandwidth of 1 Hz and detector area of 1 cm^2. The normalized detectivity is called the specific detectivity D^* and is interpreted as the SNR for 1 W of incident radiant power on a 1 cm^2 detector at a noise equivalent bandwidth of 1 Hz. D^* is expressed in Jones [cm.$H_z^{1/2}$.W] [18, 23]:

$$D^* = \frac{\sqrt{A\Delta f}}{NEP} = \frac{\sqrt{A\Delta f}}{\Phi_w} SNR \tag{2.12}$$

where Φ_w is the incident radiant power [W], A is the detector area (or pixel size) [cm^2], and Δf is the frequency bandwidth [Hz], at which the circuits of the detector operate.

For an ideal photon detector with a detectivity independent of any other parameter (not background limited), D^* linearly evolves with the wavelength and is often provided at the cutoff wavelength λ_p, where the responsivity (or SNR) is the highest [18, 24]:

$$\begin{cases} D^*_\lambda = D^*_{peak} \dfrac{\lambda}{\lambda_p} \\ D^*(\lambda > \lambda_p) = 0 \end{cases} \tag{2.13}$$

2.3.4 Noise Equivalent Difference Temperature (NEDT)

An object must produce a sufficient object-to-background apparent temperature difference to be detected and subsequently identified. The NEDT characterizes the thermal sensitivity of an IR system. It is the minimal detectable temperature difference (between the target and its background) generating a unity SNR, called "thermal resolution". The NEDT is related to the SNR by [18, 24]:

$$NEDT \equiv \frac{\Delta T}{V_s/V_N} = \frac{T_O - T_B}{SNR} \tag{2.14}$$

The minimum detectable temperature difference between the target (T_t [°K]) and the background (T_B [°K]) is obtained when $\Delta T = T_t - T_B = NEDT$, leading to a SNR of 1.

2.4 Dual-band IR detectors state of the art

Due to the bulkiness of IR optical systems and the number of IR applications requiring at least dual-band imaging, there is an increasing number of emerging multicolour IR detectors. These detectors have built-in spectral selection capabilities and can produce multicolour images without optical splitting devices (beamsplitter, filter wheel, filter array), allowing for compact IR dual-band optical systems. This section, based on [22, 23], provides a short review of the state-of-the-art multicolour IR detectors that are commercially available or in development, giving their advantages and drawbacks.

Third-generation IR detectors emerged in 1990, bringing enhanced capabilities such as a higher number of pixels, higher frame rates, better thermal resolution, and multicolour functionality [23].

2.4.1 HgCdTe (Mercury Cadmium Telluride or MCT) dual-band FPAs

HgCdTe detectors combine two photodiodes p-n junctions, separated by an interface with different energy gaps, allowing for spectral selectivity. The unit cell consists of several collocated detectors (layers), each sensitive to a different spectral band as depicted in Figure 2.5:

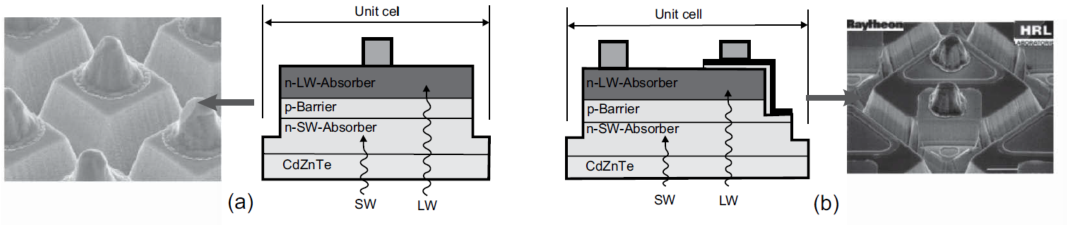


Figure 2.5: Two architectures for an n-p-n back-illuminated dual-band detector pixel. The short-wave IR flux is absorbed in the first layer (n), while the long-wave flux is transmitted through the next layer. Thin barriers (p) separate the absorbing bands. a) Bias-selectable structure (Raytheon) and b) Simultaneous design (Raytheon). Original source [23].

Back illumination refers to placing the wiring arrangement behind the sensitive part to avoid reflection losses. The sequential mode detector has a single indium bump per unit cell that permits sequential bias selectivity of the spectral bands, using the polarity of the bias voltage. Switching times within the detector can be relatively short, on the order of microseconds, so detection of slowly changing targets or images can be achieved by switching rapidly between the MW and LW modes [23]. The crosstalks between the two bands are a significant drawback of this simpler architecture.

Simultaneous dual-band detector architectures are often required for high-speed target imaging. The response of the LW photodiode is maximum when the response of the MW photodiode is minimum and vice versa. The MW photodiode acts as a bandpass filter for the LW photodiode, considerably reducing the spectral crosstalk. The simultaneous architecture needs an additional readout circuit in each unit cell allowing for independent control over the two photodiodes. Note that the Silicon ROIC must be custom-designed for each application since the flux levels can be very different in the two bands [23].

Based on the specification tables found in [22, 23] (Leonardo Condor II [25], Sofradir, Selex MWIR/LWIR detectors), HgCdTe dual-band FPAs generally exhibit the following characteristics:

- Spectral range: MWIR (dedicated), LWIR (dedicated), dual-band
- 640 × 512 pixels with 24 µm pixel size
- NEDT (see Section 2.3): 20-30 mK
- Operating temperature: 80K
- Quantum efficiency: 70% without AR coating
- Charge capacity (dedicated): $\simeq 10^7$ electrons

2.4.2 Dual-band quantum well infrared photoconductors (QWIPs)

The quantum well structure is formed in semiconductors when a material, for example, the GaAs alloy, is introduced between two layers of a material with a wider bandgap, as is it could be the case of aluminium or indium [22]. QWIPs provide ideal two-colour IR FPAs because they only absorb IR radiation in a narrow spectral band. Thus a QWIP provides zero spectral crosstalk when two spectral bands are more than a few microns apart [23]. Dual-band QWIPs can be stacked vertically, similarly to HgCdTe dual-band detectors, as shown in Figure 2.6:

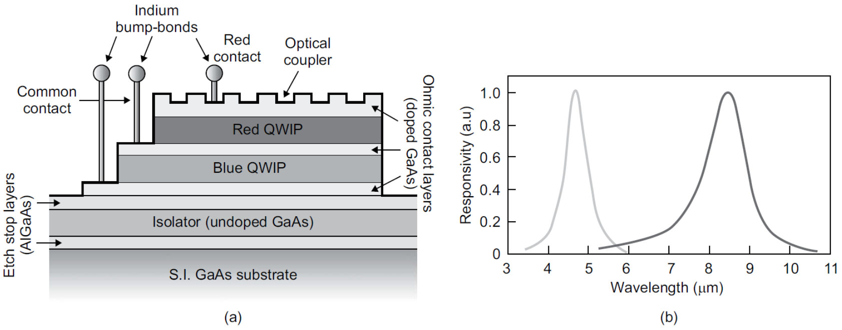


Figure 2.6: a) Schematic representation of a dual-band QWIP detector structure and (b) typical responsivity spectra at 77°K recorded simultaneously for two QWIPs at the same pixel. Original source [23].

QWIPs have been principally designed by the Jet Propulsion Laboratory (JPL), QmagiQ, the Army Research Laboratory, Goddard, Thales and AIM [23].

According to [26], the operating temperature of QWIP detectors is 40 K. It can go up to 100 K with a reduced signal-to-noise ratio (SNR), caused by the predominance of thermally stimulated carriers over optically produced carriers. Multicolor QWIPs share the advantages and disadvantages of QWIP single unit detectors, as the low quantum efficiency (<10%) or the high dark current in comparison with HgCdTe FPAs. QWIPs are intrinsically compatible with multi-band imaging and are much simpler to manufacture than HgCdTe FPAs. Finally, the NETD is lower than 35 mK, similarly to typical photodetectors. In conclusion, QWIPs are promising dual-band alternatives when low SNR is not an issue.

2.4.3 Type-II InAs/GaInSb superlattice (T2SL) detectors

The concept of superlattice refers to a structure made of stacked layers of different materials. In the last decade, type-II InAs/GaInSb superlattices (T2SLS) have emerged as the third candidate for third-generation IR detectors. One advantage of the T2SL is its reduced pixel vertical structure (5 µm) compared to HgCdTe pixels (15 µm), allowing higher photon transit time and, therefore, better performance at higher operating temperature [26]. T2SL performance is excellent in LWIR, but the German company AIM has made dual-band MWIR/MWIR detectors. The typical pixel pitch is 40 µm, and the average operating temperature can go up to 80 K [22].

2.4.4 Multicolor microbolometers

Microbolometers cannot separate multiple bandwidths but can be designed to operate on a wide bandwidth (typically from 3 to 14 µm). The use of spatially distributed band-pass filters on the FPA is necessary to make a multicolour microbolometer. Filtering often uses selective transmission, interferences, liquid crystal and birefringence. The filtering method needs to be chosen adequately, in line with the imaging strategy (snapshot or scan) [22]. In the case of snapshot imaging, dual-band Fabry-Perot filters are widely used for narrow or large wavebands. For example, a four-band LWIR microbolometer from the French manufacturer ULIS has been designed for the MISTIGRI (Micro Satellite for Thermal Ground Surface Imaging) satellite [27]. This device is displayed in Figure 2.7:

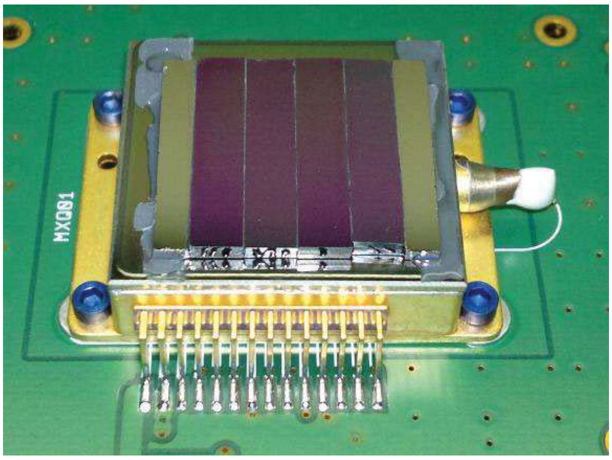
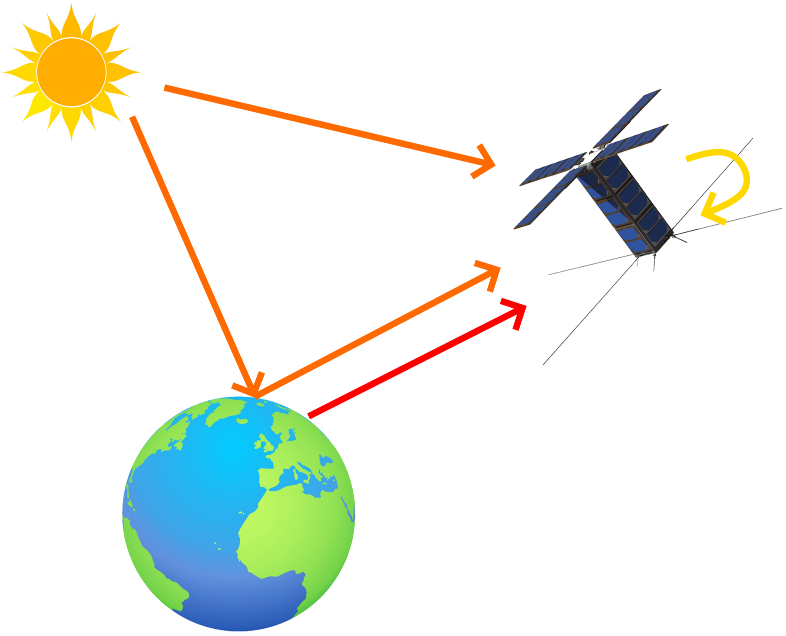


Figure 2.7: Four bands LWIR microbolometer from the French manufacturer ULIS, used in the MISTIGRI satellite. The filter is manufactured by assembling four linear sliced filter elements, so-called strips, each corresponding to one spectral band. Original source [27].



Chapter 3

Radiometry

Earth remote sensing requires the combination of high-level science requirements (Chapter 1), an optical assembly, and a detector (Chapter 2), providing valuable images. Based on the physics of radiometry, this chapter develops the many equations that define the optical system and ensure the feasibility of the selected applications. We perform a radiometric study in the bandwidth of interest and low Earth orbit for the challenging crop monitoring application, considering accurate ground emission and atmospheric transmission models. Through the combination of detection figures of merit, paraxial optics, science requirements, and pixel size trade-off, we can fully constrain the fundamental optical parameters (F-number, aperture, pixel size). A sufficient SNR is obtained in LWIR and MWIR, respectively, considering the low-temperature crop monitoring application and the high-temperature wildfire/volcano monitoring applications. This comprehensive study demonstrates the theoretical feasibility of these applications by a dual-band optical system, yet to design.

Contents

3.1 Blackbody radiation

A blackbody absorbs all incident radiations. Conversely, according to Kirchhoff's law, a blackbody is also a perfect radiator. Its emitted energy is theoretically maximal for a given surface temperature T. Most thermal radiation sources can be assimilated to blackbodies (Sun, wildfires, volcanoes), making the blackbody radiation laws the main starting point of radiometric calculations [23]. Planck's law describes the spectral emitted radiance $L(\lambda, T)$ of a perfect blackbody as a function of its surface temperature T [°K] and of the emitted radiation wavelength λ [µm]:

$$\begin{cases} L_W(\lambda, T) = \dfrac{2\hbar c^2}{\lambda^5 \left(\exp\left[\dfrac{\hbar c}{\lambda k_b T} \right] - 1 \right)} \\ L_P(\lambda, T) = L_W.\lambda/\hbar c \end{cases} \tag{3.1}$$

where $\hbar = 6.62607004.10^{-34}$ J.s is the Planck's constant, $k_b = 1.38064852.10^{-23}$ J/K is the Boltzmann's constant and $c = 3.10^8$ m/s is the speed of light in vacuum. The product $\hbar c/\lambda$ is expressed in [J/photon] and allows to convert the "power" emitted radiance L_W [W.m^{-2}.sr^{-1}.μm^{-1}] into a "photon" emitted radiance L_P [Photon.s^{-1}.m^{-2}.sr^{-1}.μm^{-1}]. Physically, the emitted radiance is the flux ([W] or [Photon/s]) emitted by a blackbody at surface temperature T, per unit of surface, per solid angle [sr] and wavelength.

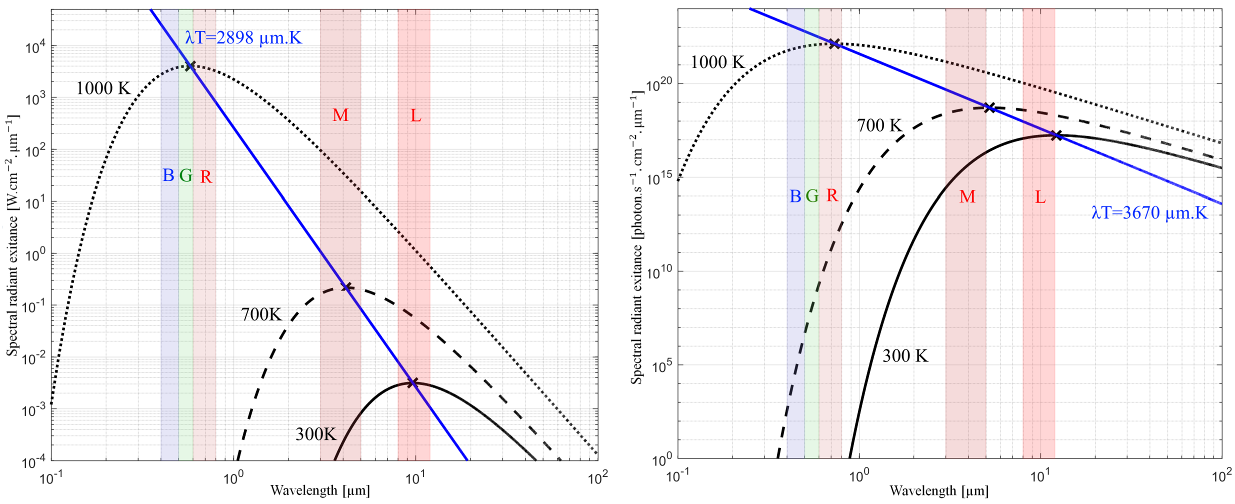
The spectral exitance (in Watt and photon/s) is defined as:

$$\begin{cases} M_W(\lambda, T) = \pi L_W \\ M_P(\lambda, T) = \pi L_P \end{cases} \tag{3.2}$$

$M_\lambda(T)$ does not depend on the solid angle. A blackbody at surface temperature T has its peak exitance at a certain wavelength (λ_W for M_W and λ_P for M_P). The thermal variation of the peak exitance wavelength is given by Wien's displacement law:

$$\begin{cases} \dfrac{\partial L_W}{\partial \lambda} = 0 \Leftrightarrow (x-5)e^x + 5 = 0 \Leftrightarrow \lambda_W = 2898/T \\ \dfrac{\partial L_P}{\partial \lambda} = 0 \Leftrightarrow (x-4)e^x + 4 = 0 \Leftrightarrow \lambda_P = 3670/T \\ x = \dfrac{\hbar c}{\lambda k_b T} \end{cases} \tag{3.3}$$

Figure 3.1 depicts Planck's law and Wien's displacement law:



(a) Power exitance units.

(b) Photon flux exitance units.

Figure 3.1: Graphical representation (log scale) of Planck's law (Equation 3.2) and Wien's displacement law (Equation 3.3) considering: a) Power (W) units and b) Photon flux (Photon/s) units. Five bandwidths are displayed: the visible B (blue), G (green) and R (red) bands, along with the infrared M (MWIR) and L (LWIR) bands.

From Figure 3.1, we deduce that a blackbody at 5800 °K, such as the Sun, will predominantly emit around 580 nm, which corresponds to the visible (green) light. The Earth's average temperature is around 300 °K, so it primarily emits light around 10 µm. Hotter temperature events, such as wildfires (700 °K), will therefore have their peak emission intensity in the MWIR waveband.

The total integrated exitance $M_{\Delta_\lambda}(T)$ inside a waveband $\Delta_\lambda = \lambda_b - \lambda_a$ depends on the considered blackbody temperature:

$$M_{\Delta_\lambda}(T) = \int_{\lambda_a}^{\lambda_b} \frac{2\hbar c^2}{\lambda^5 \left(\exp\left[\dfrac{\hbar c}{\lambda k_b T} \right] - 1 \right)} \mathrm{d}\lambda \tag{3.4}$$

It is shown in [23] that $M_{LWIR} > M_{MWIR}$ as long as the blackbody source has a temperature lower than 600 °K, meaning that the LWIR waveband is advantageous for low-temperature targets. When $\Delta_\lambda \to \infty$, we retrieve the Stefan Boltzmann law $M_\infty(T) = \sigma T^4$, with $\sigma = 5.67.10^{-12}$ [W.cm^{-2}.K^{-4}]. However, the integrated exitance is not the only criterion to consider.

The thermal variation of the spectral exitance $M_W(\lambda, T)$ can be approximated by [28]:

$$\frac{\partial M_W(\lambda, T)}{\partial T} = \frac{2\pi \hbar^2 c^3 \exp\left[\dfrac{\hbar c}{\lambda k_b T} \right]}{k_b \lambda^6 T^2 \left(\exp\left[\dfrac{\hbar c}{\lambda k_b T} \right] - 1 \right)^2} \simeq \frac{\hbar c}{\lambda k_b T^2} M_W(\lambda, T) \tag{3.5}$$

Equation 3.5 assumes that $\lambda << \hbar c / k_b T$, which is valid for $\lambda \simeq \lambda_{Wien}$ defined in System 3.3 [23]. It is important for a system operating in a certain waveband to compute the wavelength for which the source (target) exitance changes the most with the temperature.

This question is fundamental for the sensitivity of an IR system and is referred to as "thermal contrast" [23]. For a source of temperature 300 °K, the maximum contrast occurs at $\lambda = 8$ µm, which is not the wavelength corresponding to the peak exitance.

3.2 Spectral radiance at the top of atmosphere

This section describes the IR spectral radiance received by a low Earth orbit (LEO) observation satellite. This radiance is denoted "top of atmosphere" (TOA) radiance and computed in the middle wave (MWIR) and in the long wave (LWIR) wavebands. Classically, the MWIR waveband spans from 3 to 6 µm, while the LWIR waveband spans from 6 to 14 µm. A LEO spacecraft receives IR signals from four sources: The Sun direct and reflected light (albedo), the Earth and the spacecraft itself. These fluxes are depicted in Figure 3.2:

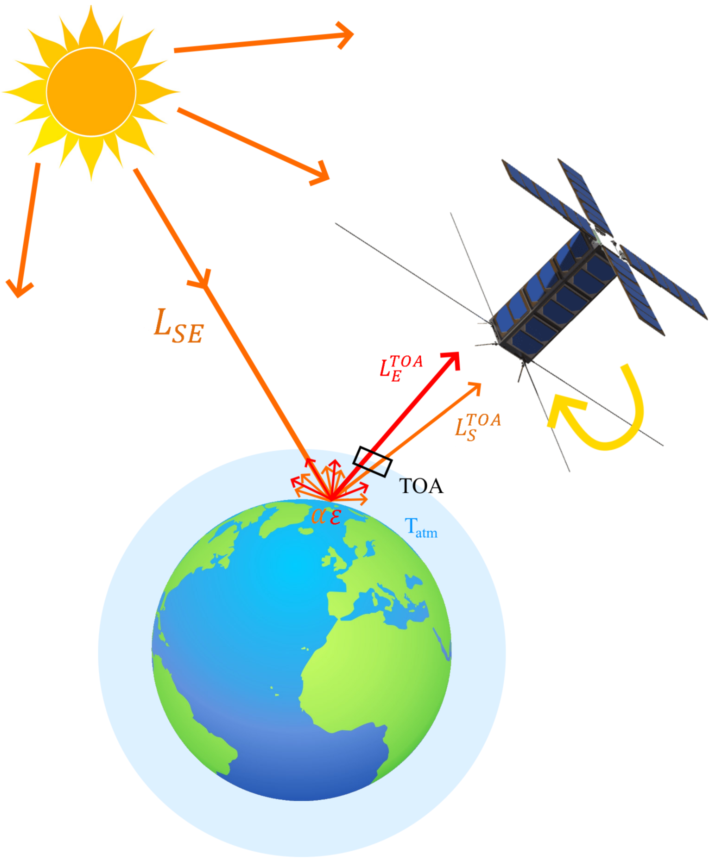


Figure 3.2: Considered IR fluxes for a LEO spacecraft. The orange arrows show the sun emitted spectral radiances: L_{SE} occurs between the Sun and the Earth, and L_S^{TOA} occurs at the top of the atmosphere (TOA) after reflection on the ground (albedo α). The TOA Earth thermal emission L_E^{TOA} depends on the spectral emissivity ϵ and is displayed in red. T_{atm} is the atmospheric transmission. Finally, the yellow arrow describes the spacecraft's internal heat dissipation flux. The flux coming directly from the Sun is assumed null since the spacecraft is turned towards the Earth.

The spacecraft emits IR light due to its internal dissipation. In this thesis, for simplicity, we consider that the optical system and the detector are thermally shielded from their surrondings. The lenses composing the optical system have a high infrared transmission and therefore do not get heated by absorption. We suppose that they do not generate parasitic infrared light, resulting in zero internal flux contribution. The spectral modeling of ϵ, α and T_{atm} is taken from [29]. In addition, the emissivity and albedo models include various ground materials (grass, ice, trees).

As previously said, the Sun is considered to be a perfect blackbody, with an average temperature of 5800 °K. Planck's law gives the emitted Sun radiance per solid angle and unit of power: $L_S(\lambda)$ [W.m^{-2}.sr^{-1}.µm^{-1}]. The Sun-Earth solid angle ω_{S-E} [sr] is expressed as:

$$\omega_{S-E} = \pi \left(\frac{R_S}{D_{SE}} \right)^2 = 6.7905.10^{-5} \tag{3.6}$$

using $R_S = 695508$ km (Sun radius) and $D_{SE} = 1.496.10^8$ km (Sun-Earth distance). The sun exitance $L_{SE}(\lambda)$ [W.m^{-2}.µm^{-1}] reaching 1 m^2 of Earth ground is expressed by:

$$L_{SE}(\lambda) = L_S.\omega_{S-E}.T_{atm} \tag{3.7}$$

For instance, the average solar irradiance on Earth is around 1400 W/m^2 at $\lambda = 0.5$ µm and only 0.5 W/m^2 at $\lambda = 8$ µm ($T_{atm} = 0.75$). A portion of this radiance is reflected in space. Considering that the ground acts as a Lambertian source (i.e. it reflects lights in a 2π sr hemisphere), the TOA solar albedo contribution $L_S^{TOA}(\lambda)$ [W.m^{-2}.sr^{-1}.µm^{-1}] is given by:

$$L_S^{TOA}(\lambda) = \frac{1}{2\pi} L_{SE}.\alpha.T_{atm} = \frac{1}{2\pi} L_S.\omega_{S-E}.\alpha.T_{atm}^2 \tag{3.8}$$

We use the same reasoning to compute the Earth's thermal emission. However, unlike the Sun, considered as a perfect blackbody, the Earth's surface is inhomogeneous, with variable temperatures and ground types. Thereby, the Earth is a "grey-body", with a thermal radiance L_E given by Planck's law and modulated by an emissivity function $\epsilon(\lambda)$. The TOA Earth IR radiance, $L_E^{TOA}(\lambda)$ [W.m^{-2}.sr^{-1}.µm^{-1}] is expressed as:

$$L_E^{TOA}(\lambda) = L_E.\epsilon.T_{atm} \tag{3.9}$$

L_E^{TOA} and L_S^{TOA} are displayed in Figure 3.3 for low temperature targets and in Figure 3.4 for high temperature targets. The considered ground surface is a mixing of "soil" and "grass" [29].

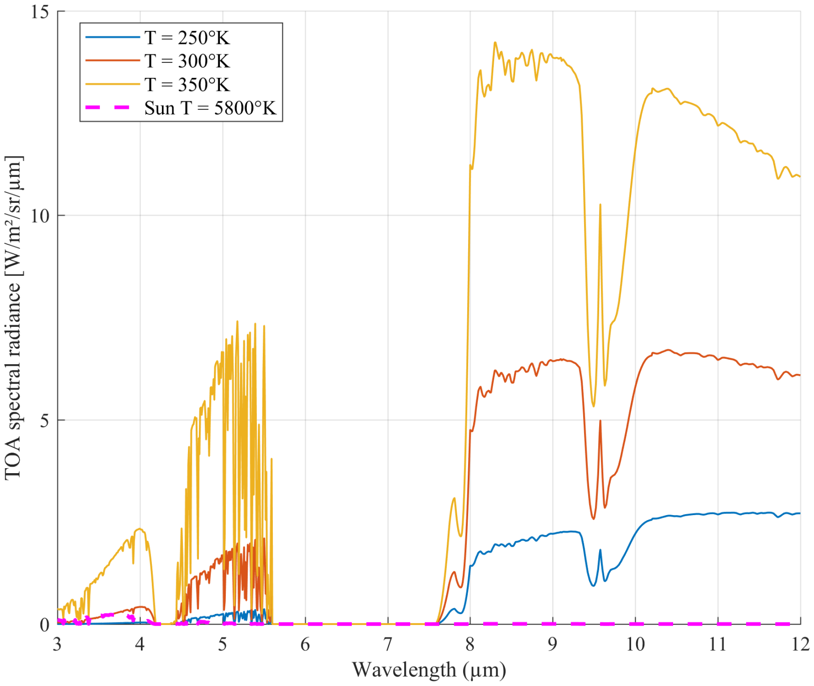


Figure 3.3: Earth TOA IR spectral radiance for various low temperature ground targets. The emissivity and albedo are deduced from [29], considering a mix of "soil" and "grass" scene. The Sun reflection contribution is displayed in magenta.

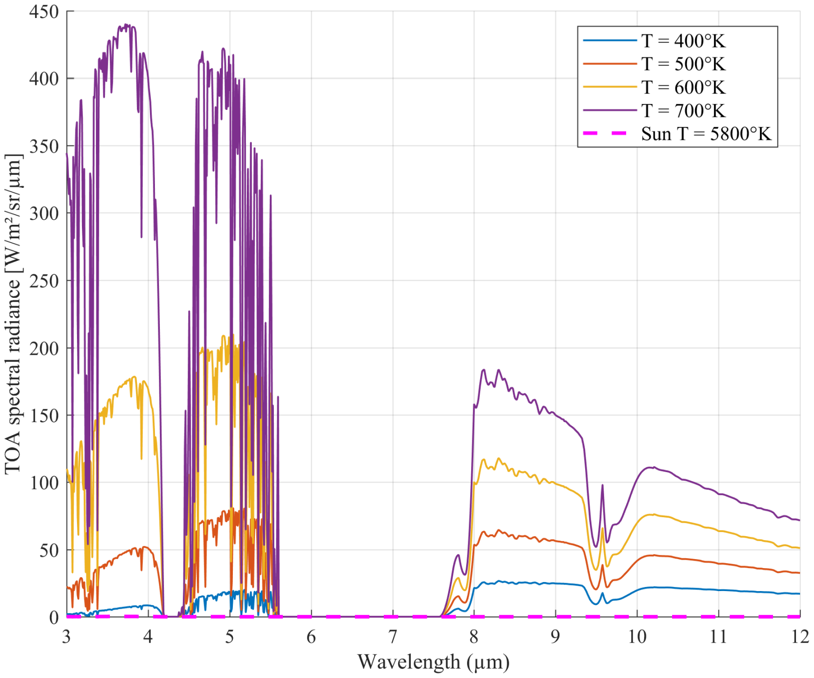


Figure 3.4: Earth TOA IR spectral radiance for various high temperature ground targets. The emissivity and albedo are deduced from [29], considering a mix of "soil" and "grass" scene. The Sun reflection contribution is displayed in magenta.

The solar component is negligible for wavelengths larger than 4.4 µm, or scene temperatures $T_s \geq 350$ °K. Two opaque atmospheric windows are present between 4.2 and 4.4 µm and between 5.5 and 8 µm. The first window is due to the absorption by CO_2 and N_2O. The second absorption window is caused by N_2O and CH_4. Consequently, this thesis defines two appropriate Earth IR imaging wavebands: the MWIR spans from 4.4 to 5.5 µm and the LWIR spans from 8 to 14 µm.

3.2.1 Advantages of LWIR/MWIR dual-band

According to [14], the 8–14 µm band is preferred for high-performance thermal imaging because of its higher sensitivity to ambient temperature objects (300 °K, see Equation 3.3) as well as its better transmission through mist and smoke. However, the 3–5 µm band may be more appropriate for hotter objects or if sensitivity is less critical than contrast.

The thermal contrast C $[K^{-1}]$ is an important parameter for IR imaging devices. It is given by [23]:

$$C = \frac{1}{M(\lambda,T)}\frac{\partial M(\lambda,T)}{\partial T} \simeq \frac{\hbar c}{\lambda k_b T^2} \tag{3.10}$$

A thermal image arises from temperature variations or differences in emissivity within a scene. When the temperatures of a target and its background are similar, detection becomes very difficult. Figure 3.5 displays the thermal contrast in both MWIR and LWIR wavebands for multiple scene temperatures:

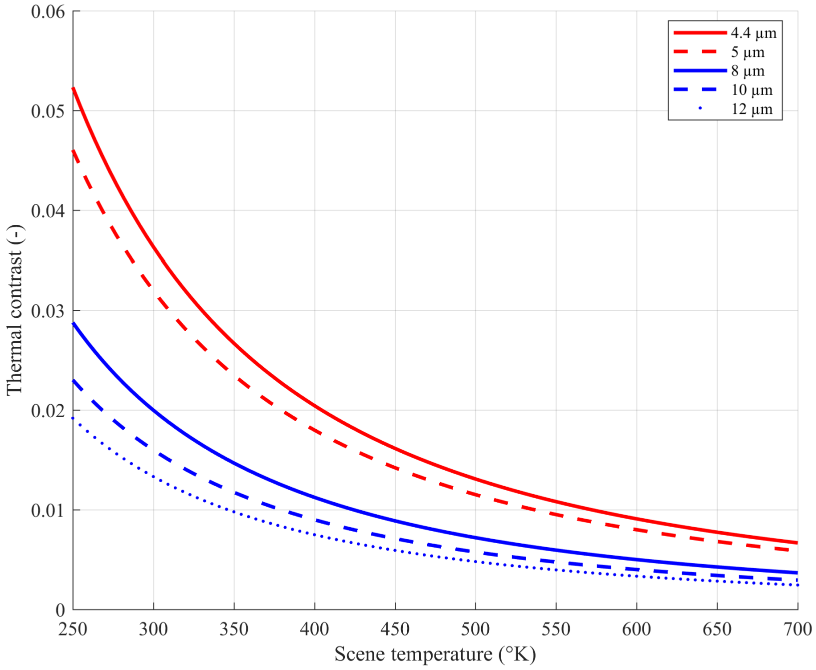


Figure 3.5: Thermal contrast in MWIR (red) and LWIR (blue) for various scene temperatures (Eq. 3.10).

For instance, the contrast in the MWIR band for a 300 °K target is 3.5% while it is only 1.6% in the LWIR band. The LWIR band has a high sensitivity for ambient temperature objects, but the MWIR band has better contrast. The other advantage of the MWIR band is the smaller aperture diameter required to obtain a particular resolution (ground sampling distance GSD) due to weaker diffraction effects. Some detectors may operate at higher temperatures (thermoelectric cooling) in MWIR than in LWIR, where cryogenic cooling is often required (about 77 °K for LWIR photon detectors).

In summary, MWIR and LWIR spectral bands differ substantially concerning the background flux, scene characteristics, thermal contrast and atmospheric transmission under diverse weather conditions. Factors favouring MWIR applications are higher contrast, superior clear-weather performance (favourable weather conditions in most countries of Asia and Africa), higher transmittivity in high humidity and higher resolution due to smaller diffraction blur spot. Factors favouring LWIR applications are better performance in fog and dust conditions, winter haze (typical weather conditions in West Europe, North USA, and Canada), higher immunity to atmospheric turbulence and reduced sensitivity to solar glints and fire flares. The possibility of achieving a higher signal-to-noise ratio (SNR) due to greater radiance levels in the LWIR spectral range is not persuasive since the background photon fluxes are higher to the same extent [14].

3.3 Parametric study

3.3.1 Radiant power and electrical signal

In this section, we express the radiant power $\Phi_{w,\lambda}$ and the electrical signal S detected by one pixel based on the main optical parameters of a dual-band camera. The optical system is modelled in the simple paraxial regime, where all parameters are geometrically related to each other, as illustrated in Figure 3.6:

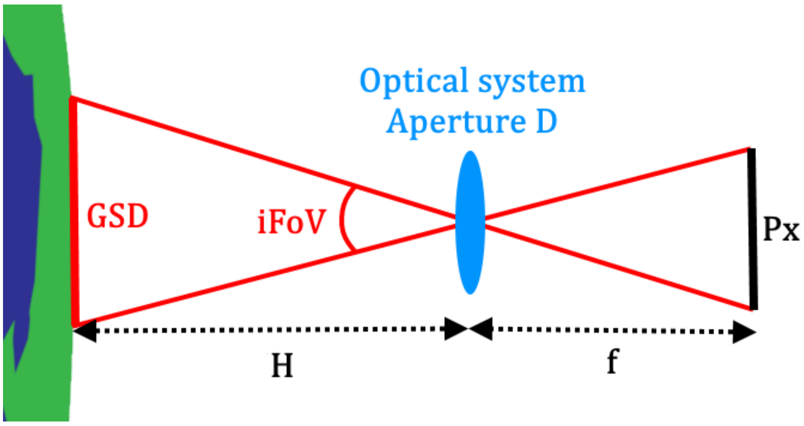


Figure 3.6: Geometric representation of an imaging system in the paraxial regime.

The fundamental paraxial relation derived from Figure 3.6 is:

$$2\tan\left(iFoV/2\right) \simeq \tan\left(iFoV\right) = \frac{GSD}{H} = \frac{P_x}{F/\sharp D} \tag{3.11}$$

where GSD is the ground sampling distance, H is the spacecraft altitude, D is the pupil diameter, and P_x is the pixel size. $iFoV$ represents the instantaneous field of view (FoV), i.e. the FoV seen by exactly one pixel. The spacecraft integration time can be calculated as well, using:

$$T_i = \frac{GSD}{V_{sat}} = \frac{GSD}{\dfrac{2\pi}{T_0} r - R_e \omega_e} \tag{3.12}$$

R_e is the Earth radius, $r = R_e + H$ is the orbit radius and $T_0 = 2\pi\sqrt{r^3/\mu_e}$ is the Earth rotation period. $\mu_e = 398600.4418$ km^3/s^2 is the Earth gravitational parameter and $\omega_e = 7.2921159.10^{-5}$ rad/s is the Earth angular velocity. Here we take the assumption of a passive spacecraft with constant speed, making a linear scan of the GSD. Assuming a LEO altitude, the spacecraft speed is $\simeq 7.1$ km/s. According to Equation 3.12, the integration time is proportional to the GSD, for instance, with a 100 m GSD, the integration time is $T_i \simeq 14$ ms.

The Earth TOA thermal radiance L_E^{TOA} [W.m^{-2}.sr^{-1}.µm^{-1}] has been calculated in Equation 3.9. Only a portion of this radiance effectively enters the optical system and reaches the detector. The concept of etendue conservation is used to compute the radiant power on the detector. It is based on the geometric representation depicted in Figure 3.7:

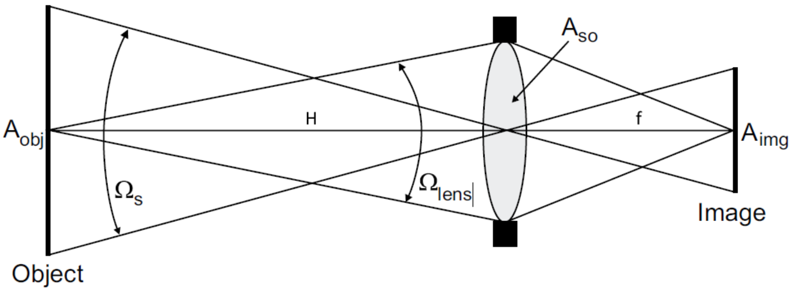


Figure 3.7: Radiant power collected by an optical system using the conservation of the etendue $E = A\Omega = cte$, where A is an area $[m^2]$ and Ω the related solid angle [sr].

The etendue characterizes how "spread out" the light is (area and solid angle) and is a constant inside an optical system. Based on Figure 3.7 scheme, the etendue can be written as:

$$E = A_{obj}\Omega_{lens} = A_{so}\Omega_{obj} = \frac{A_{so}A_{obj}}{H^2} = \frac{A_{so}A_{img}}{f^2} = \frac{\pi p_x^2}{4F/\sharp^2} \tag{3.13}$$

where A_{so} and A_{img} are respectively the collection area of the optics (for a circular aperture, $A_{so} = \pi D^2/4$) and the pixel area p_x^2. The radiant power $\Phi_{w,\lambda}$ [W.µm.m^{-1}] reaching a pixel is therefore expressed as [23]:

$$\Phi_{w,\lambda} = L_E^{TOA}.E.T_{opt} = \frac{p_x^2 T_{opt}}{4F/\sharp^2} M_\lambda(T) \tag{3.14}$$

T_{opt} represents the total transmission coefficient of the optical system, including optical interface transmission and diffraction losses. The atmosphere spectral transmission and the ground spectral emissivity are included in the radiant exitance M_λ.

Condidering a photodetector, this radiant power is converted into an electron flux by the photoelectric effect. The quantum efficiency η represents the ability of the pixel to convert an incoming photon signal into electrical charges. The integrated electron flux Φ_{e^-} [e$^-$.s^{-1}] reaching a detector pixel per second is expressed as:

$$\Phi_{e^-} = \int_\lambda \Phi_{w,\lambda} \frac{\lambda}{\hbar c} \eta \, d\lambda \tag{3.15}$$

During an exposure time t [ms], the total electrical signal S [e^-] generated and stored in each pixel is:

$$S = t.\Phi_{e^-} \tag{3.16}$$

If the total integration time $t = T_i$ is used, the electrical signal is directly related to the GSD.

3.3.2 NEDT and SNR

We provide the NEDT expression based on the optical system and detector parameters in the following. Here, the NEDT is defined and expressed in terms of voltage (thermal detector), but the calculation is identical using a current (photon detector). The integrated signal voltage V_s is expressed using Equations 2.8, 2.9 and 2.12:

$$V_s = \int\limits_{\lambda} \mathcal{R}_v(\lambda)\Phi_{w,\lambda}\,\mathrm{d}\lambda = \int\limits_{\lambda} \frac{V_N}{NEP}\Phi_{w,\lambda}\,\mathrm{d}\lambda = \frac{V_N}{\sqrt{A\Delta_f}}\int\limits_{\lambda} D^*\Phi_{w,\lambda}\,\mathrm{d}\lambda \tag{3.17}$$

The spectral radiant power $\Phi_{w,\lambda}$ [W/µm] is expressed in Equation 3.14. The specific detectivity D^* does not depend on the target temperature, leading to the expression of the NEDT [23, 24]:

$$NEDT \equiv V_N\frac{\partial T}{\partial V_s} = V_N\left(\frac{\partial V_s}{\partial T}\right)^{-1} = \frac{4F/\sharp^2}{T_{opt}}\sqrt{\frac{\Delta_f}{A}}\left[\int\limits_{\lambda}\frac{\partial M_\lambda}{\partial T}(T_B)D^*\,\mathrm{d}\lambda\right]^{-1} \tag{3.18}$$

In the ideal case of a photon detector with D^* independent of other parameters (the detector is not background limited), D^* can be simplified using Equation 2.13 [24]. In addition, Equation 3.5 can also be used to remove the temperature derivative:

$$\frac{\partial M_\lambda}{\partial T}(T_B)D^* \simeq \frac{\hbar c}{\lambda_p k_b T_B^2}M_\lambda D^*_{peak}(\lambda_p) \tag{3.19}$$

$k_b = 1.38064852.10^{-23}$ J/K is the Boltzmann's constant and T_B is the background temperature. λ_p denotes the peak detectivity wavelength, defined in Equation 2.13. The final NEDT expression is [24]:

$$NEDT \simeq \Lambda F/\sharp^2\sqrt{\frac{\Delta_f}{A}} \tag{3.20}$$

where the constant $\Lambda = \dfrac{4k_b\lambda_p T_B^2}{T_{opt}\hbar c D^*_{peak}}\left[\int\limits_{\lambda_1}^{\lambda_p} M_\lambda(T_B)\,\mathrm{d}\lambda\right]^{-1}$ depends only on the background temperature T_B. The T_B^2 dependency is confusing at first sight since the NEDT should decrease for higher background temperatures. In fact we have $T_B^2 << \int\limits_{\lambda_1}^{\lambda_p} M_\lambda(T_B)\,\mathrm{d}\lambda$, making the NEDT much smaller for high background temperatures.

The best thermal sensitivity corresponds to the lowest NEDT and, therefore, to the highest spectral exitance. It is obtained at the peak wavelength of the spectral exitance M_λ. The $\sqrt{\Delta_f}$ dependency is intuitive, since it is shown in Section 2.2 that the RMS noise is proportional to $\sqrt{\Delta_f}$. In addition, better NEDT results from lower $F/\sharp$ because more flux is captured by the detector, increasing the SNR. The $1/\sqrt{A}$ dependency is critical: The RMS noise increases as the square root of the pixel area. A trade-off must be made between thermal sensitivity (large pixels) and spatial resolution (small pixels).

For a scanning system, the frequency bandwidth Δf can be estimated using the dwelltime τ_d as [24]:

$$\Delta_f = \frac{\pi}{4\tau_d} = \frac{\text{FoV}_X.\text{FoV}_Y}{\text{FoV}_i^2}\frac{\pi F_R}{4N\eta_{sc}} \tag{3.21}$$

where FoV_X, FoV_Y and FoV_i denote the detector X and Y fields of view and the instantaneous pixel field of view. F_R is the frame rate, N is the number of pixels per scan line and $\eta_{sc} = 0.8^2$ is the scan efficiency.

In the following, we study the LWIR NEDT in a particular detection scenario. The background temperature is set to 300 °K, and we define $\lambda_1 = 8$ µm and $\lambda_p = 12$ µm. We take $D^*_{peak} = 10^{11}$ [cm.Hz$^{1/2}$.W^{-1}] as standard peak detectivity value. We arbitrarily set 500 scan lines, containing each $N = 600$ pixels, and a 30 Hz frame rate leading to $\Delta f = 10$ kHz. The lens transmission is 0.8 while the pixel encircled energy E_c is computed for any $(p_x, F/\sharp)$ couple using a Bessel function. We obtain a total transmission of $T_{opt} = 0.8.E_c$. Figure 3.8 displays the resulting LWIR NEDT map for various pixel sizes and F-numbers.

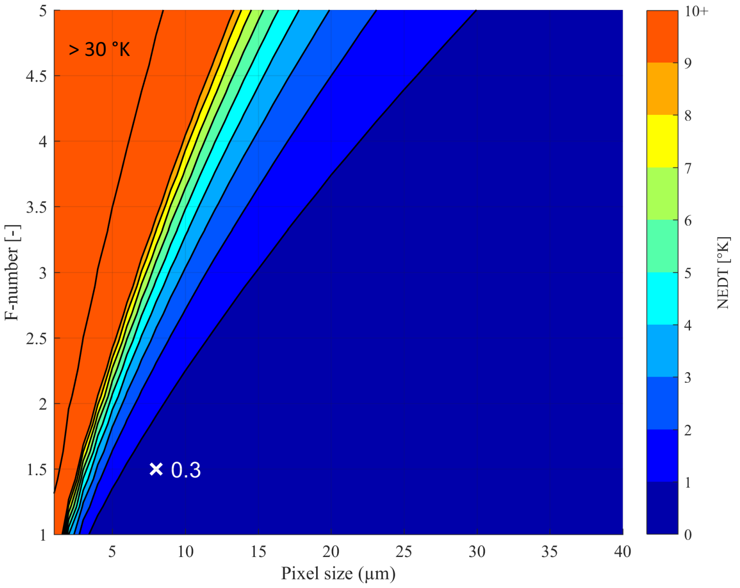


Figure 3.8: NEDT [°K] contour map computed using Equation 3.20 in LWIR. The white cross correponds to the NEDT value considering an F/1.5 system with a 8 µm pixel pitch, as explained in Section 3.4.1. The considered detector operates in a frequency bandwidth $\Delta_f = 10$ kHz and a peak detectivity $D^*_{peak} = 10^{11}$ [cm.Hz$^{1/2}$.W^{-1}]. The background temperature T_B is set to 300 °K.

The MWIR NEDT contour plot is displayed in Section 3.5 Figure 3.14 for completeness.

3.4 Radiometric budget

In this section, we use the previously derived quantities to evaluate the radiometric budget of a LEO dual-band mission. We study this budget by considering the challenging crop water assessment application (Section 1.2). The requirements defined in Table 1.1 stated a $\simeq 10$ m GSD, with 1 °K thermal resolution. Crop field mapping is ideally done in both MWIR and LWIR wavebands, to ensure c-high contrast, meteorological robustness.

3.4.1 Pixel size trade-off

The pixel size is a fundamental parameter involved in every radiometric relation. The optics (O), detector (det) and optical system (SO) modulation transfer functions (MTFs) are expressed as [30].

$$\begin{cases} MTF_O = \dfrac{2}{\pi}\left[\cos^{-1}\left(\dfrac{u}{u_O}\right) - \dfrac{u}{u_O}\sqrt{1-\left(\dfrac{u}{u_O}\right)^2}\right] \\ MTF_{det} = \mathrm{sinc}\left(\pi\dfrac{u}{u_d}\right) \\ MTF_{SO} = MTF_O \times MTF_{det} \end{cases} \tag{3.22}$$

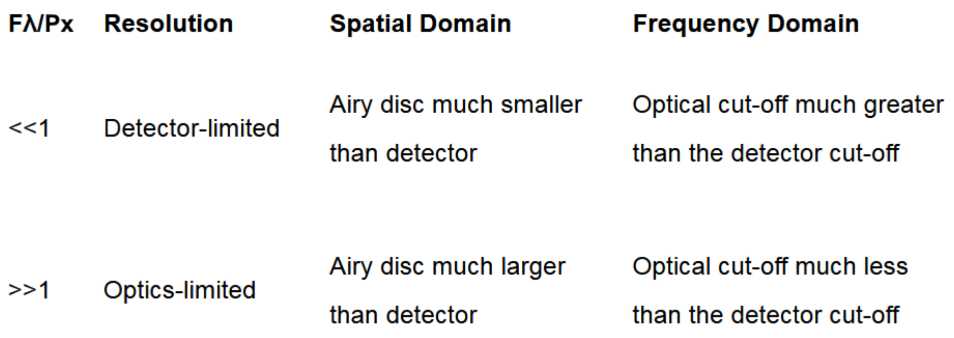
where u [cycles/mm] represents a spatial frequency. The optical and detector cutoff frequencies u_O and u_d are expressed as:

$$\begin{cases} u_O = \dfrac{1}{\lambda F} \\ u_d = \dfrac{1}{p_x} \end{cases} \tag{3.23}$$

where F is the optics F-number, and p_x is the pixel size. Note that u_d is defined at the first zero of MTF_{det} since this function is defined for $u \in [-\infty; \infty]$ [30]. Camera resolution depends upon the optical blur diameter and the detector size. In the spatial domain, λ/p_x is the ratio of the Airy disc diameter to the detector size. In the frequency domain, it is the ratio of the detector cutoff to the optics cutoff:

$$\frac{u_d}{u_O} = \frac{\lambda F}{p_x} \tag{3.24}$$

Table 3.1 depicts two important limiting cases:



FΛ/Px	**Resolution**	**Spatial Domain**	**Frequency Domain**
<<1	Detector-limited	Airy disc much smaller than detector	Optical cut-off much greater than the detector cut-off
>>1	Optics-limited	Airy disc much larger than detector	Optical cut-off much less than the detector cut-off

Table 3.1: Detector-limited VS optics-limited performance. The denomination "detector" represents a pixel in this case. Original source [31].

$\lambda F/p_x$ uniquely defines the shape of the optics/detector MTF combination. The sampling frequency u_S is equal to the detector cutoff frequency u_d, and the Nyquist frequency is therefore defined as $u_N = u_d/2$. The optics-limited and detector-limited condition are illustrated in Figure 3.9:

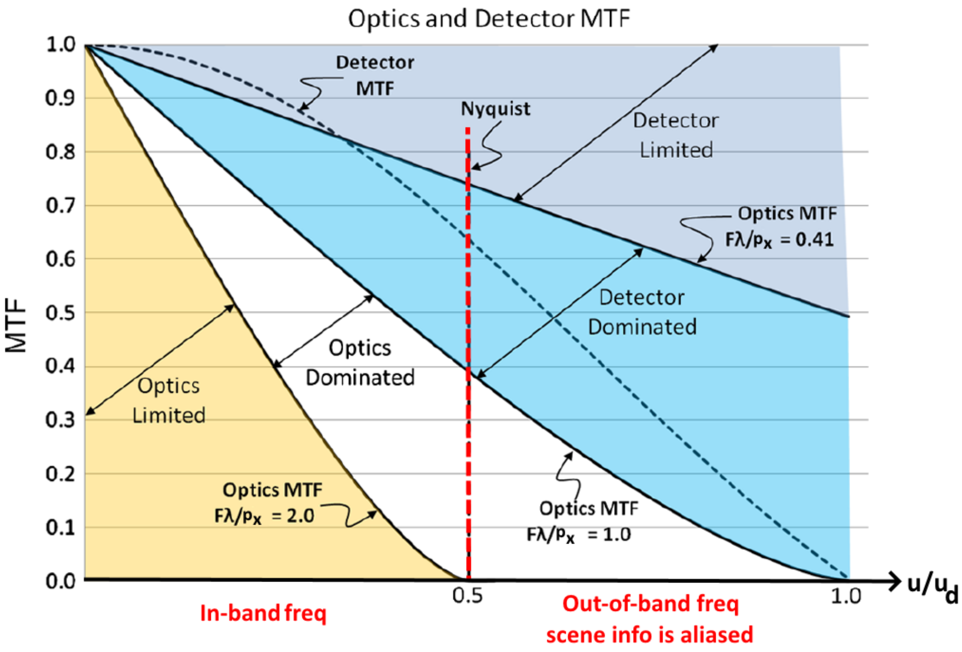
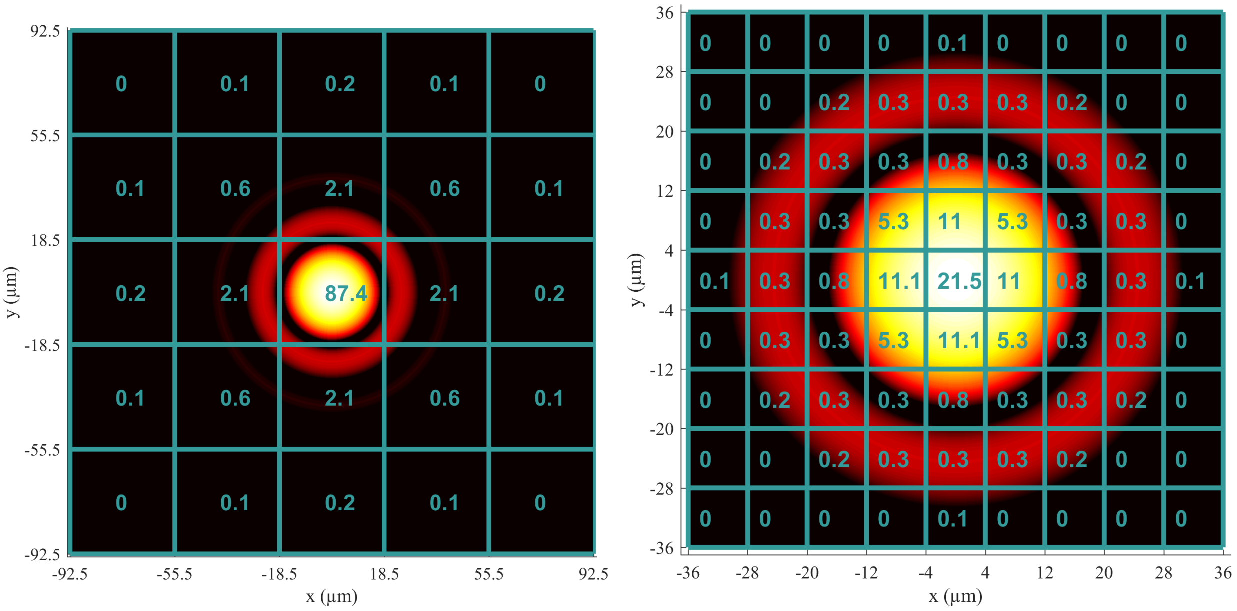


Figure 3.9: Optics MTF (MTF_O) and detector MTF (MTF_{det}) curves for various $F\lambda/p_x$ conditions describing the different regions within the design space. Spatial frequencies are normalized to the detector cutoff u_d. Original source [32]. Recall that $F\lambda/p_x = 2 \Leftrightarrow u_O = u_N$ and that 2.44=1/0.41. Each optical MTF curve assumes a diffraction-limited system with no aberrations.

Figure 3.9 presents a $F\lambda/p_x$ trade-off between sharpness and aliasing. High $F\lambda/p_x$ values reduce aliasing and in-band MTF_O, reducing contrast and edge sharpness in the final image. This effect is increased in reality since a reconstruction filter must be used to convert digital data into viewable analogic information. Reconstruction filters can be of many types [33] and create artefacts, reducing further the image quality. However, these artefacts frequencies (or spurious response) are mostly out-of-band for high $F\lambda/p_x$ [30]. Therefore, in the following, the pixel size is locked to the F-number by $F\lambda/p_x = 2$ in both MWIR and LWIR wavebands. It leads to an Airy spot sampled by 4.88 pixels, with a central pixel receiving 20% of the total energy, as displayed in Figure 3.10:



0	0.1	0.2	0.1	0
0.1	0.6	2.1	0.6	0.1
0.2	2.1	87.4	2.1	0.2
0.1	0.6	2.1	0.6	0.1
0	0.1	0.2	0.1	0

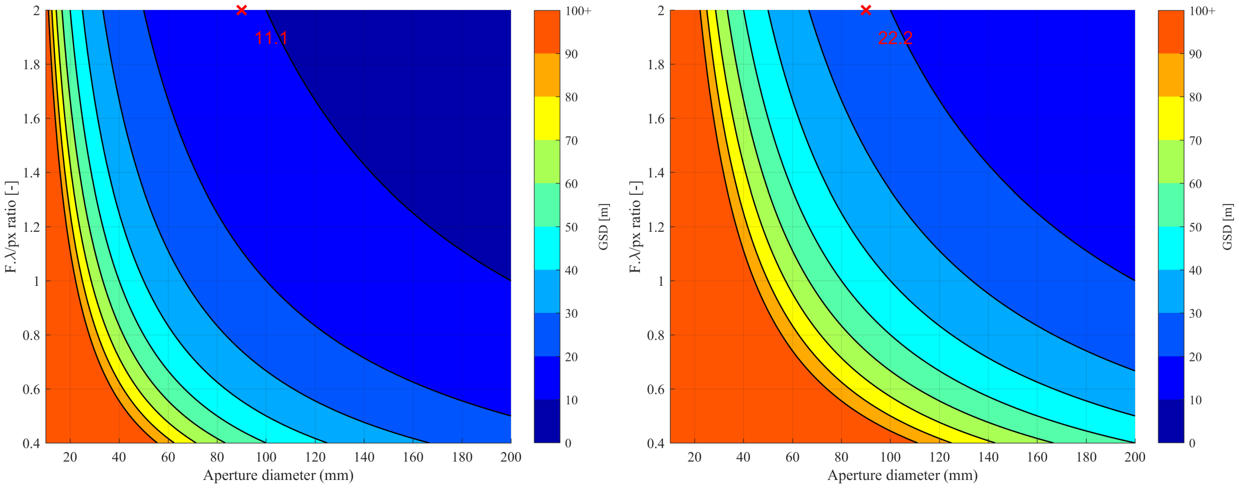
0	0	0	0	0.1	0	0	0	0
0	0	0.2	0.3	0.3	0.3	0.2	0	0
0	0.2	0.3	0.3	0.8	0.3	0.3	0.2	0
0	0.3	0.3	5.3	11	5.3	0.3	0.3	0
0.1	0.3	0.8	11.1	21.5	11	0.8	0.3	0.1
0	0.3	0.3	5.3	11.1	5.3	0.3	0.3	0
0	0.2	0.3	0.3	0.8	0.3	0.3	0.2	0
0	0	0.2	0.3	0.3	0.3	0.2	0	0
0	0	0	0	0.1	0	0	0	0

(a) Detector-limited: $F\lambda/p_x = 0.41$.

(b) Optics-limited: $F\lambda/p_x = 2$.

Figure 3.10: Logarithm of the normalized airy pattern (PSF): a) Detector-limited: $F\lambda/p_x = 0.41$ and b) optics-limited $F\lambda/p_x = 2$. For simplicity, the incoming radiation coincides with the pixel centre. The worst case wavelength $\lambda_0 = 10$ µm is used with $F/\sharp = 1.5$. The green rectangles delimit the square pixels. The energy integrated inside each pixel is displayed in green, showing the energy distribution in %. In the chosen case b), 4.88 pixels sample the Airy blur spot.

We consider a 5 µm MWIR wavelength and a 10 µm LWIR wavelength and we lock the ratio $\lambda_0 F/p_x = 2$. Figure 3.11 illustrates the GSD evolution in MWIR and LWIR:



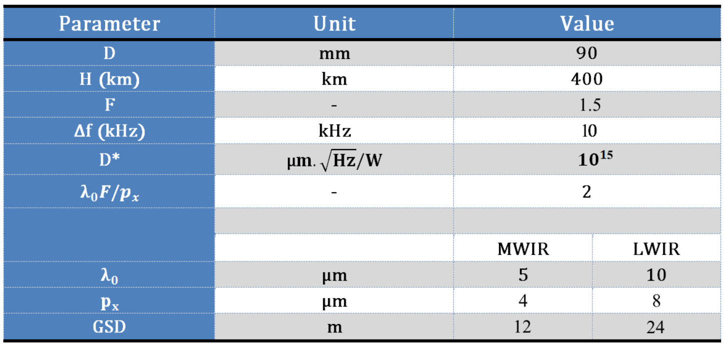
(a) MWIR waveband: $\lambda_0 = 5$ µm.

(b) LWIR waveband: $\lambda_0 = 10$ µm.

Figure 3.11: GSD evolution with the aperture diameter D (mm) and the $F\lambda/p_x$ ratio, represented in a contour map. We consider an F/1.5 system at a 400 km LEO altitude. Particular GSD vaules are indicated with a red cross, assuming a 90 mm diameter aperture, with $\lambda_0 F/p_x = 2$.

3.4.2 SNR calculation

Based on the previous parametric studies, we fix most of the system parameters as shown in Table 3.2:

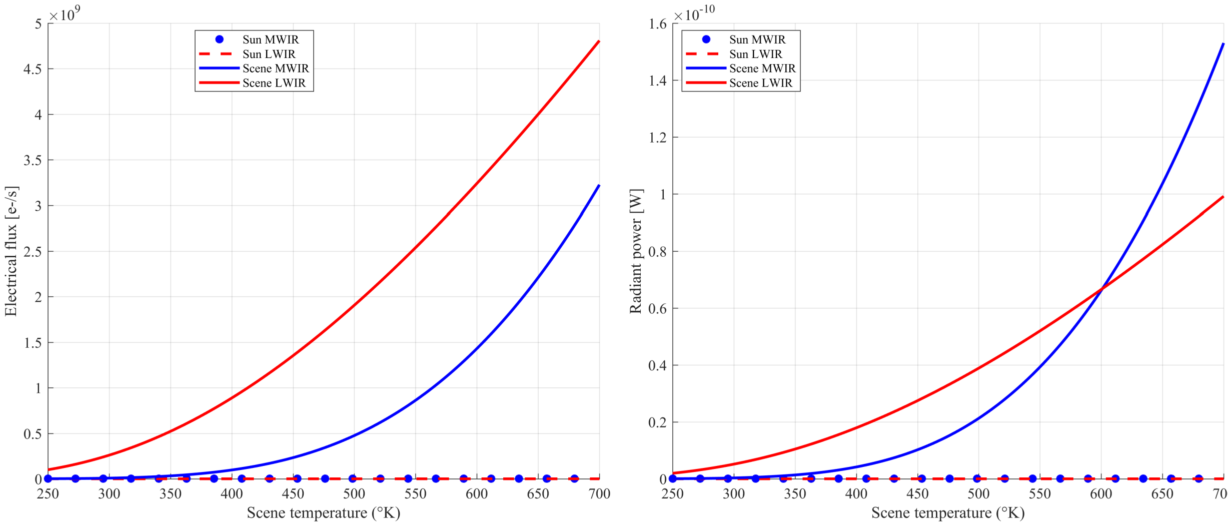


Parameter	Unit	Value	
D	mm	90	
H (km)	km	400	
F	-	1.5	
Δf (kHz)	kHz	10	
D*	$\mu m.\sqrt{Hz}/W$	10^{15}	
$\lambda_0 F/p_x$	-	2	
		MWIR	LWIR
λ_0	µm	5	10
p_x	µm	4	8
GSD	m	12	24

Table 3.2: Optical parameters fixed for the SNR calculation.

We define the total optical transmission T_{opt} as the average "glass transmission" times the pixel's "ensquared energy", given in Figure 3.10b): $T_{opt} = 0.8 \times 0.2 = 0.16$.

Figure 3.12 shows the electrical flux reaching each pixel for various scene temperatures. The electrical flux is integrated for MWIR and LWIR wavebands (in [e-/s] and [W]) and comes from the Earth emission and the Solar reflection.



(a) Electrical flux in $[e^-/s]$.

(b) Radiant power [W].

Figure 3.12: a) Electrical flux [e-/s] and b) radiant power [W] reaching each pixel, integrated in MWIR and LWIR wavebands, for various scene temperatures. The red and blue dotted lines represent the solar reflection contribution.

Figure 3.12 Assumes an homogeneous GSD, made of 70% cropland [29]. Finally, the SNR is illustrated in Figure 3.13 for the LWIR waveband, based on the NEDT (Figure 3.8) for various scene and background temperatures.

The MWIR SNR contour plot is displayed in Section 3.5 Figure 3.15 for completness. Considering the LWIR waveband and the crop water monitoring application, the resulting SNR is around 25 with a 0.3 °K thermal resolution (i.e. NEDT in Figure 3.8). The theoretical GSD is 24 m, providinga high resolution LWIR image, even if it exceeds the 10 m resolution threshold. However, the latter is achieved in MWIR, with a higher thermal resolution of 8 °K but at the cost of a very low SNR of 1.

In conclusion, the small pixel size choice nearly fulfils the challenging requirement $GSD \simeq 10$ m. Sampling the PSF at the Nyquist frequency removes aliasing while keeping a

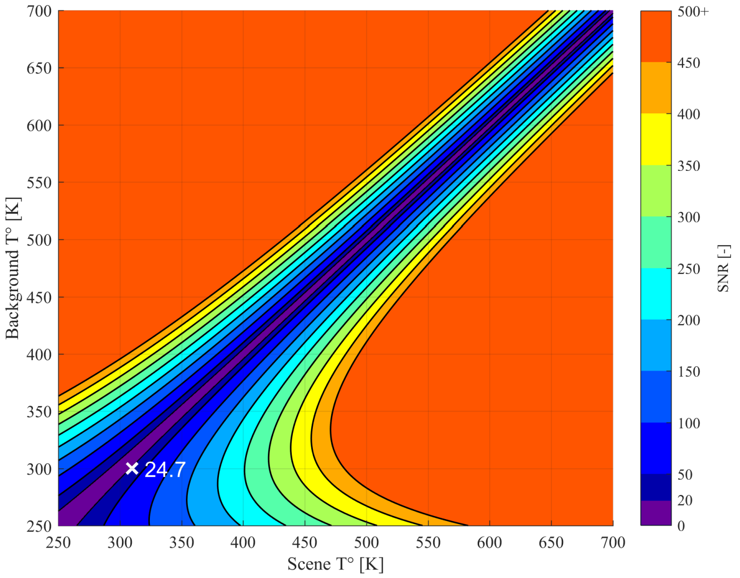


Figure 3.13: LWIR SNR contour plot for various scene and background temperatures. The white cross displays the SNR value of the considered optical configuration, assuming a 300 °K background and 310 °K scene.

relatively high 25 LWIR SNR. The SNR could be further improved by increasing the pixel size, for instance, until $F\lambda/p_x = 1$, at the cost of some aliasing. However, the strong GSD requirement could not be fulfilled, at least in LWIR. The MWIR waveband enables high resolution images at the cost of the SNR if the background an scene temperatures are kept low (around 300 °K). However, as depicted in Figure 3.15, when the scene has a high temperature, the MWIR SNR highly increases, making the optical system very efficient for high resolution and high scene temperatures imaging. Another lever to increasing the SNR would be modifying the detector's operating frequency bandwidth Δf or the peak detectivity. Since the selected detectivity is already high, only increasing the detector's operating frequency would really impact the SNR.

3.5 Annex

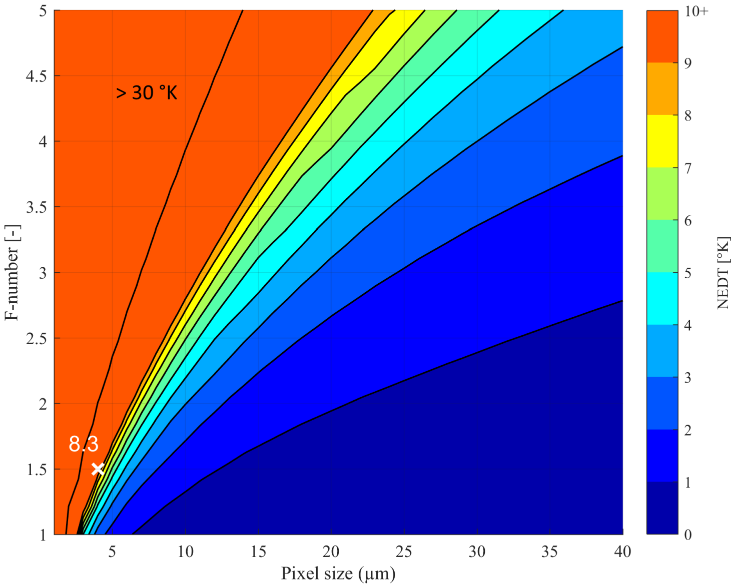


Figure 3.14: NEDT [°K] contour map computed using Equation 3.20 in MWIR. The white cross correponds to the NEDT value considering an F/1.5 system with a 8 µm pixel pitch, as explained in Section 3.4.1. The considered detector operates in a frequency bandwidth $\Delta_f = 10$ kHz and a peak detectivity $D^*_{peak} = 10^{11}$ [cm.Hz$^{1/2}$.W^{-1}]. The background temperature T_B is set to 300 °K.

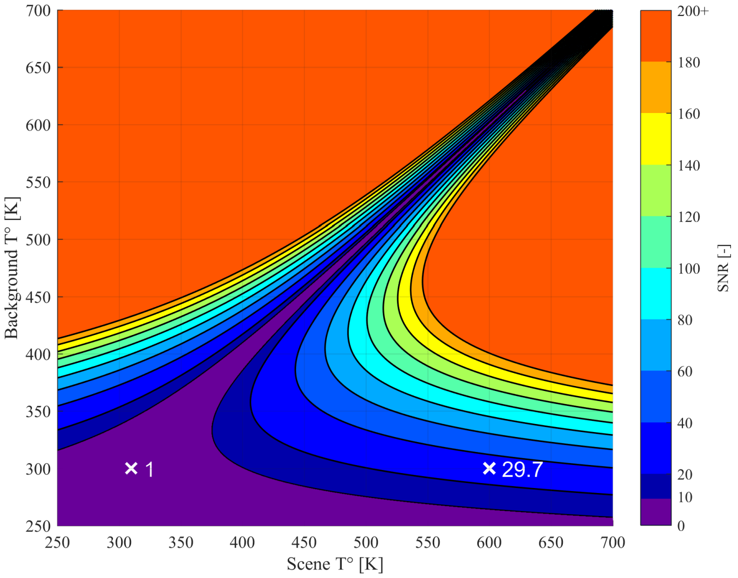
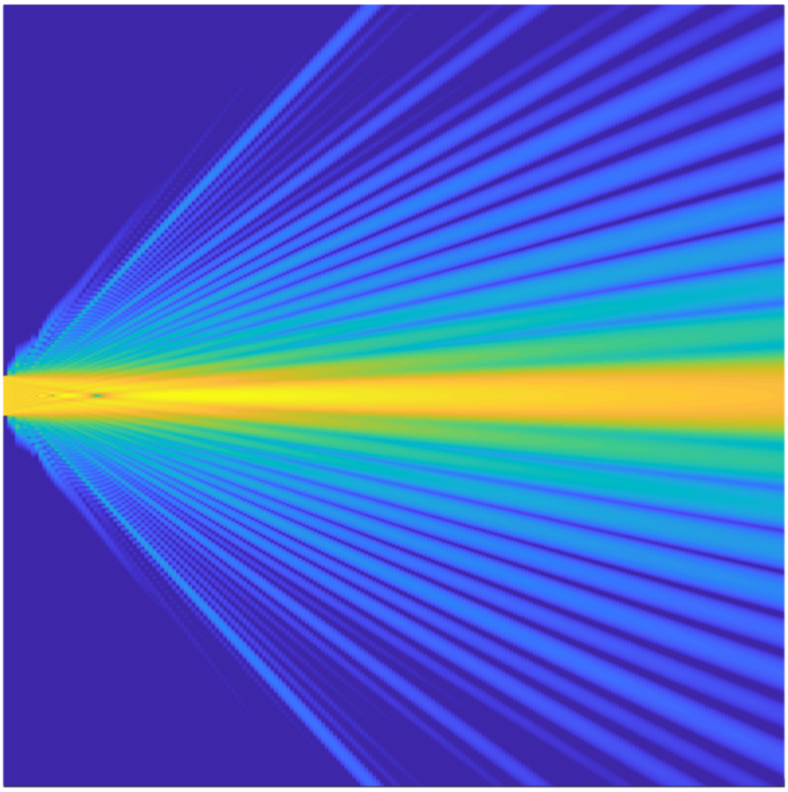


Figure 3.15: MWIR SNR contour plot for various scene and background temperatures. White crosses display the SNR value of the considered optical configuration, assuming a 300 °K background and 310 °K or 600 °K scenes.

Part II

Part 2: Modelling tools



Chapter 4

Introduction to Fourier Optics

The following chapters introduce modelling tools and methods for a specific application: the multilayer diffractive optical element (MLDOE), defined in Chapter 9. After designing a MWIR camera in Chapter 7, following the requirements established in Part I, we have been unable to extend it to LWIR using standard diffractive lenses (Chapter 8). Fortunately, the MLDOE constitutes a prominent candidate for building a dual-band hybrid and compact camera. The modelling of this complex diffractive element is the topic of this chapter, in which We develop a free-space wave propagation tool using the angular spectrum method (Fourier optics). This computer-friendly tool is very fast and accurate, calculating a wavefront phase and amplitude at any position along the propagation direction. Consequently, it constitutes our go-to method for retrieving the MLDOE focal length and its various figures of merit, such as the Strehl ratio. Since this method only works in free space, it cannot describe how an MLDOE deforms a plane wavefront. We end this chapter by describing various improvements to the method, related to its sampling and off-axis version.

Contents

In this Chapter, vectors are written in bold.

4.1 Angular spectrum representation

An electromagnetic wave propagating in a medium free of charge and current obeys the fundamental Maxwell's equations, given in MKS units:

$$\begin{cases} \nabla.\epsilon\mathbf{E} = 0 \\ \nabla.\mu\mathbf{H} = 0 \\ \nabla \times \mathbf{E} = -\mu\dfrac{\partial\mathbf{H}}{\partial t} \\ \nabla \times \mathbf{H} = \epsilon\dfrac{\partial\mathbf{E}}{\partial t} \end{cases} \tag{4.1}$$

E is the electric vector field, with scalar components (E_x,E_y,E_z), and **H** is the magnetic vector field, with components (H_x,H_y,H_z). ϵ and μ respectively are the permeability and permittivity of the dielectric medium in which the wave propagates. The **E** and **H** fields are functions of both position **P** and time t. The symbols × and . represent the vector cross product and the vector dot product. The nabla operator is written: $\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial x}\mathbf{j} + \frac{\partial}{\partial x}\mathbf{k}$, where **i**, **j** and **k** are unit vectors in the x, y, and z directions, respectively.

We assume that the propagation medium has the following properties:

- Dielectric
- Linear: the superposition property of linear systems holds
- Homogeneous: The permittivity (ϵ) is constant throughout the wavelength region of propagation
- Nondispersive: In the region of propagation, the permittivity (ϵ) is wavelength independent.
- Nonmagnetic: the magnetic permeability $\mu = \mu_0$, the permeability of vacuum

Such medium will be referred to as "free space" in the following. Under those assumptions, Maxwell's equations take a simplified form and are uncoupled. The analytical study of Maxwell's equation in free space, shown in the following, involves Fourier transforms. This research field is therefore called "Fourier Optics" and was introduced by Joseph W. Goodman [34]. This section explains the main steps to developing the rigorous free-space propagation method called "The angular spectrum of plane waves" (AS). It is exclusively based on the work of Goodman [34].

4.1.1 Helmholtz equation

We first introduce two well-known quantities: the velocity of propagation in vacuum c and the refractive index n:

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{4.2}$$

$$n = \sqrt{\epsilon/\epsilon_0} \tag{4.3}$$

where ϵ_0 is the vacuum permittivity. We also recall a well-known vector identity:

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla.\mathbf{E}) - \nabla^2\mathbf{E} \tag{4.4}$$

where ∇^2 is the Laplacian operator. Applying the $\nabla\times$ operation to the left and right sides of Equation 4.1 for the previously described free space medium yields:

$$\nabla^2\mathbf{E} - \frac{n^2}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = \overrightarrow{0} \tag{4.5}$$

This equation also applies to the magnetic field **H**. Since the vectorial wave Equation 4.5 is obeyed for both electric and magnetic fields, it is also obeyed for all the scalar components (E_x,E_y,E_z,H_x,H_y,H_z). Therefore it is possible to summarize the behaviour of all components of **E** and **H** through a single scalar wave equation:

$$\nabla^2 u(\mathbf{P},t) - \frac{n^2}{c^2}\frac{\partial^2 u(\mathbf{P},t)}{\partial t^2} = 0 \tag{4.6}$$

where $u(\mathbf{P},t)$ represents any of the scalar fields components, with explicit dependency in the position **P** and the time t. Recall that Equation 4.6 is only valid in free space. For example, the permittivity $\epsilon(\mathbf{P})$ becomes position-dependent if the medium is inhomogeneous. The wave Equation 4.5 becomes:

$$\nabla^2\mathbf{E} + 2\nabla(\mathbf{E}.\nabla \ln n(\mathbf{P})) - \frac{n^2P(\mathbf{P})}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = \overrightarrow{0} \tag{4.7}$$

The refractive index is now position-dependent, and the added term introduces a coupling between the various components of the electric field: E_x, E_y, and E_z may no longer satisfy the same wave equation. This coupling is critical, for example, when light propagates through a "thick" dielectric diffraction grating. A similar effect occurs at the boundary with another medium. At this boundary, a coupling is also introduced between **E** and **H** and between their respective scalar components. Consequently, even though Equation 4.6 is exact inside a homogeneous medium, it becomes an approximation at the boundaries. The error will remain small if the boundary condition affects a small part of the propagation area (see Section 4.2.2).

In the following, we consider a monochromatic wave. A scalar field of optical frequency ν can be explicitly expressed as:

$$u(\mathbf{P},t) = A(\mathbf{P})\cos\left(2\pi\nu t + \Phi(\mathbf{P})\right) \tag{4.8}$$

where $A(\mathbf{P})$ and $\Phi(\mathbf{P})$ are the amplitude and phase of the wave, at position **P**. The field complex "phasor" is defined as:

$$U(\mathbf{P}) = A(\mathbf{P})\exp\left[i\Phi(\mathbf{P})\right] \tag{4.9}$$

The scalar field is often written using the complex phasor notation:

$$u(\mathbf{P},t) = Re(U(\mathbf{P})\exp\left[i2\pi\nu t\right]) \tag{4.10}$$

If the real disturbance u(**P**,t) represents an optical wave propagating inside a free space medium, it must satisfy the scalar wave Equation 4.6. Using the exponential notation of the field (Equation 4.9) leads to the time-independent Helmholtz equation:

$$(\nabla^2 + k^2)U(\mathbf{P}) = 0 \tag{4.11}$$

where k is called the "wave number", defined by:

$$k \equiv 2\pi n \frac{\nu}{c} = \frac{2\pi}{\lambda} \tag{4.12}$$

λ is the wavelength of the wave propagating inside the dielectric medium ($\lambda = c/n\nu$). In conclusion, the complex phasor of any monochromatic scalar field propagating in a free space medium must satisfy the time-independent Helmholtz equation. This equation is exact as long as no boundary is met and as long as the medium satisfies the free space conditions.

4.1.2 Angular spectrum propagation method

This section describes an exact wave propagation method based on the Helmholtz equation. Let us assume that a scalar field $U(x, y, 0)$ is incident on a plane $z = 0$, and we want to propagate this field in free space medium until a parallel plane $z > 0$ along the z-axis (optical axis).

The initial scalar field $U(x, y, 0)$ has a two-dimensional Fourier transform given by:

$$A(u, v, 0) = \iint\limits_{-\infty}^{+\infty} U(x, y, 0) \exp\left[-i2\pi(ux + vy)\right] \mathrm{d}x\mathrm{d}y \tag{4.13}$$

The inverse Fourier transform physically describes the field $U(x, y, 0)$ as a collection of simpler complex-exponential functions: plane waves.

$$U(x, y, 0) = \iint\limits_{-\infty}^{+\infty} A(u, v, 0) \exp\left[i2\pi(ux + vy)\right] \mathrm{d}u\mathrm{d}v \tag{4.14}$$

To understand the relation of the above integrand functions with plane waves, we need to consider the general formulation of a plane wave propagating along the wave vector **k**. We are only interested in the complex phasor of this wave, written as:

$$\begin{aligned} PW(x, y, z) &= \exp[i\mathbf{k}.\mathbf{P}] \\ &= \exp\left[i\frac{2\pi}{\lambda}(\alpha x + \beta y)\right] \exp\left[i\frac{2\pi}{\lambda}\gamma z\right] \end{aligned} \tag{4.15}$$

where (α, β, γ) are the direction cosines of the wave vector **k** of magnitude $|\mathbf{k}| = \frac{2\pi}{\lambda}$. Note that the direction cosines are interrelated through:

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2} \tag{4.16}$$

Now the exponential functions of Equation 4.14 appear has plane waves propagating along the wave vector $\overrightarrow{\text{k}} = \frac{2\pi}{\lambda}(\alpha\mathbf{x} + \beta\mathbf{y} + \gamma\mathbf{z})$ with:

$$\begin{aligned}\alpha &= \lambda u\\ \beta &= \lambda v\\ \gamma &= \sqrt{1-(\lambda u)^2-(\lambda v)^2}\end{aligned} \tag{4.17}$$

We define the Angular spectrum of the disturbance $U(x, y, 0)$ as:

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0) = \iint\limits_{-\infty}^{+\infty} U(x, y, 0) \exp\left[-i2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)\right] \mathrm{d}x\mathrm{d}y \tag{4.18}$$

The unknown field $U(x, y, z)$ in a plane $z > 0$ can also be represented by a Fourier transform in the same way as in Equation 4.18. The inverse Fourier transform leads to:

$$U(x, y, z) = \iint\limits_{-\infty}^{+\infty} A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) \exp\left[i2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)\right] \mathrm{d}\frac{\alpha}{\lambda}\mathrm{d}\frac{\beta}{\lambda} \tag{4.19}$$

$U(x, y, z)$ must also satisfy the Helmholtz equation (4.11), leading to the differential equation:

$$\begin{aligned}&\iint\limits_{-\infty}^{+\infty} [\nabla^2 + k^2] \left(A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) \exp\left[i2\pi(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)\right]\right) \mathrm{d}\frac{\alpha}{\lambda}\mathrm{d}\frac{\beta}{\lambda} = 0\\ &\Leftrightarrow \frac{d^2}{dz^2}A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) + (\frac{2\pi}{\lambda})^2[1-\alpha^2-\beta^2]A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = 0\end{aligned} \tag{4.20}$$

The solution is elementary:

$$A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z) = A(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; 0) \exp\left[i\frac{2\pi}{\lambda}z\sqrt{1-\alpha^2-\beta^2}\right] \tag{4.21}$$

Two cases arise from this equation:

- ($\alpha^2 + \beta^2 < 1$): The effect of propagating a field over a distance z is to apply a phase shift to every plane wave component of the field. Each plane wave component propagates at a different angle and will travel a different optical path, introducing a relative phase delay. The propagated field is the superposition of all propagated plane waves.
- ($\alpha^2 + \beta^2 > 1$): The field necessarily has a boundary condition where this situation can arise. α and β are no more interpreted as direction cosines. In this situation, the wave components are rapidly attenuated and are called "evanescent waves" and carry no energy away from the boundary (aperture).

Finally, the angular spectrum method can be written in terms of Fourier transforms:

$$\begin{cases} U(x,y;z) = \mathcal{F}^{-1}\{\mathcal{F}\{U(x,y;0)\}H_z(u,v)\} \\ H_z(u,v) = \mathcal{F}\{h_z(x,y)\} = \exp\left[ikz\sqrt{1-(\lambda u)^2-(\lambda v)^2}\right] \end{cases} \tag{4.22}$$

The spatial frequencies (u,v) must satisfy :

$$u^2+v^2 < \frac{1}{\lambda^2} \tag{4.23}$$

In conclusion, the angular spectrum method is an exact optical propagation tool for a scalar wave evolving in free space, between two parallel planes.

4.2 Fourier optics propagation on Matlab

We have developed a Matlab Fourier optics propagator based on the angular spectrum method (AS) to study the imaging performances of some innovative diffractive lenses. The Matlab implementation of the angular spectrum method is greatly inspired by the Matlab tutorial [35]. In addition to the examples shown hereafter, our Matlab propagator has been validated using many optical cases (grating, Fresnel zone plate) originating from [35]. This section explains the simulation process behind the propagator and displays the results obtained in some particular optical examples.

Figure 4.1 represents the treatment of a monochromatic scalar field:

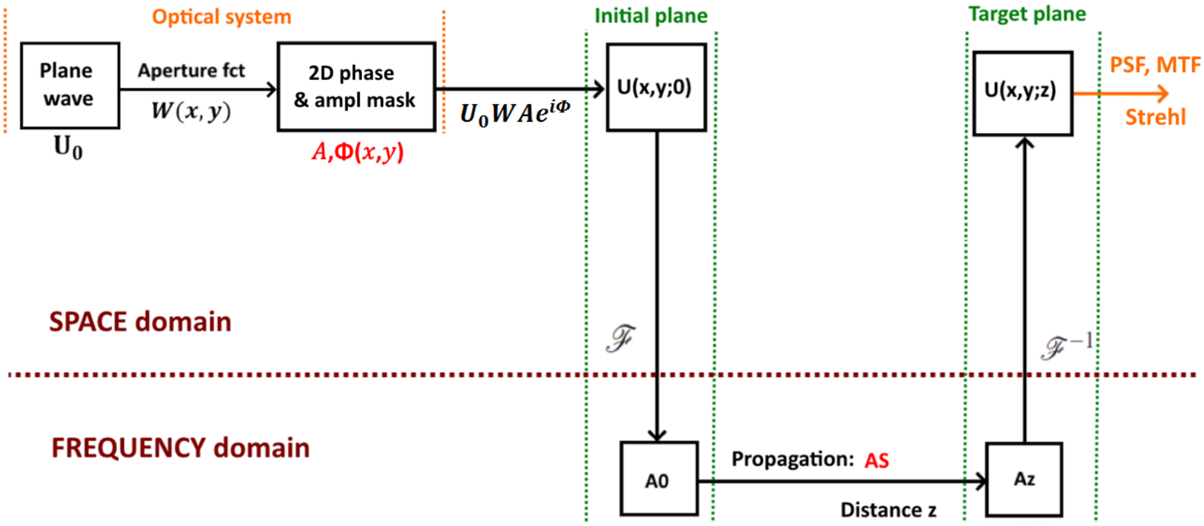


Figure 4.1: Diagram showing the treatment of a complex monochromatic scalar field using Fourier optics. The studied optical element is represented by a 2D phase and amplitude function: $\Phi(x,y)$ and $A(x,y)$, inside an aperture $W(x,y)$. The resulting scalar field in the input plane ($z=0$) is propagated in the frequency domain, using the angular spectrum (AS) method [34]. The field $U(x,y;z)$ is evaluated in the target plane ($z>0$) using different metrics (point spread function (PSF), modulation transfer function (MTF), Strehl ratio).

The propagation is divided into two steps: Obtaining the field after the optical element and propagating this field to an image plane. The last step is described by System 4.22, involving Fourier transform. In Figure 4.1, an optical component is described by a complex field phasor $U(x, u; 0) = U_0 W(x, y) A(x, y) \exp[i\Phi(x, y)]$. $W(x, y)$ is the aperture function defining the physical extension of the field. $A(x, y)$ and $\Phi(x, y)$ are the amplitude and phase functions of the field phasor. This diagram describes how the phase delay introduced by the optical component leads to a wavefront in the image plane. Note that one can apply this propagation process regardless of the expression of the functions $A(x, y)$ and $\Phi(x, y)$. Finally, at the image plane, various optical performance metrics are computed. The Point Spread Function (PSF) is defined as:

$$PSF = |U(x, y; z)|^2 \tag{4.24}$$

The PSF gives the impulse response in the image plane of a plane wave affected by an imaging element. This impulse response is limited by the diffraction pattern created by the aperture edges on the image plane for an aberration-free element.

The MTF (Modulation Transfer Function) is defined as the Fourier transform of the PSF. The MTF describes how spatial frequencies are "transmitted" through an optical system. The MTF gives a contrast value (from 0 to 1) to each frequency in the image plane. While initially maximal for an infinite number of frequencies, the contrast drops drastically due to the diffraction-limited nature of optical systems (i.e. the PSF is not a Dirac). This effect strengthens when aberrations are present. As a result, spatial frequencies progressively lose contrast when they increase until a cutoff frequency. This frequency is the maximal possible frequency that can be imaged and is defined for an incoherent source as:

$$f_c = \frac{1}{\lambda F/\sharp} \tag{4.25}$$

For an incoherent diffraction-limited imaging system, the analytical MTF is defined as [35]:

$$MTF_{an}(\nu) = \frac{2}{\pi} \mathrm{Real}\left[\arccos\left(\frac{\nu}{\mathrm{f_c}}\right) - \frac{\nu}{\mathrm{f_c}}\sqrt{1 - \left(\frac{\nu}{\mathrm{f_c}}\right)^2}\right] \tag{4.26}$$

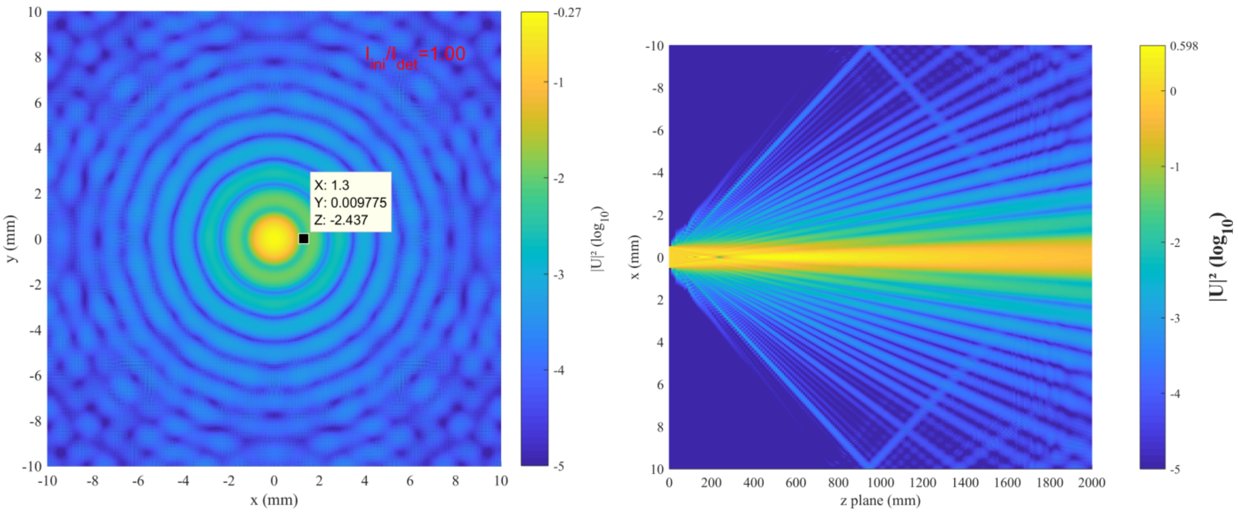
where ν a the spatial frequency. The analytical MTF is null for frequencies $\nu > f_c$.

Finally, the Strehl ratio is defined as the ratio of the MTF over the analytical MTF. Diffraction-limited optical components have a Strehl ratio of 1. This ratio is often used to describe the amount of aberrations present in an optical system but can also give interesting clues about the "focalization efficiency". In the case of diffractive optics, we use the Strehl ratio to evaluate the diffraction efficiency at the focal plane.

4.2.1 Validation of the propagator: circular aperture

In this Section, the propagator described in Figure 4.1 and implemented using Matlab is tested. An initial field propagates through a tiny circular aperture and is evaluated far from the source plane. We expect the pattern observed in the image plane to be a Fraunhofer diffraction pattern: the so-called Airy-disk. The simulation setup consists of a hole of 1 mm diameter (D) centred inside a 20 mm squared simulation window, with $M = 1024$ samples in the x and y directions. The image plane is parallel to the source plane and situated 2000 mm away. We use a wavelength of 532 nm, which gives a theoretical Airy radius of:

$$R_{Airy} = 1.22\lambda\frac{f}{D} \simeq 1.3 \text{ mm} \tag{4.27}$$



(a) PSF at the plane z = 2000 mm.

(b) PSF along the optical axis.

Figure 4.2: Angular spectrum of a plane wave propagating in free space through a 1 mm diameter circular aperture (plane $z = 0$ mm). a) PSF (logarithmic scale) at the plane $z = 2000$ mm. b) PSF (logarithmic scale) along the optical axis. The first zero of the Airy pattern is obtained at a radius of 1.3 mm. Edge effects are present in both figures causing unwanted reflections.

Figure 4.2 a. shows that the first zero of the Airy function is obtained for a radius of 1.3 mm, as expected by Equation 4.27. However, the Airy pattern is disturbed at the edges of the simulation window. This disturbance originates from the field reflecting at the edges of the simulation window (Figure 4.2 b.).

4.2.2 Bandlimited angular spectrum

In this section, based on the work of K. Matsushima [36], we discuss some sampling precautions to take when using the angular spectrum method (Equation 4.22).

The angular spectrum method involves Fourier transforms, which are calculated in practice using discrete Fourier transforms (DFTs). However, DFTs are numerically exact when periodic fields are involved in space and frequency domains. In this case, the convolution with the transfer function H_z is "circular" (signal processing term). Therefore, for aperiodic functions such as the aperture function or the transfer function H_z, the circular convolution causes numerical errors at the edges of the computation window. When the extent of the field is comparable to the computation window size, the angular spectrum is no longer considered accurate (Figure 4.2 b.). The solution discussed in [36] is to double the input window size using zero-adding artificially. By doing so, the initial field extent is sufficiently small compared to the sampling window to convert the circular convolution into a linear one when propagating the field. After the propagation, the resulting field is clipped to retrieve the original sampling. This process is depicted in Figure 4.3:

As seen in Figure 4.2, some aliasing artefacts occur in the output plane and during the propagation. This effect is not caused by the previously mentioned circular convolution issue since the sampling window is 20 times higher than the aperture (hole of 1 mm

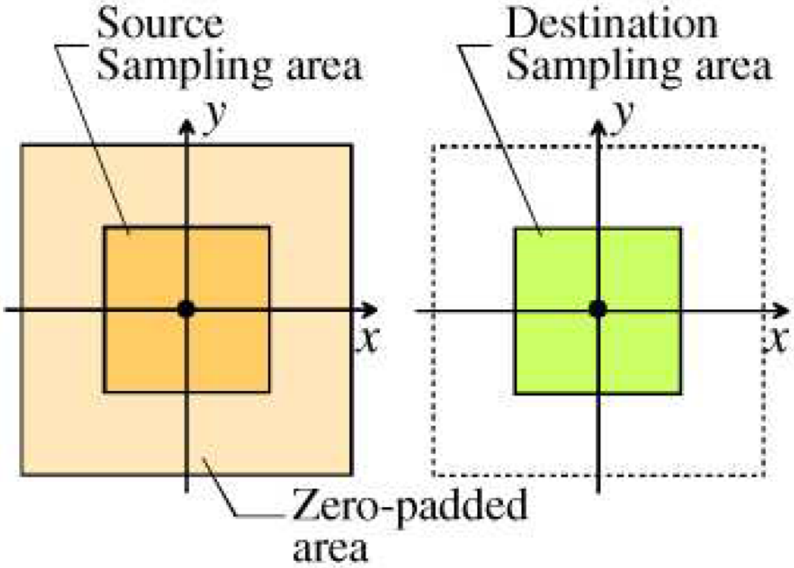


Figure 4.3: Zero-padding process: conversion of a circular convolution into a linear convolution. The initial sampling window is extended to avoid any numerical error at the edges. The output field is clipped to retrieve the original sampling. Original source [36].

diameter). As explained by Matsushima [36], the aliasing issue comes from the sampling of the transfer function H_z. The transfer function $H_z(u,v)$, defined in the frequency domain, is rewritten here for clarity:

$$H_z(u,v) = \exp\left[i2\pi z\Phi\right] = \exp\left[i2\pi z\sqrt{\lambda^{-2} - u^2 - v^2}\right] \tag{4.28}$$

u and v are spatial frequencies. Figure 4.4 shows a one dimensional plot of the transfer function:

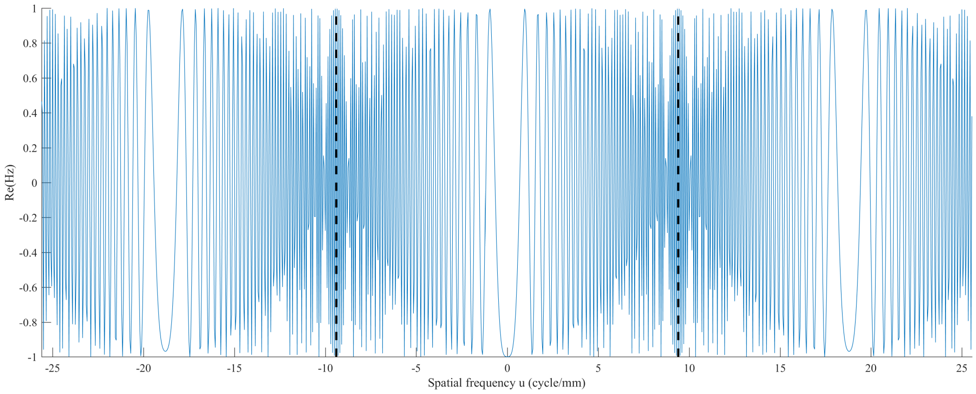


Figure 4.4: One dimensional transfer function $H_z(u)$. After a certain high frequency (black dashed curves), the transfer function becomes aliased, causing numerical artefacts in the propagated field.

The aliasing effect shown in Figure 4.4 cannot be solved by increasing the number of samples because it will also increase the maximal spatial frequency. The transfer function is a kind of *chirp* function with respect to u: the signal frequency increases exponentially with u. The denomination "signal frequency" represents the frequency of peaks and valleys in the function $H_z(u)$ and is not a physical frequency. The solution brought by Matsushima

[36] is to artificially limit the transfer function to a maximal spatial frequency, depicted with black dashed lines in Figure 4.4. The calculation of this maximal sampling frequency u_{limit} is detailed in [36] but the main steps are written here for completeness.

According to [34], the local signal frequency of the function $H_z(u) = \exp\left[i\Phi(u)\right]$, is given by:

$$f_u = \frac{1}{2\pi}\frac{\partial \Phi}{\partial u} \tag{4.29}$$

with $\Phi(u) = 2\pi z\sqrt{\lambda^{-2} - u^2}$. The spatial frequency limit u_{limit} can be retrieved using the Nyquist sampling theorem:

$$\Delta u^{-1} \geq 2|f_u| \tag{4.30}$$

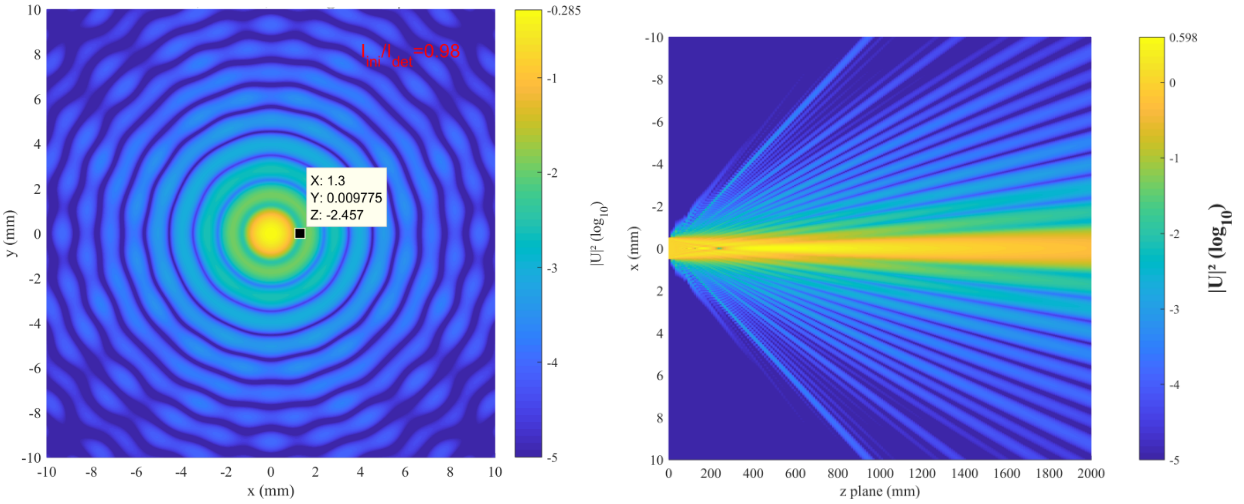
Here, note that the sampling interval of the spatial frequency u: Δ_u is defined as $\Delta_u = (2S)^{-1}$ because of the zero-padding process of the sampling window previously described. Finally, the spatial frequency limitation that avoids aliasing is given by:

$$u_{limit} \equiv [\lambda\sqrt{(2\Delta_u z)^2 + 1}]^{-1} \tag{4.31}$$

As a result, the transfer function must be clipped within a bandwidth of $2u_{limit}$. In two dimensions, the bandlimited transfer function $H_z(u, v)$ is defined as:

$$H_z'(u, v) = H_z(u, v)\mathrm{rect}(\frac{u}{2u_{limit}})\mathrm{rect}(\frac{v}{2v_{limit}}) \tag{4.32}$$

where $v_{limit} = [\lambda\sqrt{(2\Delta_v z)^2 + 1}]^{-1}$ and $\mathrm{rect}(x)$ is a rectangular function of unity half width [37]. Using this clipping of the transfer function, Figure 4.5 is obtained, following the exact same setup as Figure 4.2:



(a) PSF at the plane z = 2000 mm.

(b) PSF along the optical axis.

Figure 4.5: Angular spectrum of a plane wave propagating in free space through a 1 mm diameter circular aperture (plane $z = 0$ mm). a) PSF (logarithmic scale) at the plane $z = 2000$ mm. b) PSF (logarithmic scale) along the optical axis. The first zero of the Airy pattern is obtained at a radius of 1.3 mm. Aliasing has been removed using the bandlimited transfer function (Equation 4.32.)

So far, we have described a necessary condition to avoid the aliasing of the transfer function. However, this bandlimited region may be insufficient to propagate the field. In [36], Matshumima provides a geometrical model to give the value of the highest spatial frequency that is "physically" needed. He showed that, in any case, this required frequency was included in the rectangle region previously computed (Equation 4.32). Therefore, the presented method always satisfies the sampling theorem and the minimum bandwidth condition.

4.2.3 Off-axis propagation using the AS method

The angular spectrum method suffers from the intrinsic sampling limitations detailed in Section 4.2.2. However, they are not the only limitations of this wave propagation method. The Fourier transform definition (Equation 4.22) of the angular spectrum implies that the source and target planes have the same size, are parallel to each other and are perpendicular to the propagation direction. We can only simulate a field with divergence smaller than the sampling window. Therefore, the sampling window size must be increased to encompass the whole beam in case of a diverging or off-axis field. This process of extending the sampling window is depicted in Figure 4.6 a):

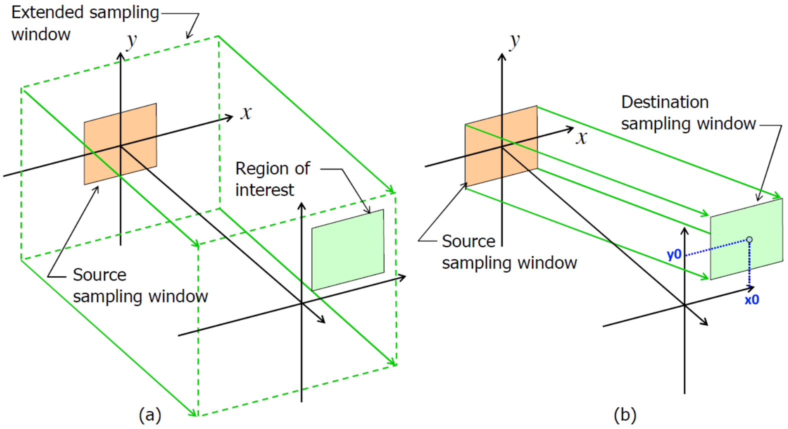


Figure 4.6: Calculation of the diffracted field using (a) extended sampling window and (b) shifted angular spectrum propagation, where the region of interest is apart from the optical axis in the destination pane. In Figure 4.6 a., the output field must be clipped to recover a resulting field of the same size as initially.

A more efficient propagation method, developed by K. Matsushima in [38] is called shifted angular spectrum (Figure 4.6 b)). It allows for computing the angular spectrum for off-axis propagation. Instead of extending the input and output planes to encompass the field, this method includes shifting the target plane in the x and y directions. Therefore, this method does not require extending the size of the sampling window.

The source field $u(x, y, 0)$ is given in the source plane $(x, y, 0)$ while the destination field $u(\widehat{x}, \widehat{y}, z_0)$ is given in the destination plane $(\widehat{x}, \widehat{y}, z_0)$. The origin of the lateral coordinates $\widehat{x}$ and $\widehat{y}$ is shifted from the source coordinates as follows:

$$\begin{cases} \widehat{x} = x - x_0 \\ \widehat{y} = y - y_0 \end{cases} \tag{4.33}$$

The angular spectrum method (Equation 4.22) is rewritten as:

$$\begin{cases} U(\widehat{x}, \widehat{y}; z_0) = \mathcal{F}^{-1}\{\mathcal{F}\{U(x, y; 0)\}H(u, v; z_0)\} \\ H(u, v; z_0) = \exp\left[i2\pi(x_0 u + y_0 v + z_0 w)\right] \\ w(u, v) = \sqrt{\lambda^{-2} - u^2 - v^2} \end{cases} \tag{4.34}$$

The demonstration of this expression is detailed in [38]. The The demonstration of this expression is done in detail in [38]. The simple numerical implementations of Equation 4.34 will likely cause huge sampling problems if the shifted transfer function is not bandlimited. Using the same reasoning as in Section 4.2.2, the bandlimited angular spectrum (Equation 4.32) is rewritten as:

$$H'(u, v; z_0) = H(u, v; z_0)\mathrm{rect}\left(\frac{\mathrm{u - u0}}{\mathrm{u_{width}}}\right)\mathrm{rect}\left(\frac{\mathrm{v - v0}}{\mathrm{v_{width}}}\right) \tag{4.35}$$

where rect(x) is a rectangular function of unity half width ([37]), u_0, v_0, u_{width}, v_{width} are defined in Table 4.1 [38]:

Case	u_0	u_{width}	Case	v_0	v_{width}
$S_x < x_0$	$(u_{\mathrm{limit}}^{(+)} + u_{\mathrm{limit}}^{(-)})/2$	$u_{\mathrm{limit}}^{(+)} - u_{\mathrm{limit}}^{(-)}$	$S_y < y_0$	$(v_{\mathrm{limit}}^{(+)} + v_{\mathrm{limit}}^{(-)})/2$	$v_{\mathrm{limit}}^{(+)} - v_{\mathrm{limit}}^{(-)}$
$-S_x \leq x_0 < S_x$	$(u_{\mathrm{limit}}^{(+)} - u_{\mathrm{limit}}^{(-)})/2$	$u_{\mathrm{limit}}^{(+)} + u_{\mathrm{limit}}^{(-)}$	$-S_y \leq y_0 < S_y$	$(v_{\mathrm{limit}}^{(+)} - v_{\mathrm{limit}}^{(-)})/2$	$v_{\mathrm{limit}}^{(+)} + v_{\mathrm{limit}}^{(-)}$
$x_0 \leq -S_x$	$-(u_{\mathrm{limit}}^{(+)} + u_{\mathrm{limit}}^{(-)})/2$	$u_{\mathrm{limit}}^{(-)} - u_{\mathrm{limit}}^{(+)}$	$y_0 \leq -S_y$	$-(v_{\mathrm{limit}}^{(+)} + v_{\mathrm{limit}}^{(-)})/2$	$v_{\mathrm{limit}}^{(-)} - v_{\mathrm{limit}}^{(+)}$

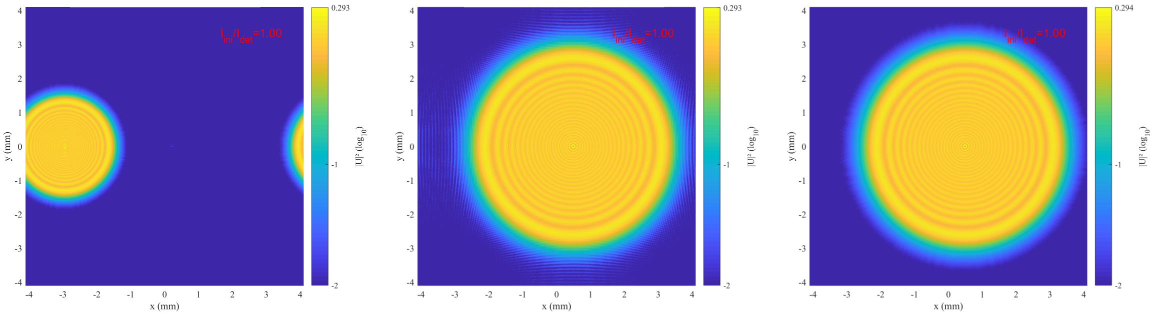
Table 4.1: Constants used for the bandlimiting process to avoid aliasing errors.

S_x and S_y are the sampling window sizes in both the x and y directions. Finally, the constants u_{limit} and v_{limit} have a similar expression than in Equation 4.31:

$$\begin{cases} u_{limit}^{(\pm)} \equiv \left[z_0^2\left(x_0 \pm \dfrac{1}{2\Delta u}\right)^{-2} + 1\right]^{-1/2} \lambda^{-1} \\ v_{limit}^{(\pm)} \equiv \left[z_0^2\left(y_0 \pm \dfrac{1}{2\Delta v}\right)^{-2} + 1\right]^{-1/2} \lambda^{-1} \end{cases} \tag{4.36}$$

When $x_0 = 0$ and $y_0 = 0$, Equation 4.36 reduces to the previous definition of u_{limit} and v_{limit} (Equation 4.31) in Section 4.2.2. Recall that the frequency sampling interval $\Delta u = (2S_x)^{-1}$ is due to the zero padding process to avoid edges numerical errors.

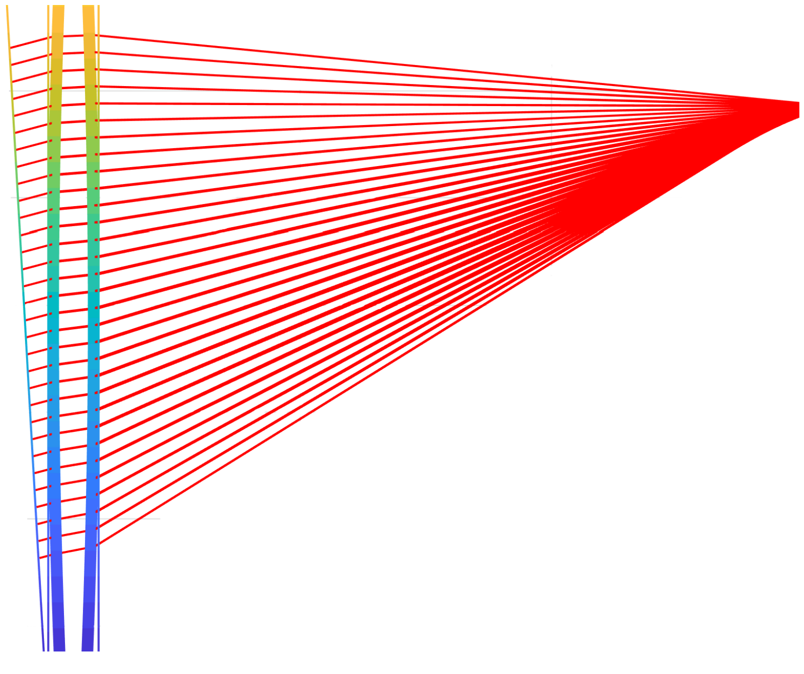
We now apply the shifted angular spectrum method to an off-axis plane wave going through a circular aperture. We use the same setup as in [38] for validation: the window size $S = S_y = S_x = 8.192$ mm, the aperture diameter $D = 6$ mm and the wavelength $\lambda = 532$ nm. The target plane is located as $z_0 = 400$ mm, and we consider a shift of $x_0 = 10$ mm. The incidence angle $\theta = 1.5°$. The results are displayed in Figure 4.7:



(a) Extended AS (b) shifted AS not bandlimited. (c) shifted bandlimited AS.

Figure 4.7: Angular spectrum of an off-axis ($\theta = 1.5°$) plane wave propagating in free space through a 6 mm diameter circular aperture (plane $z = 0$ mm). The PSF at the destination plane ($z = 400$) is computed in three different cases: a) conventional AS with 3x extended sampling window, b) shifted AS not bandlimited, and c) shifted bandlimited AS. The extended sampling in a) is insufficient to recover the field divergence, and aliasing can be seen in b)

Figure 4.7a shows that increasing the sampling window three times is still insufficient to recover the target field in this case. The computation effort is, therefore, very limiting if the conventional angular spectrum method is considered for off-axis simulations. Figure 4.7b displays the result of the shifted method: the target plane is centred around $\widehat{x} = x - x0$, encompassing the output field. However, aliasing can be seen because the transfer function has not been bandlimited (Equation 4.35), contrary to Figure 4.7c. The numerical implementation of the shifted angular spectrum is therefore validated, the results being identical to those obtained in [38] with the same simulation setup.



Chapter 5

Analytic optical path length calculator

Following the development of MLDOE modelling tools started in Chapter 4, we propose using a combination of ray-tracing and Fourier optics, called the ray model, to approximate the phase delay introduced by an MLDOE (Chapters 10, 12). This model is a significant innovation of this thesis and enables estimating the performance/behaviour of an MLDOE using a fast and parameterisable approach. This chapter details the development of an analytical ray-tracing calculator implemented on Matlab. For speed and parametrisation reasons, we have created this calculator specifically for MLDOEs and validated it using professional three-dimensional ray-tracing tools (ASAP). We analytically and simultaneously retrieve all the rays' optical path length (OPL) using Snell's law of refraction. This is possible since the surface equation of an MLDOE is perfectly known, rendering the determination of intersections, normal derivatives, and refractions analytical. Our calculator is accurate since very few approximations are made, and extremely fast since only element-wise matrix multiplications are needed to treat thousands of rays.

Contents

5.1 Introduction

This chapter explains the development of an analytical optical path length (OPL) calculator based on ray tracing. This software is dedicated to studying specific optical components called multilayer diffractive optical elements (MLDOEs) but is also applicable to refractive lenses as a validation use case. Its goal is to provide a fast and specific OPL calculator that can analytically compute thousands of rays in a few seconds, to be used as an MLDOE optimisation tool. "Analytic" refers to the calculation of rays, whose directions and intersections with optical interfaces are handled analytically. Processing the OPL of diffractive elements based on ray-tracing is a method already used in [39, 40]. It is discussed in Chapter 10, providing an alternative phase modelling tool for MLDOEs.

With high speed and accuracy, powerful existing ray-tracing software such as CodeV and ASAP allows one to optimise/model any refractive optical systems, even the most complex ones. However, they are not specifically designed to compute OPLs or treat complex diffractive elements. While it is possible in ASAP to build an MLDOE, trace rays and export the optical path lengths, it is not a convenient method. It requires a high amount of time, and the results must be exported to Matlab, where the main phase modelling and optical propagation code are implemented (See Chapters 4 and 10). In addition, using a single software such as ASAP does not guarantee the validity of the results since many mistakes can occur during the definition of the MLDOE layout, the source and during the Matlab importation process. This chapter provides the mathematical and algorithmic descriptions of the analytical OPL calculator developed in Matlab. It has been validated using both CodeV and ASAP in the case of refractive lenses and MLDOEs. This software has four requirements:

- To be much faster than ASAP or CodeV for a high number of rays
- To remain simple and parametric
- To provide accurate OPL values.
- To be specific for MLDOEs and refractive lenses

The Matlab calculator is compared to CodeV or ASAP, using optical metrics such as the spot, the MTF or the PSF. Optical metrics are retrieved by computing the diffracted field at the focal plane using Fourier optics. The initial scalar fields are defined using the OPL calculated by Matlab, CodeV and ASAP.

For MLDOEs, we make some approximations for particularly complex ray paths (multiple refractions on interfaces, total internal reflection). However, the Matlab software is accurate as long as the optical results after Fourier optics propagation match the exact ASAP or CodeV outputs. In addition, this analytic OPL calculator only gives *in fine*, an MLDOE phase approximation that will always be much less accurate than rigorous electromagnetic calculations. Therefore, analytical ray tracing is more than sufficient.

The following sections describe the algorithmic and mathematical principle behind the Matlab analytic OPL calculator and the validation using CodeV and ASAP for refractive lenses and MLDOEs.

5.2 Ray-tracing formalism for Matlab

5.2.1 Fundamental steps and equations

This section describes the main steps and fundamental equations involved in the analytic ray-tracing algorithm. The vectors are written in bold and are considered unitary. Each interface is described by an analytical equation $f(x, y, z) = 0$, allowing one to compute its precise intersection with any ray. The "analytic" requirement eliminates any sampling issue, ensuring the speed of the implemented OPL calculator. The main steps are described in Figure 5.1:

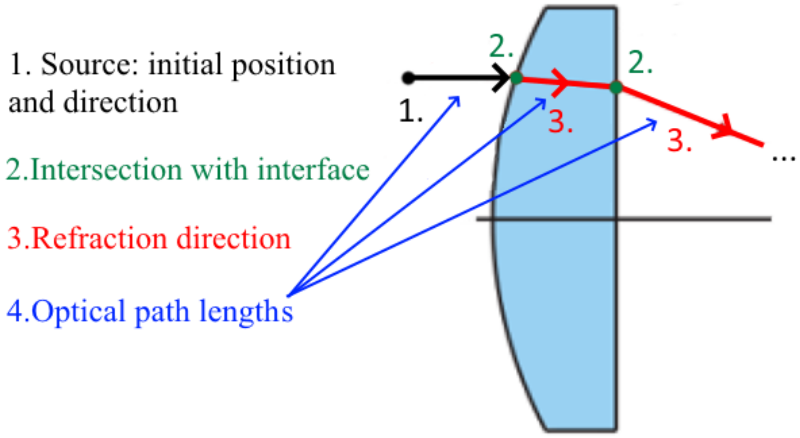


Figure 5.1: Ray-tracing steps implemented in the analytical calculator to obtain the optical path length (OPL) of any ray inside an optical element.

In the following, we mathematically describe the four main steps illustrated in Figure 5.1:

Step 1:

The source plane is the plane where all the OPL calculations originate. Therefore, it must bust orthogonal to the initial beam (parallel rays). For on-axis calculation, the source plane is defined in $z = 0$ and constitutes the aperture stop (AS). All the rays coming from various fields of view intersect by definition at the source plane $z = 0$. The first optical interface also starts in $z = 0$. The direction vector of any ray is trivially $\mathbf{I}(0, 0, 1)$. For off-axis incidences, the source plane is tilted to remain perpendicular to the beam, as displayed in Figure 5.2:

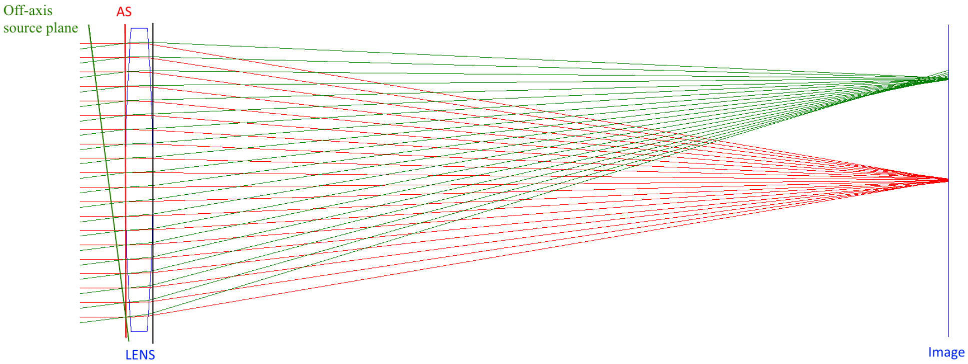


Figure 5.2: Example of a CodeV ZnS F/3 biconvex lens. On-axis (red) and off-axis (green) incidences are represented. The off-axis angle is $\theta_x = 7.125°$. The red plane at $z = 0$ is the aperture stop, where all the FoVs intersect, and is also the on-axis source plane for OPL calculations. The tilted green plane is perpendicular to the green beam and defines the off-axis OPL source. Finally, OPL computations end at the black plane, located at the back of the biconvex lens.

The director coefficient of the off-axis beam $\mathbf{U}(a, b, c)$ is expressed for any incidence angles (θ_x, θ_y) as:

$$\begin{cases} \mathbf{U}(a,b,c) = R_x * R_y * \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \\ R_x = \begin{bmatrix} 1 & 0 & 0 \\ \cos\theta_x & -\sin\theta_x & 0 \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \\ R_y = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \end{cases} \tag{5.1}$$

We normalize the vector $\mathbf{U}$ so that: $a^2 + b^2 + c^2 = 1$. Any off-axis incident ray is normal to the source plane $\mathcal{P}$, defined by the equation:

$$ax + by + cz + d = 0 \Leftrightarrow d = -ax_p - by_p - cz_p \tag{5.2}$$

The parameter d is defined by the intersection point $P(x_p, y_p, z_p)$ between the source plane $\mathcal{P}$ and the aperture stop. When the source plane $\mathcal{P}$ is titled, the initial position (x_i, y_i, z_i) of each ray on this plane must be computed. We start with a grid of rays (x_0, y_0, z_0) defined at the aperture stop plane (vertical plane $z_0 = 0$). The titled initial direction coefficient $\mathbf{U}$(a,b,c) has been given in Equation 5.1. We then compute all the intersection points (x_i, y_i, z_i) between the rays and the source plane:

$$\begin{cases} x_i = \dfrac{a}{c}(z_i - z_0) + x_0 \\ y_i = \dfrac{b}{c}(z_i - z_0) + y_0 \\ z_i = (-ax_i - by_i - d)/c \\ \Leftrightarrow z_i = -\dfrac{a^2+b^2}{c^2} z_i + \dfrac{a^2+b^2}{c^2} z_0 - \dfrac{ax_0 + by_0 + d}{c} \\ \Leftrightarrow z_i = z_0(1 - c^2) - c(ax_0 + by_0 + d) \end{cases} \tag{5.3}$$

The first OPL value, between the source plane and the aperture stop $z = 0$ is simply given by: $OPL = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2 + (z_0 - z_i)^2}$.

Step 2:

The intersection point $P(x, y, z)$ between a ray and an optical interface $f(x, y, z) = 0$ is obtained by solving an intersection equation. A ray is described in 3D by a line, directed by $\mathbf{I}(a_i, b_i, c_i)$, and starting at the previous intersection point $P_0(x_0, y_0, z_0)$. We have:

$$\begin{cases} b_i(z - z_0) - c_i(y - y_0) = 0 \\ c_i(x - x_0) - a_i(z - z_0) = 0 \\ a_i(y - y_0) - b_i(x - x_0) = 0 \\ f(x, y, z) = 0 \end{cases} \tag{5.4}$$

The first three equations mean that $P(x, y, z)$ is on the line passing by P_0 with direction vector $\mathbf{I}$. The last equation is the interface analytical equation $z = f(x, y)$. System 5.4 can be solved by substitution to find $P(x, y, z)$.

Optical interfaces with circular symmetry display a general profile called "aspheric". The expression of this profile in the z direction for any radius r is [41]:

$$z(r) = \frac{r^2}{R + R\sqrt{1 - (1+k)\dfrac{r^2}{R^2}}} + \alpha_4 r^4 + \alpha_6 r^6 + ... \tag{5.5}$$

R represents the curvature radius of the sphere associated with the aspheric profile. The first term of Equation 5.5 describes a conic profile, with conic constant k. The additive aspheric terms α_i are often set to 0 unless specific aberration correction is needed. In the following, for simplicity, we consider conic interfaces only ($\alpha_i = 0$). Two particular profiles arise: the spherical profile ($k = 0$) and the parabolic profile ($k = -1$):

$$\begin{cases} z_{sph}(r) = \dfrac{r^2}{R + \sqrt{R^2 - r^2}} \\ z_{parab}(r) = \dfrac{r^2}{2R} \end{cases} \tag{5.6}$$

Note that solving system 5.4 using a conic (spherical or parabolic) profile always leads to two solutions (one is out of the design bounds).

Step 3:

An incident ray, intersecting an interface of equation $f(x, y, z) = 0$ at the coordinate points $P_0(x_0, y_0, z_0)$ with an initial 3D direction vector $\mathbf{I}(a_i, b_i, c_i)$ will be refracted in the direction $\mathbf{R}(a_r, b_r, c_r)$ following Snell's law of refraction. This direction is defined according to the normal leaving the interface at point P_0: $\mathbf{N}(a_n, b_n, c_n)$. The normal $\mathbf{N}$ direction vector is mathematically defined by:

$$\mathbf{N}(a_n, b_n, c_n) = \nabla_f * (x_0, y_0, 1) = \left(x_0 \frac{df}{dx}(x_0), y_0 \frac{df}{dy}(y_0), \frac{df}{dz}(z_0)\right) \tag{5.7}$$

The vectors **I, R, N** are considered unitary in the following. Using the vectorial expression of Snell's law [42] is more practical to eliminate sign problems coming from trigonometric function inversions. We define n_1 and n_2 as the refractive indexes of the media, respectively, before and after the interface. θ_i and θ_r are the incident and refracted angles defined relative to the normal. The vectorial Snell's law is expressed as:

$$\mathbf{R} = \frac{n_1}{n_2}.\mathbf{I} + \left[\frac{n_1}{n_2}.\cos(\theta_i) - \cos(\theta_r)\right].\mathbf{N} \tag{5.8}$$

With the expression of the cosines being :

$$\begin{cases} \cos(\theta_i) = \mathbf{I}.\mathbf{N} \\ \cos(\theta_r) = \sqrt{1 - (\frac{n_1}{n_2})^2(1 - \cos^2(\theta_i)} \end{cases} \tag{5.9}$$

Step 4:

The optical path length (OPL) followed by a ray between two interfaces i and $i+1$ is defined as the euclidean length of the line joining these interfaces times the refractive index n of the propagating medium:

$$OPL(i, i+1) = n\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2} \tag{5.10}$$

The two points $P_i(x_i, y_i, z_i)$; $P_{i+1}(x_{i+1}, y_{i+1}, z_{i+1})$ define the intersections of the ray respectively with interfaces i and $i+1$. They are sufficient to define the entire ray between these interfaces.

These four steps are very general and can be applied to any interface of equation $f(x, y, z) = 0$. Only the intersection point and interface normal direction **N** depend on the specific optical layout (lens, MLDOE).

5.2.2 Resolution in the case of a conic lens

In the case of a conic lens of conic constant k, whose interface is defined by Equation 5.5, System 5.4 becomes:

$$\begin{cases} x = \frac{a_i}{b_i}(y - y_0) + x_0 \\ z = \frac{c_i}{b_i}(y - y_0) + z_0 \\ z = \dfrac{r^2}{R + R\sqrt{1 - (1+k)\frac{r^2}{R^2}}} + S \\ r^2 = x^2 + y^2 \end{cases} \tag{5.11}$$

S represents a profile shift along the optical axis z, defining the profile position on the z axis. We simplify the z equation, the symbol $\sqrt{.}$ representing the content of the square root for clarity:

$$\begin{cases} z = \dfrac{r^2}{R + R\sqrt{.}} + S \\ \Leftrightarrow (z-S)R + (z-S)R\sqrt{.} = r^2 \\ \Leftrightarrow \sqrt{.} = \dfrac{r^2}{(z-S)R} - 1 \\ \Rightarrow 1 - (1+k)\dfrac{r^2}{R^2} = \dfrac{r^4}{(z-S)^2R^2} - 2\dfrac{r^2}{(z-S)R} + 1 \\ \Leftrightarrow (1+k)(z-S)^2 - 2R(z-S) + r^2 = 0 \end{cases} \tag{5.12}$$

Two cases occur depending on the values of the direction vector (a_i, b_i, c_i). Note that $c_i \neq 0$ is always true if the rays travel in the Z direction.

Case: $b_i \neq 0$

This case occurs for most of the rays. The expressions of r^2 and z^2 are rather "bulky" y^2 and y polynomials. The second-degree equation to solve is given by:

$$\begin{cases} x = \dfrac{a_i}{b_i}(y - y_0) + x_0 \\ z = \dfrac{c_i}{b_i}(y - y_0) + z_0 \\ Ay^2 + By + C = 0 \\ \Rightarrow A = (1 + \dfrac{a_i^2}{b_i^2}) + (1+k)\dfrac{c_i^2}{b_i^2} \\ \Rightarrow B = -2R\dfrac{c_i}{b_i} + \left(2x_0\dfrac{a_i}{b_i} - 2y_0\dfrac{a_i^2}{b_i^2}\right) + (1+k)\left[2(z_0 - S)\dfrac{c_i}{b_i} - 2y_0\dfrac{c_i^2}{b_i^2}\right] \\ \Rightarrow C = -2R\left[(z_0 - S) - y_0\dfrac{c_i}{b_i}\right] + \left(y_0^2\dfrac{a_i^2}{b_i^2} - 2x_0y_0\dfrac{a_i}{b_i} + x_0^2\right) \\ + (1+k)\left[(z_0 - S)^2 + y_0^2 - 2y_0(z_0 - S)\dfrac{c_i}{b_i}\right] \end{cases} \tag{5.13}$$

Case: $b_i = 0$

In this case, using Systems 5.4 and 5.12, we have a second degree equation in z to solve:

$$\begin{cases} y = y_0 \\ x = \dfrac{a_i}{c_i}(z - z_0) + x_0 \\ (1+k)(z-S)^2 - 2R(z-S) + (x^2 + y_0^2) = 0 \Leftrightarrow Az^2 + Bz + C = 0 \\ \Rightarrow A = 1 + k + \dfrac{a_i^2}{c_i^2} \\ \Rightarrow B = -2S(1+k) - 2R - 2z_0\dfrac{a_i^2}{c_i^2} + 2x_0\dfrac{a_i}{c_i} \\ \Rightarrow C = (1+k)S^2 + 2RS + z_0^2\dfrac{a_i^2}{c_i^2} + y_0^2 + x_0^2 - 2z_0x_0\dfrac{a_i}{c_i} \end{cases} \tag{5.14}$$

Solving the second-degree equation allows precisely defining the intersection point $P(x, y, z)$ between a ray and any conic interface. This analytic solution is fast and efficient since it does not require sampling or numerical resolution.

Finally, we provide the partial derivative expressions of the conic profile necessary to compute the normal (Equation 5.7). Recall that the interface is defined by an equation of the type $f(x, y, z) = 0$.

$$\begin{cases} \dfrac{df}{dx} = \dfrac{x}{R\sqrt{1-(1+k)\dfrac{x^2}{R^2}}} \\ \dfrac{df}{dy} = \dfrac{y}{R\sqrt{1-(1+k)\dfrac{y^2}{R^2}}} \\ \dfrac{df}{dz} = -1 \end{cases} \tag{5.15}$$

5.2.3 Validation in the case of a conic lens

In this section, we validate the implemented software in the case of an F/3 ZnS lens. CodeV optimises the curvature radius and image plane distance, using an aperture diameter $D = 10$ mm. The selected lens is biconvex, with a parabolic shape for simplicity. Two incidence angles are considered: $\theta_x = 0°$ and $\theta_x = 7°$. A macro is defined in CodeV to compute the OPL of 1024*1024 rays inside the aperture (*Real ray trace* function inside a loop) and to record them in a text file. A similar process is done in ASAP (functions *LIST POS* and *IO OUTPUT*). ASAP generates an output file with three columns, recording the ray position (x and y) and OPL at the last interface for each ray. The Matlab function *ScatteredInterpolent* allows to retrieve the OPL map from the 3 vectors (x, y, OPL). CodeV gives a similar output but only provides the length of the ray and its coordinate at each interface. Therefore, the total OPL must be calculated afterwards.

When computing OPLs in CodeV and ASAP, one must be careful to define the source accordingly to step 1: tilting the source for off-axis incidences and defining the *STOP* surface at the lens vertex. The layout generated in Matlab is displayed in Figure 5.3:

The phase delay generated by the ZnS lens is directly computed in Matlab for the three software: analytical ray-trace, ASAP and CodeV, for both on-axis and off-axis incidences. The PSF and MTF at the detector plane result from the Fourier propagation of an initial scalar field, defined by the OPL, following the process described in Figure 4.1. The MTF for each software is shown in Figure 5.4:

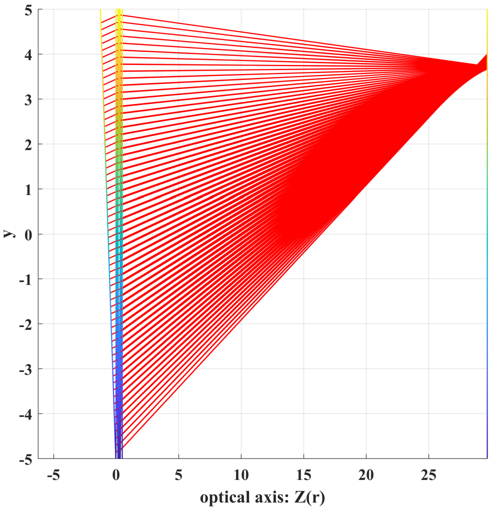


Figure 5.3: 3D ZnS lens layout generated in Matlab. The source plane is tilted to an angle $\theta = 7°$, perpendicular to the beam. The OPL is recorded on the plane interface just after the lens. The focal plane is located at $z = 30$ mm.

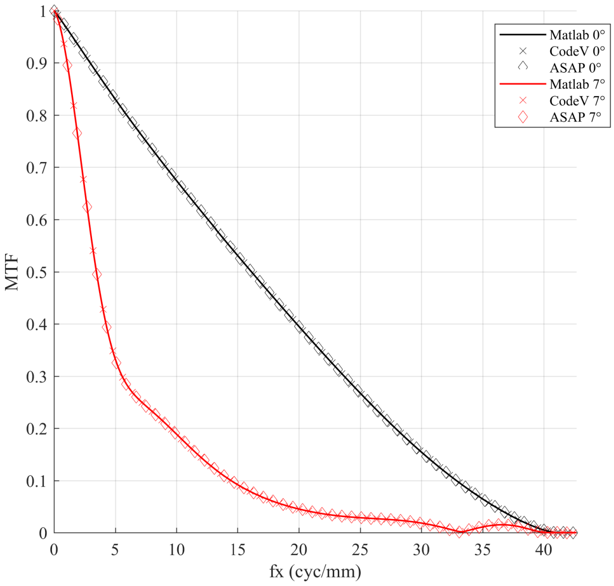


Figure 5.4: On-axis and off-axis MTF for Matlab, CodeV and ASAP OPL calculations. The MTF is retrieved at the focal plane using Fourier optics (see 4.1), based on the OPLs provided by each software. The curves perfectly match, showing that all three software provide the same accuracy.

Figure 5.4 MTF curve is based on the OPL maps recorded by each software (Matlab analytical OPL calculator, CodeV and ASAP). The time needed to obtain such result greatly varies depending on the software:

- CodeV: **45 min** have been necessary to produce the on-axis and off-axis file data, plus **40s** to read the output file and get the OPL map in Matlab. This huge time demand comes from calculating the OPL in CodeV, ray by ray, using a loop and the *real ray trace* function.
- ASAP: **160 s** have been necessary to produce the on-axis and off-axis file data, plus **40s** to read the output file and get the OPL map in Matlab. This high time demand

is caused by the ray trace process, which is quite long for such a high number of rays.

- Matlab analytic engine: **3s** are necessary to provide the full 1024*1024 OPL map for both on-axis and off-axis incidences. This speed comes from the analytic resolution of the ray-tracing process, which is equivalent to adding and multiplying matrices element-wise. Besides, no time or action by the user is required to read an output file.

This performance and time comparative study demonstrates the Matlab analytical calculator potential for OPL calculation, which can be easily used as an optimisation tool.

5.3 Adaptations in the case of MLDOEs

MLDOEs are more challenging than lenses because they are composed of multiple, shifted portions of parabolas. We have demonstrated the MLDOE profile equation for all configurations in Equation 9.40 (Chapter 9. In this chapter, we use the profile equation of the first layer, assumed to be converging, for simplicity. It is rewritten here for clarity:

$$Z(r \in [r_m; r_{m+1}[) = -H\left(m - \frac{r^2}{r_1^2}\right) + S \tag{5.16}$$

According to Equation 9.40, if the zone is diverging, m should be replaced by $m+1$. Similarly, if the second layer is considered instead of the first, reverse the sign of H (whose value also depends on the layer).

This profile equation describes a parabola with a radius of curvature $\mathcal{R} = r_1^2/2H$, shifted by an amount $-Hm + S$ that varies depending on the diffractive zone m. Similarly to the lens case, a shift S is added to place the MLDOE layer at any position $z > 0$ on the optical axis. r_1 is the radius of the central zone and $H(i)$ is the constant height of the diffractive zones, considering the layer $i = 1, 2$. In this section, we consider a more general case, where the height $H = H_m$ and the shift $S = S_m$ vary for each zone, thus also making the curvature radius variable. This general case corresponds to the extended scalar theory optimisation discussed in detail in Chapter 12.

Only two things differ from the ray-tracing algorithm depicted in Section 5.2: The analytical intersection with an MLDOE layer (step 3) and the expression of the interface normal (step 2). Equation 5.16 raises the issue of the diffractive zone knowledge: the equation can be solved for many zones m, but only one gives the correct intersection. Consequently, the correct zone must be found before computing the intersection, which seems paradoxical. The possible intersections between an MLDOE layer and a ray are depicted in Figure 5.5.

The layout depicted in Figure 5.5 provides a simple way to retrieve the right intersection zone. The intersection of the initial point A_I with the plane Z_{II}, located just after the layer, is computed: A_{II}. The right zone is retrieved by comparing the coordinates of A_{II} with the zone radii r_{m-1}, r_m and r_{m+1}. As shown in Figure 5.5, the right intersection zone is not trivial and many cases can occur:

- The ray hits the parabola of the same zone m
- The ray hits the parabola of another zone $m-1$ (or $m+1$ if diverging layer)

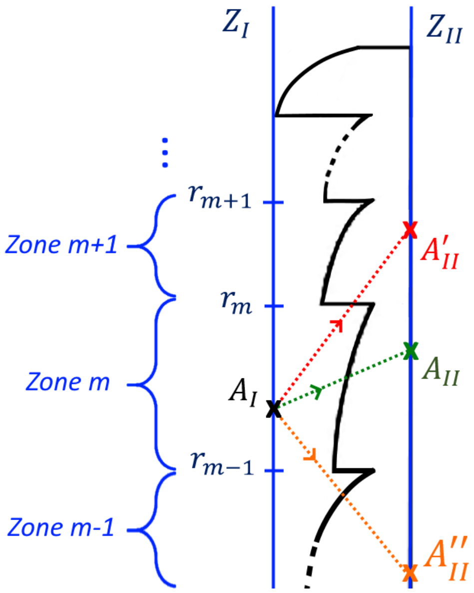


Figure 5.5: Possible intersections between a ray and an MLDOE layer. Depending on the ray position and direction, the intersection can occur: With a parabola of the same zone m: A_{II} (green), with a parabola of another zone $m-1$ (or $m+1$): A'_{II} (red) or with a vertical edge: A''_{II} (orange). The analytical calculator does not handle more complex options, such as multiple refractions on the same zone or total internal reflection. The right zone is given by comparing the coordinates of points A_{II}, A'_{II} and A''_{II} with the zone radii r_{m-1}, r_m and r_{m+1}, before solving the intersection.

- The ray hits a vertical edge between two parabolas at radius r_m (or r_{m-1} if diverging layer)
- The ray is refracted multiple times by a zone or its edge
- The ray is reflected (total internal reflection)

The first three cases provide the value of the correct zone and are handled analytically, which is not the case in the last two cases. These are more complex, thus, very hard to handle mathematically without increasing the OPL calculator complexity. Such rays are called "missed" in the following. The OPL of a missed ray is approximated to its last value (before being marked as "missed"). Recall that this analytical OPL calculator is used as an approximation to describe diffractive optics behaviours; thus, perfect ray tracing is not mandatory. In addition, these situations only appear for a minority of the rays. Thereby, less than 1% of the rays hit another zone (A'_{II} and A''_{II} in Figure 5.5), and only a few of these rays cannot be handled analytically. Therefore, their impact on the final OPL map is negligible, as shown in the following. Step 3 becomes, in the case of an MLDOE:

- Compute the ray intersections with plane z_{II}: A_{II}
- Find the right zone m and the intersecting interface profile type: parabola or $r = cte$ edge.
- Compute the height H_m and the shift S_m for the previously found zone m
- Compute the intersection point between the ray (starting at A_I) and the MLDOE interface

- Compute the derivatives and the normal **N** (step 2) based on the interface profile

The exact intersection point computation (for a parabolic zone or a vertical edge) and the normal derivative determination are detailed in Annex 17 in the case of an MLDOE.

5.4 Validation of MLDOE OPL calculation using ASAP

In this section, we validate the implemented calculator in the case of a ZnS-Air-Ge F/15 MLDOE with 15 diffractive zones. The selected three-dimensional MLDOE layout is created in ASAP as explained hereafter in Section 5.4.1. Section 5.4.2 studies the constant zone height case: H_1 and H_2 are designed following the MLDOE TEA profile design of Equation 9.11. Section 5.4.3 analyses similar figure of merits but considering an MLDOE with variable zone heights $H_1(m)$ and $H_2(m)$ (See the MLDOE extended scalar theory profile design in Equations 12.4 and 12.5). For each case, two incidence angles are considered: $\theta_x = 0°$ and $\theta_x = 5°$.

Apart from the layout, we use the same source definition, OPL recording functions and Matlab reading function as for the refractive lens case, ensuring the validity of the ASAP output.

5.4.1 ASAP 3D MLDOE layout creation

Without going into too many script details, this section describes the main steps to creating a three-dimensional MLDOE layout in ASAP to perform a rigorous OPL calculation. For visualisation purposes, only in the presented layout of Figure 5.7 the number of zones has been decreased to 7 (F/20).

First, we recall to the reader the main MLDOE design parameters:

- F-number ($F/\sharp$), aperture diameter D and number of diffractive zones N. These parameters are linked by Equation 9.31, rewritten here for clarity:

$$N = \frac{D}{8\lambda F/\sharp_\lambda} \tag{5.17}$$

- The design wavelengths are $\lambda_1 = 4.7$ µm and $\lambda_2 = 10.4$ µm according to Section 9.2.1.
- Layer 1, 2 and the gap can be made of any IR material. Their indices are respectively named n_1, n_2 and n_g. The right MLDOE configuration must be selected as described in Section 9.5.
- The diffractive zones m are annular rings of parabolic shape and thickness $H_i(m)$, defined by Equation 9.11. They are defined by a curvature radius $\mathcal{R}_m$, a radius $X_{II}(m)$ and a central obscuration $X_I(m)$. Using Relations 9.1 and 9.11, we have:

$$\begin{cases} X_{II}(m>1) = \sqrt{m X_{II}(1)} \\ X_I(m>1) = X_{II}(m-1) \\ X_I(1) = 0 \\ X_{II}(1) = \sqrt{2\lambda_1 f(\lambda_1)} \\ \mathcal{R}_m = \dfrac{X_{II}^2(1)}{2H_i(m)} \end{cases} \tag{5.18}$$

Note that the sign of the curvature radius $\mathcal{R}_m$ depends on the layer configuration and must be adapted in Equation 5.18.

- If the MLDOE is designed according to the thin element approximation (Chapter 9), the thickness of layer i: H_i is constant for the entire layer, and given by Equation 9.11.
- On the contrary, if the thin element approximation is not considered, the layer thickness $H_i(m)$ can vary for each zone m (Chapter 12).

Based on the previous design parameters, The parametric creation of a three-dimensional MLDOE in ASAP follows multiple steps, described in Figure 5.6:

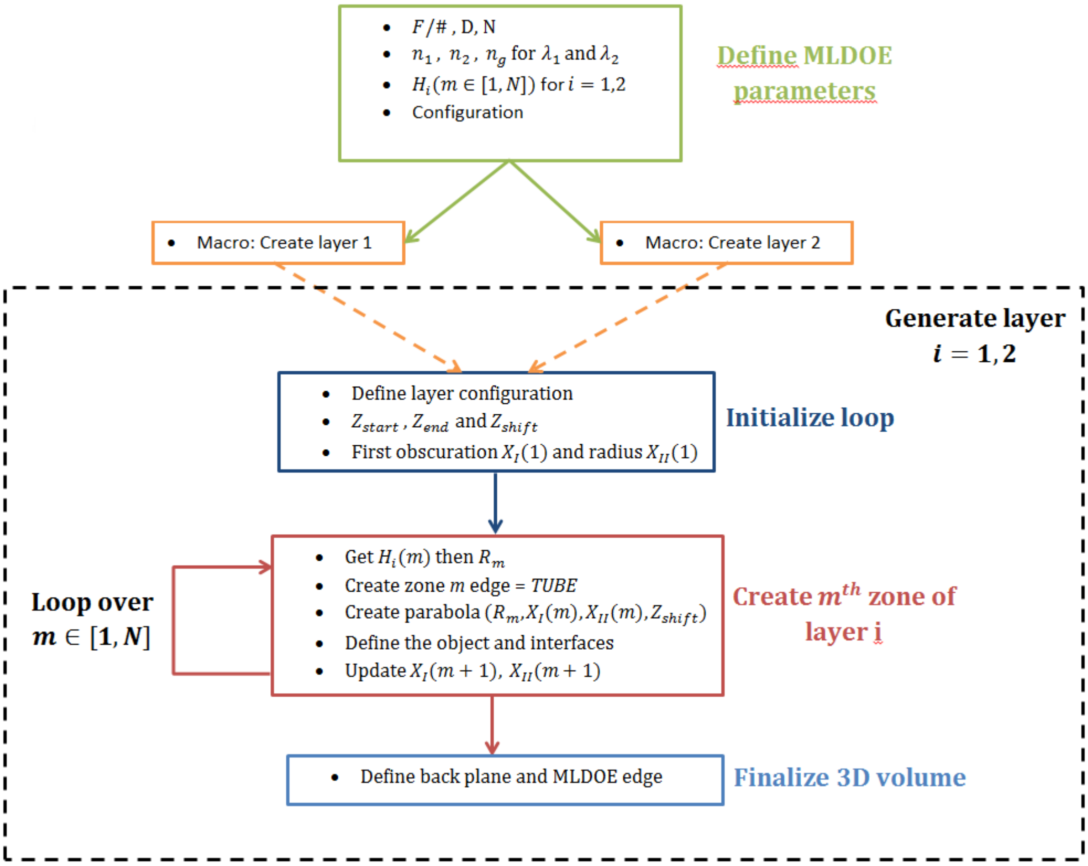


Figure 5.6: 3D MLDOE layout creation process in ASAP. A macro is defined to generate an HDOE layer. Each zone is an annular parabolic ring, defined by a curvature radius $\mathcal{R}_m$, a radius $X_{II}(m)$ and a central obscuration $X_I(m)$. A shift along Z: Z_{shift} is necessary to obtain a Fresnel zone pattern, (See for instance Figure 6.72). Finally, a closed volume is generated by adding a backplane and an edge to the MLDOE

The resulting three-dimensional MLDOE is displayed in Figure 5.7, based on the ASAP visualization panel.

5.4.2 Results for constant heights

As previously mentioned, the time needed to generate the MLDOE 1024*1024 OPL map in Matlab, using the ASAP software, is around 200 s, while it takes only a few seconds for the analytical calculator. After generating the OPL maps, similarly to the lens case, the

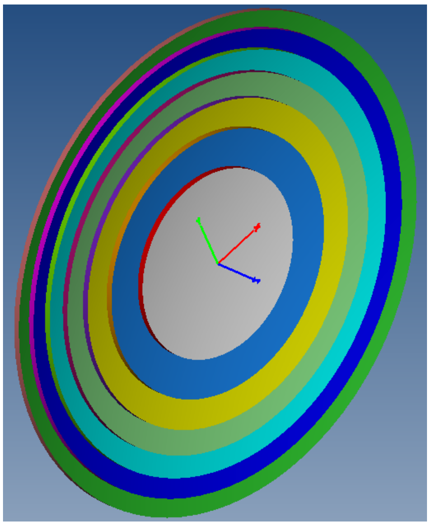


Figure 5.7: 3D ZnS-Air-Ge MLDOE layout viewed in ASAP. We have hidden the Ge layer to see the ZnS layer layout. Only seven diffractive zones are considered (F/20 design) for visualisation purposes. The displayed layer is diverging, as shown by the orientation of the parabolas. A tube along the z direction makes the link between two adjacent zones, as explained in Figure 5.6 process.

PSF and MTF are provided by the Fourier propagation of the phase map (initial field) to the detector plane (output field), following the process described in Figure 4.1. The MTF at the best focal plane $z = 110$ mm is shown in Figure 5.8.

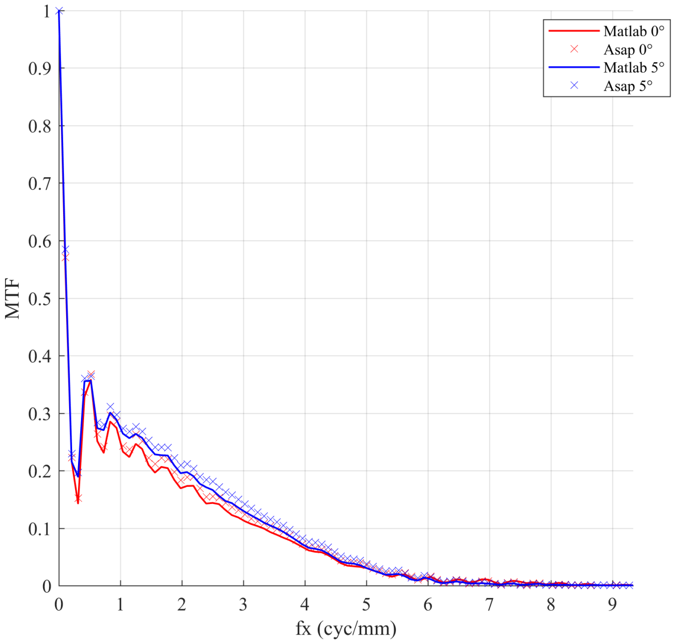
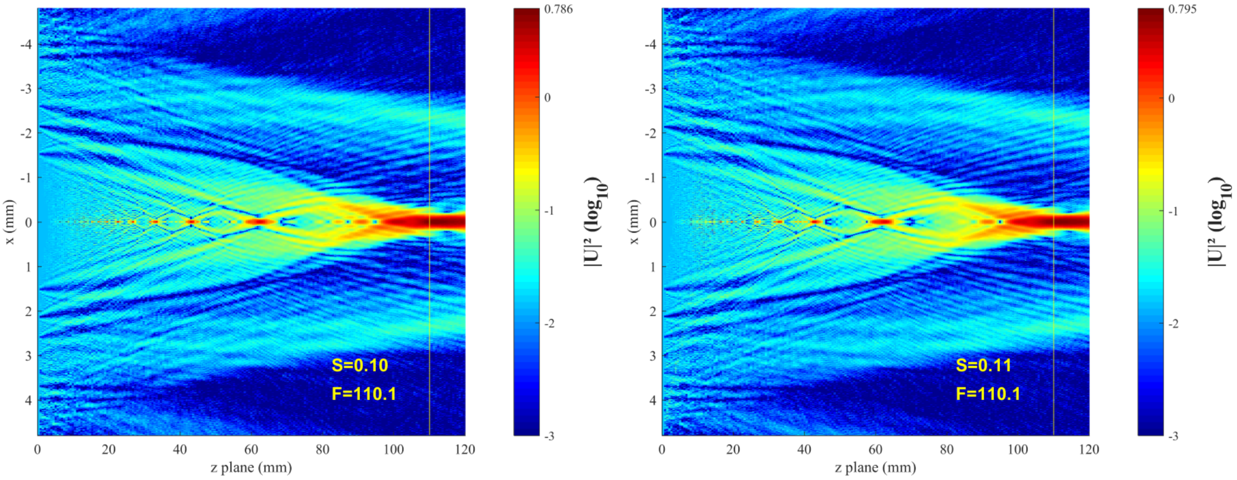


Figure 5.8: On-axis and off-axis MTF comparison between Matlab and ASAP OPL calculations. We consider a ZnS-Air-Ge F/15 MLDOE with constant zone height (Equation 9.11). The MTF is retrieved at the focal plane using Fourier optics (see 4.1), based on the OPLs provided by each software. The curves match excellently, showing that the Matlab OPL calculator is as accurate as it is fast.

The MTF curves, shown in Figure 5.8, provide a near-perfect match between the Matlab analytical calculator and the reference ASAP software. The associated PSF and Strehl ratio values along the optical axis are depicted in Figure 5.9 (on-axis case).



(a) I_z using analytical Matlab calculator.

(b) I_z using exact ASAP software.

Figure 5.9: PSF along the optical axis Z. The MLDOE is located at $z = 0$ mm, while the best focal plane, defined as the plane providing the highest Strehl ratio (S), is located at $z = 110$ mm. The PSF is displayed on a logarithmic scale to emphasise the beam details. Two software are considered: a) Matlab analytic OPL engine. b) ASAP ray-tracing engine.

The irradiance patterns shown in Figures 5.9a and 5.9b are similar, depicting the same "best" focus and very close Strehl ratios. Interestingly, this MLDOE design is not optimal, as displayed by the MTF and irradiance. The extreme zones do not contribute to the focus energy, explaining the low Strehl ratio value. This effect is caused by the high ratio between height and period for the extreme zones. Therefore, this design is a worst-case ray tracing scenario.

5.4.3 Results for variable heights

The optical performances obtained in Figures 5.8 and 5.9 can be enhanced by optimizing the height of each zone instead of having a constant value. This process is detailed in Chapter 12, and the optimised zone results are used in this section to compare the output results of the Matlab analytical OPL calculator and ASAP software. The MTF comparative study is displayed in Figure 5.10:

The irradiance, computed along axis for on-axis incidence, is pictured in Figure 5.11:

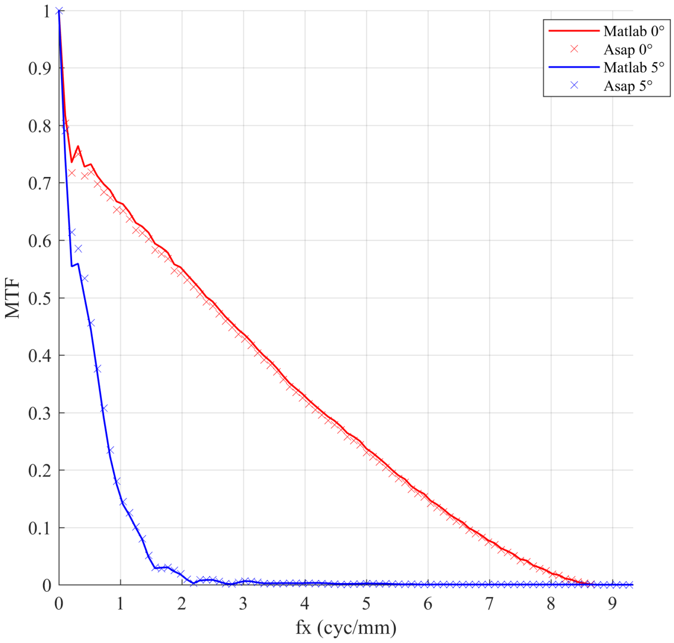
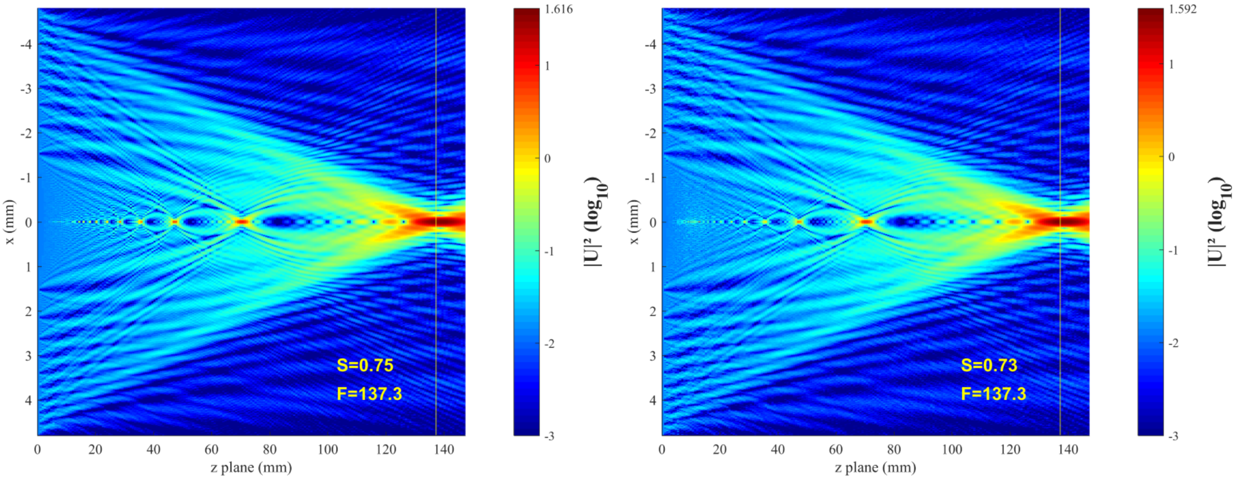


Figure 5.10: On-axis and off-axis MTF comparison between Matlab and ASAP OPL calculations. We consider a ZnS-Air-Ge F/15 MLDOE with variable zone height (Equations 12.4 and 12.5). The MTF is retrieved at the focal plane using Fourier optics (see 4.1), based on the OPLs provided by each software. The curves match excellently, showing that the Matlab calculator is perfectly accurate.

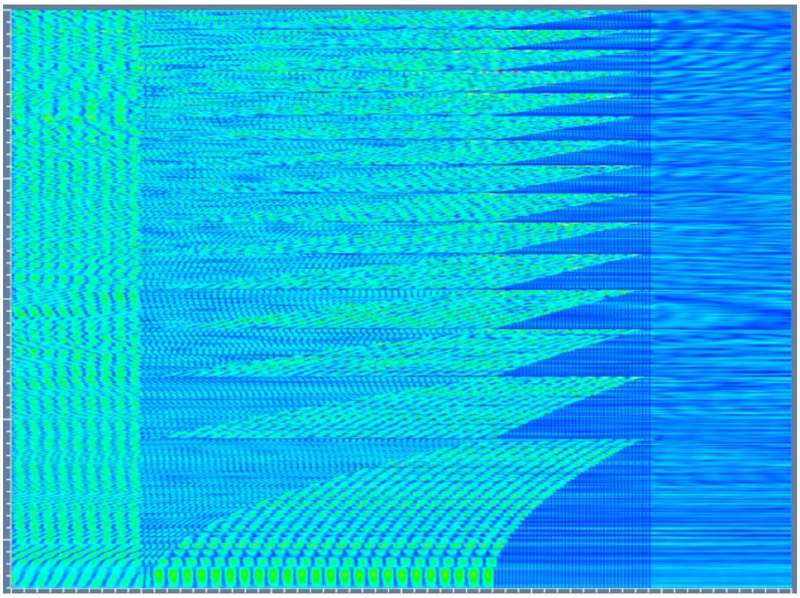


(a) I_z using analytical Matlab calculator.

(b) I_z using exact ASAP software.

Figure 5.11: PSF along the optical axis Z. The MLDOE is located at $z = 0$ mm, while the best focal plane, defined as the plane providing the highest Strehl ratio (S), is located at $z = 110$ mm. The PSF is displayed on a logarithmic scale to emphasise the beam details. Two software are considered: a) Matlab analytical OPL calculator. b) ASAP ray-tracing engine.

Resulting of this analysis, our Matlab OPL calculator gives perfectly accurate results but uses a much faster algorithm.



Chapter 6

Finite difference time domain method

This chapter concludes the tryptic about defining MLDOE modelling tools. In Chapter 4, we have implemented an exact free space wave propagator to retrieve various optical metrics at the detector plane. Chapter 5 explains and justifies the development of an analytical OPL calculator, extensively used to estimate the MLDOE internal behaviour. However, when high accuracy is needed, we use the Finite Differences Time Domain (FDTD) numerical approach instead of the OPL calculator. This numerical method offers precise near-field MLDOE modelling at a high computational effort cost. This method's numerical results are considered a reference in this thesis since. Thereby, we perform a worst-case sampling study to ensure the validity of FDTD predictions for every MLDOE layout and wavelength. To our knowledge, only direct experimentation on an MLDOE prototype can provide more trustable results. This chapter also explains the two-dimensional MLDOE layout creation in the OptiFDTD software.

Contents

6.1 FDTD principle

This section describes the formulation of the rigorous numerical method called "Finite Differences Time Domain" (FDTD). It is entirely based on the technical background documentation provided by the specialised company Optiwave for their software OptiFDTD, Version 16.0 RC, for Microsoft Windows 10 64-bit [43, 44].

6.1.1 2D FDTD equations

The FDTD approach numerically solves the time-dependent Maxwell's curl equations. The lossless and source-free expressions are:

$$\begin{cases} \frac{\partial H_x}{\partial t} = \frac{1}{\mu_0}[\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y}] \\ \frac{\partial H_y}{\partial t} = \frac{1}{\mu_0}[\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z}] \\ \frac{\partial H_z}{\partial t} = \frac{1}{\mu_0}[\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x}] \\ \frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon}[\frac{\partial H_y}{\partial z} - \frac{\partial H_z}{\partial y}] \\ \frac{\partial E_y}{\partial t} = -\frac{1}{\epsilon}[\frac{\partial H_z}{\partial x} - \frac{\partial H_x}{\partial z}] \\ \frac{\partial E_z}{\partial t} = -\frac{1}{\epsilon}[\frac{\partial H_x}{\partial y} - \frac{\partial H_y}{\partial x}] \end{cases} \tag{6.1}$$

The optical component lies in the XZ plane in two dimensions (infinite along the Y axis), and the propagation is along the Z axis. The two-dimensional domain removes any term containing a spatial derivative in Y, partially uncoupling the six Maxwell's equations. Equations 6.1 can be separated in two groups: (E_x, E_z, H_y) is the TM group while (E_y, H_z, H_x) forms the TE group. By convention, the transverse electric (TE) group goes with the electric wave equation along the axis perpendicular to the direction of propagation. In contrast, the Transverse Magnetic (TM) designation goes to the group with the magnetic wave equation along y. TE and TM waves are treated independently, meaning that any 2D FDTD simulation is either a TE or TM simulation (i.e. the full unpolarised field requires two simulations).

Each field is represented in a 2D array of indices i, k: [$E_x(i,k)$, $E_y(i,k)$, $E_z(i,k)$, $H_x(i,k)$, $H_y(i,k)$, $H_z(i,k)$]. The indices i and k respectively account for the number of space steps in the X and Z directions. Figure 6.1 shows the location of the field components in the mesh i, k in the case of TE wave:

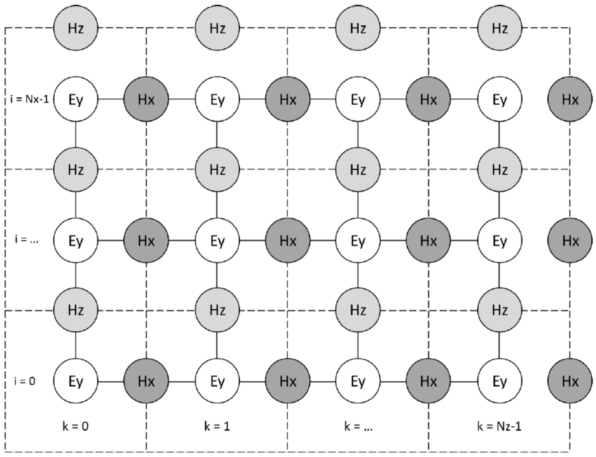


Figure 6.1: Location of the TE field components in the computational domain. The distribution of the TE fields is based upon the two-dimensional Yee Cell. The E_y field is the centre of the spatial cells defined by the dashed lines. The magnetic fields, H_x and H_z, are associated with cell edges. The fields are offset so that the derivatives within Maxwell's equations are discretised and expressed using central finite differences, ensuring second-order accuracy in both space and time [44].

TM waves work similarly to Figure 6.1, with the H_y field in the centre of the cell and both E_x, and E_z fields on the edges. In the case of TE waves, Maxwell's equations are discretised along the grid:

$$\begin{cases} \rightarrow E_y^n(i,k) = E_y^n(i,k) + \dfrac{\Delta_t}{\epsilon\Delta_z}[H_x^{n-\frac{1}{2}}(i,k+\frac{1}{2}) - H_x^{n-\frac{1}{2}}(i,k-\frac{1}{2})] - \\ \dfrac{\Delta_t}{\epsilon\Delta_x}[H_z^{n-\frac{1}{2}}(i+\frac{1}{2},k) - H_z^{n-\frac{1}{2}}(i-\frac{1}{2},k)] \\ \rightarrow H_x^{n+\frac{1}{2}}(i,k+\frac{1}{2}) = H_x^{n-\frac{1}{2}}(i,k+\frac{1}{2}) + \dfrac{\Delta_t}{\mu_0\Delta_z}[E_y^n(i,k+1) - E_y^n(i,k)] \\ \rightarrow H_z^{n+\frac{1}{2}}(i+\frac{1}{2},k) = H_z^{n-\frac{1}{2}}(i+\frac{1}{2},k) - \dfrac{\Delta_t}{\mu_0\Delta_z}[E_y^n(i+1,k) - E_y^n(i,k)] \end{cases} \tag{6.2}$$

where

- n is the discretized time step
- Δ_x is the size in real units of a space step along the X direction
- Δ_z is the size in real units of a space step along the Z direction
- Δ_t is the size in real units of a time step

Section 6.3 describes the constraints and implications on the time and spatial step choice.

6.1.2 Three-dimensional FDTD equations

In three dimensions, each field components is represented by a 3D array of indices i, k: $[E_x(i,j,k)$, $E_y(i,j,k)$, $E_z(i,j,k)$, $H_x(i,j,k)$, $H_y(i,j,k)$, $H_z(i,j,k)]$. This array, composed of so-called three-dimensional "Yee cells", is depicted in Figure 6.2

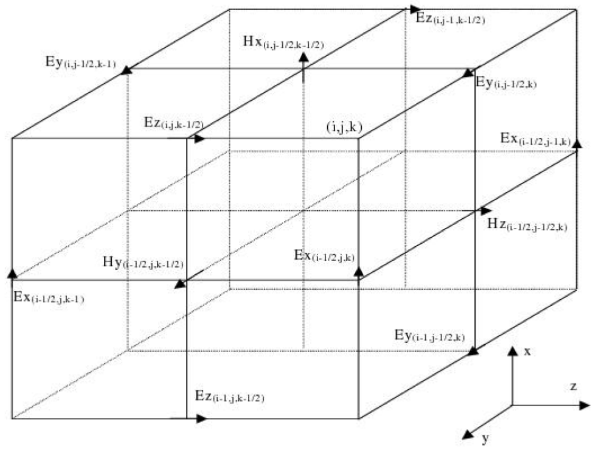


Figure 6.2: Complete three dimensional Yee cell [44]. **E** and **H** are spaced 1/2 step in both space and time and interleaved such that central difference expressions for components have the required information. In this algorithm, the **E** and **H** fields are solved in a leapfrog fashion where **E** fields are calculated using **H** from the previous half-time step and then vice versa.

The source-free time domain vectorial Maxwell's equations are given in differential form by:

$$\begin{cases} \mu \dfrac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E} \\ \epsilon \dfrac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} = \nabla \times \mathbf{H} \end{cases} \tag{6.3}$$

These equations can be discretized on the three-dimensional Yee cell, with recursive definition given in [44], similarly to Equations 6.2.

6.1.3 Boundary conditions

FDTD can only simulate and compute a vectorial field inside a wafer enclosed by physical boundaries. Fields, such as plane waves, cannot extend infinitely and must be studied inside a finite domain whose size dramatically impacts the sampling. The FDTD method proposes a set of boundary conditions, providing the required information to update the EM field components at the edges of the simulation domain. Three boundary conditions are used in this study:

- Anisotropic Perfectly Matched Layer (APML)
- Perfect Electric Conductor (PEC)
- Perfect Magnetic Conductor (PMC)

These boundaries guarantee the field's behaviour when it reaches the edge of the computational window: absorption and reflection of its components. In two dimensions, the simulation space is a rectangle, bounded on the top ($+x$), bottom ($-x$), left ($-z$) and right ($+z$). Waves travel in the Z direction.

The $-z$ and $+z$ sides are set as APML boundaries: the refractive index within the simulation domain is matched, reducing reflections, but a loss is introduced through the layer so that no field returns to the simulation domain when it is reflected. APML boundaries perfectly simulate unbounded wave propagation, acting as an absorbing boundary. However, unwanted diffraction effects at the edges of the aperture may occur if only the APML boundary is used, especially for plane waves. Figure 6.3 shows the propagation of a TM plane wave (H_y) bounded at all sides by APML condition:

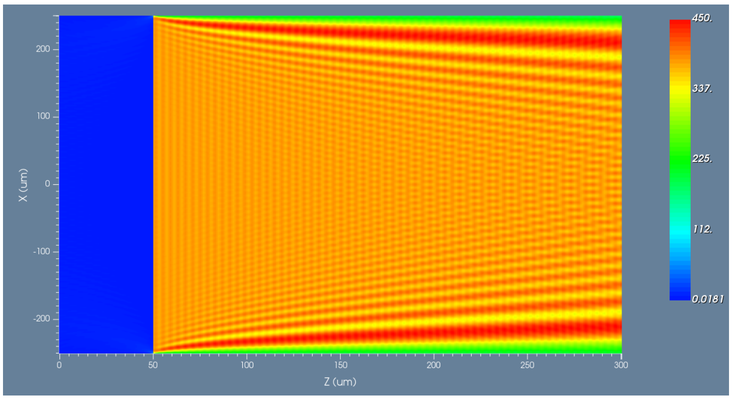


Figure 6.3: H_y component of a TM plane wave propagating in the Z direction. The boundaries ($+x$, $-x$, $+z$, $-z$) are all set with a field absorbing condition: APML. Undesired diffraction effects occur at the edge of the aperture $+x$ and $-x$. Note that no reflection occurs at sides $-z$ and $+z$ due to the APML boundaries.

The $+x$ and $-x$ sides are set to PEC (TM simulation) or PMC (TE simulation) to avoid diffraction by the computational edges, which is not physically meaningful. These boundaries act as mirrors for certain field components and are widely used for plane wave simulations. Thereby, Figure 6.4 depicts how the PEC boundary operates on the **E** and **H** fields:

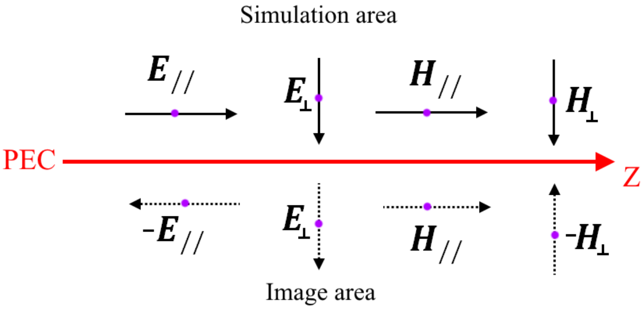


Figure 6.4: Description of the Perfect Electric Conductor (PEC) boundary. The **E** field perpendicular to the boundary can "cross", resulting in the parallel component being null. Similarly, the perpendicular **H** component is null, while the parallel **H** component is the same for each side of the line. When the wave propagation direction is Z, it means that the parallel **H** component does not "see" the boundary, avoiding any interaction (diffraction) with it.

To avoid any diffraction effect, no interaction must occur between the wave and the $+x$ and $-x$ sides. According to Figure 6.4, this is possible for the parallel **H** field using a PEC boundary. In other words, a TM wave suffers no diffraction by the edge with a PEC boundary on the $+x$ and $-x$ sides (similarly, TE waves should be used with a PMC boundary). As a result, Figure 6.5 shows the same plane wave as in Figure 6.3, but with $+x$ and $-x$ PEC boundaries:

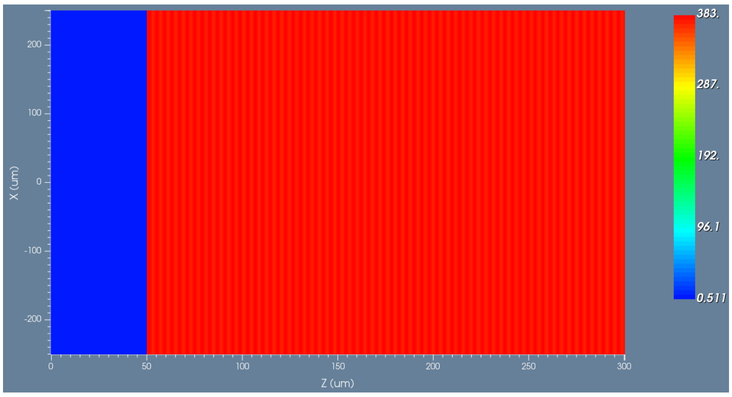


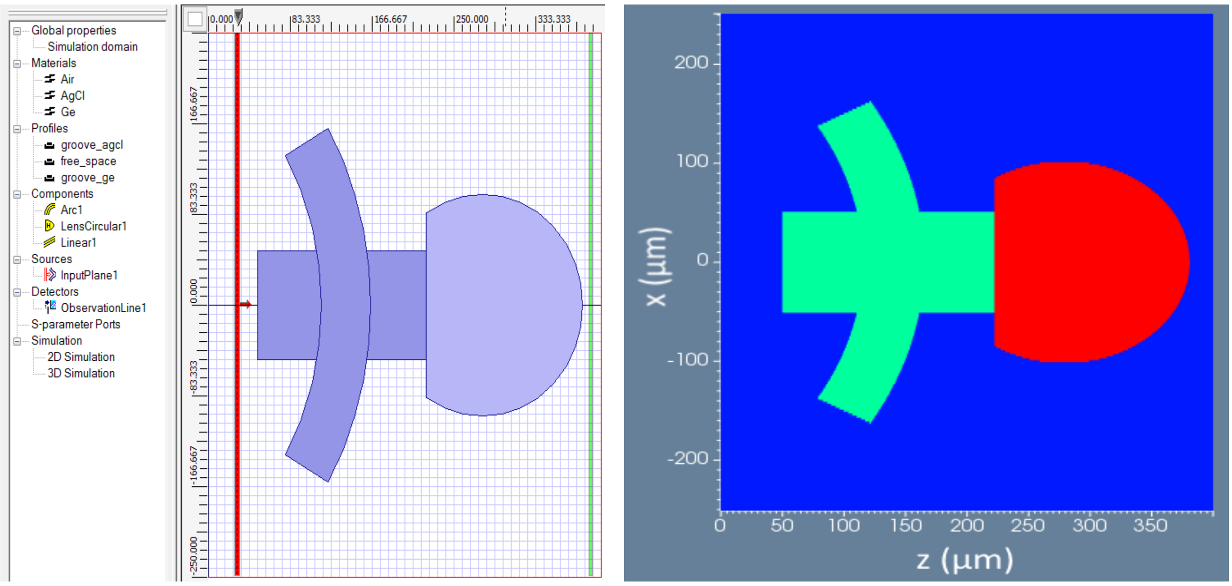
Figure 6.5: H_y component of a TM plane wave propagating in the Z direction. The boundaries $+x$ and $-x$ are set to PEC, while $+z$ and $-z$ are set to APML. No undesired diffraction effects occur at the edge of the aperture $+x$ and $-x$, and the amplitude of the field remains perfectly constant. Note that no reflection occurs at sides $-z$ and $+z$ due to the APML boundaries.

6.2 MLDOE FDTD Layout

6.2.1 Two-dimensional MLDOE layout creation

A complex FDTD layout is constituted of basic polygons such as linear or curved waveguides, elliptic or spherical lenses, cylinders, etc... Each polygon has a specific material. Superposing basic surfaces with specific refractive indices generate complex FDTD layouts. Finally, a source plane is defined (pulse source, Gaussian source, plane wave source) with some observation areas or planes (detectors). Figure 6.6 depicts a hand-made two-dimensional FDTD layout example.

Compared to the simple handmade example of Figure 6.6, more complex layouts often require a script when many built-in surfaces and precise parametrisation are needed. Thereby, we have implemented an MLDOE layout creation function in Matlab, producing an FDTD script that can directly be imported and used in OptiFDTD. MLDOEs are made of two Fresnel HDOEs (See Chapter 9) with parabolic zone shapes. Fortunately, parabolas are part of the available OptiFDTD built-in surfaces. In OptiFDTD, an MLDOE layer is formed by adding shifted and truncated parabolas. As explained in Section 9.5, depending on their materials, MLDOEs can have various layer configurations such as converging-converging, diverging-converging, converging-diverging and diverging-diverging. This configuration must be taken into account when designing the FDTD layout. For instance, we consider an arbitrary converging-converging MLDOE. The main steps leading to its creation are illustrated in Figure 6.7:



(a) FDTD layout made by adding various built-in parametric shapes

(b) Resulting refractive index variation

Figure 6.6: Creation of an arbitrary FDTD layout by adding a linear waveguide, an arc waveguide and a spherical lens. The source and observation planes are displayed respectively in red and green. The various shape profiles and materials are sorted in the left panel of Figure 6.6a. Figure 6.6b shows the two-dimensional refractive index of the layout assembly. The x coordinate represents the radial dimension, while the z coordinate represents the optical axis (light propagation direction).

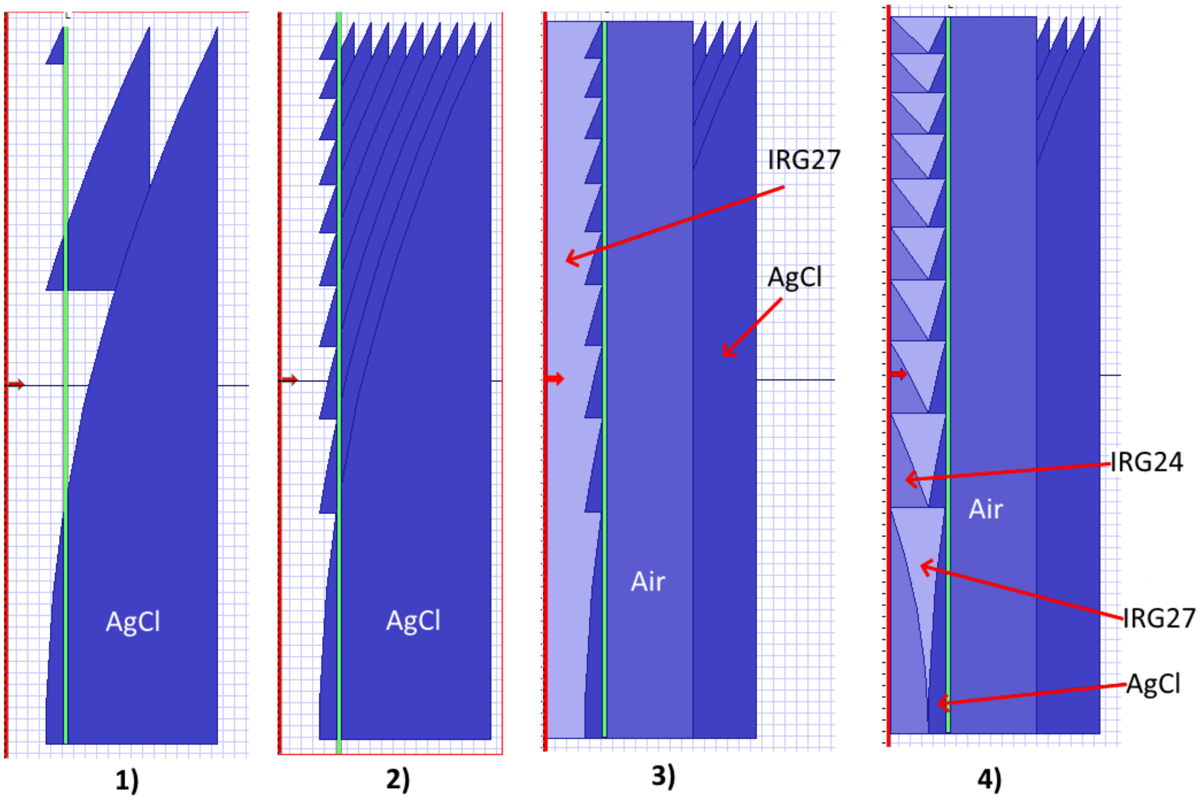


Figure 6.7: Steps leading to creating an arbitrary converging-converging IRG24-IR27-AgCl ML-DOE. The source starts on the red vertical line, whereas the detector lies on the green line. Step 1 (left): parametrise and place truncated parabolas made of AgCl (second layer first) until step 2 is reached. Step 2: Note that the shape of layer 2 appears when taking only the left side of the green observation plane. Notice that, the entire shifted parabolas used to construct layer 2 are still visible. Step 3: Remove the AgCl construction parabolas by adding a linear rectangle filled with air above them. The same principle allows placing the IRG27 gap layer on the left of the AgCl parabolas. Step 4: IRG24 parabolas are added above the IRG27 gap layer using the

Figure 6.7 explains how to create a two-dimensional MLDOE layout in FDTD using parabolas. Any MLDOE of any thickness and material can be created, except for the diverging-diverging configuration. Thereby, following the additive process described in Figure 6.7, A diverging-diverging MLDOE can not be created because the construction parabolas of one layer will always cover the other layer. Note that only a half MLDOE is needed because of its circular symmetry. This symmetry will help with the sampling specification. Finally, the MLDOE refractive index map is displayed in Figure 6.8:

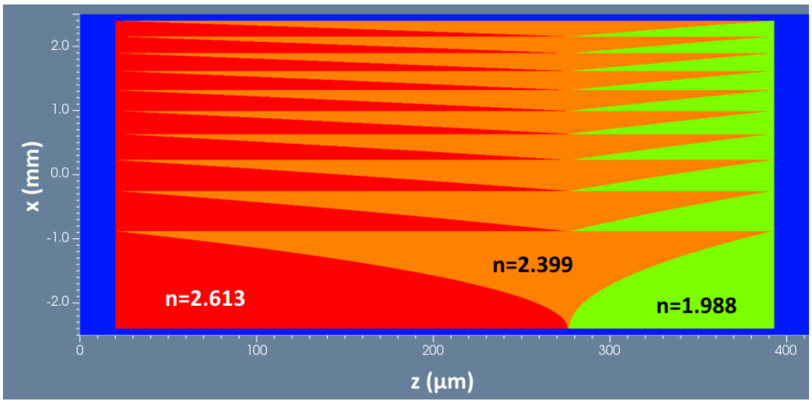


Figure 6.8: Refractive index map considering the arbitrary IRG24-IRG27-AgCl MLDOE layout resulting from Figure 6.7. Only a half MLDOE is required in the x dimension due to circular symmetry, alleviating the sampling demand. The displayed image is not on scale (z in µm and x in mm).

6.2.2 Three-dimensional MLDOE layout creation

Unfortunately, OptiFDTD does not provide as many basic shapes in three dimensions as in two. The built-in three-dimensional components are the sphere, the ellipsoid, the cylinder and the block. Accordingly, previously used two-dimensional truncated parabolas cannot be transposed into three-dimensional objects (only extruded in the Y dimension). Nonetheless, OptiFDTD allows importing a *.IGES* file (defining a volume) and enables associating it with a material. We generate three-dimensional *.IGES* files using the optical modeling software ASAP (See Figures 5.6 and 5.7). Each *.IGES* file defines an "empty" layer as a closed volume.

The importation and handling of *.IGES* files is pretty simple in OptiFDTD. Each layer can be scaled to the desired dimensions and filled with a material. A final step is required if the gap material differs from the background material. The separation between two imported volumes is not a volume and is filled by the ambient medium. Consequently, a three-dimensional cylinder volume, filled with the gap material, is added between the two layers. It is then defined as 1^{st} volume in the hierarchy ("front component"). The addition of volumes, similarly to the addition of surfaces (Figure 6.7), allows to create complex three-dimensional layouts.

6.3 FDTD sampling study

There are two fundamental constraints to the FDTD method regarding the simulation space discretisation. The first concerns the spatial step size (Δ_x, Δ_y, Δ_z), impacting the simulation accuracy. The second is the time step size (Δ_t), defining the simulation stability. This section discusses the step size constraints and the rules of thumb to apply.

A large spatial step size (coarse sampling grid) affects the simulation results in two ways:

- The resolution is not fine enough to properly characterise either the source or the structure under test (sub-wavelength feature, for instance).
- Numerical dispersion can be introduced resulting from the derivatives discretisation.

As a rule of thumb, the source is properly represented when the spatial steps Δ_x, Δ_y and Δ_z are set to one-tenth of the simulation wavelength [44]. Note that the wavelength is taken inside the medium of highest refractive index. This resolution might be increased if the layout is made of complex sub-wavelength features. This criteria defines the so-called "*auto settings*" and is written as:

$$\max\left(\Delta_x, \Delta_y, \Delta_z\right) \leq \frac{\lambda_{min}}{10 n_{max}} \tag{6.4}$$

where n_{max} is the maximum refractive index value in the computational domain. If the layout contains sub-wavelength structures, it is good practice to set the spatial resolution to one-tenth of the smallest feature size.

Sizing the time step according to the Courant-Friedrichs-Levy (CFL) condition ensures the numerical stability of the solution. This condition defines the "*auto setting*" in time for OptiFDTD simulations. The CFL condition is expressed as [44]:

$$\Delta_t \leq \frac{1}{\nu\sqrt{\dfrac{1}{\Delta_x^2} + \dfrac{1}{\Delta_y^2} + \dfrac{1}{\Delta_z^2}}} \tag{6.5}$$

where ν is the speed of the light inside the medium of the highest refractive index. For two-dimensional simulations, the term Δ_y in Equations 6.4 and 6.5 is obviously set to 0.

Convergence testing in space and time is always required to guarantee the validity of the results. In the frame of the MLDOE study using FDTD, convergence testing is not a straightforward process because the output field produced by OptiFDTD corresponds to the near-field. A convergence study process is developed as shown in Figure 6.9.

The convergence study described in Figure 6.9 leads to Figure 6.10.

Resulting from Figure 6.10, the sampling that is sufficient to ensure convergence and stability in the worst case is:

- $N_x = 20000$ samples ($\Delta_x = \dfrac{D}{2N_x} = 0.25$ µm)
- $N_y = N_x$ ($\Delta_y = \Delta_x$) in 3D because of the circular symmetry.
- $N_t = 5000$ samples
- Δ_z and Δ_t : "*auto*" setting. Their value therefore depends on the selected MLDOE layout, accordingly to Equations 6.4 and 6.5.

This study corresponds to a worst case MLDOE design, meaning that it remains valid regardless of the simulated MLDOE, as long as the F-number $F/\sharp \geq 10$, the aperture diameter $D \leq 10$ mm and the minimal wavelength $\lambda_{min} \geq 4.4$ µm.

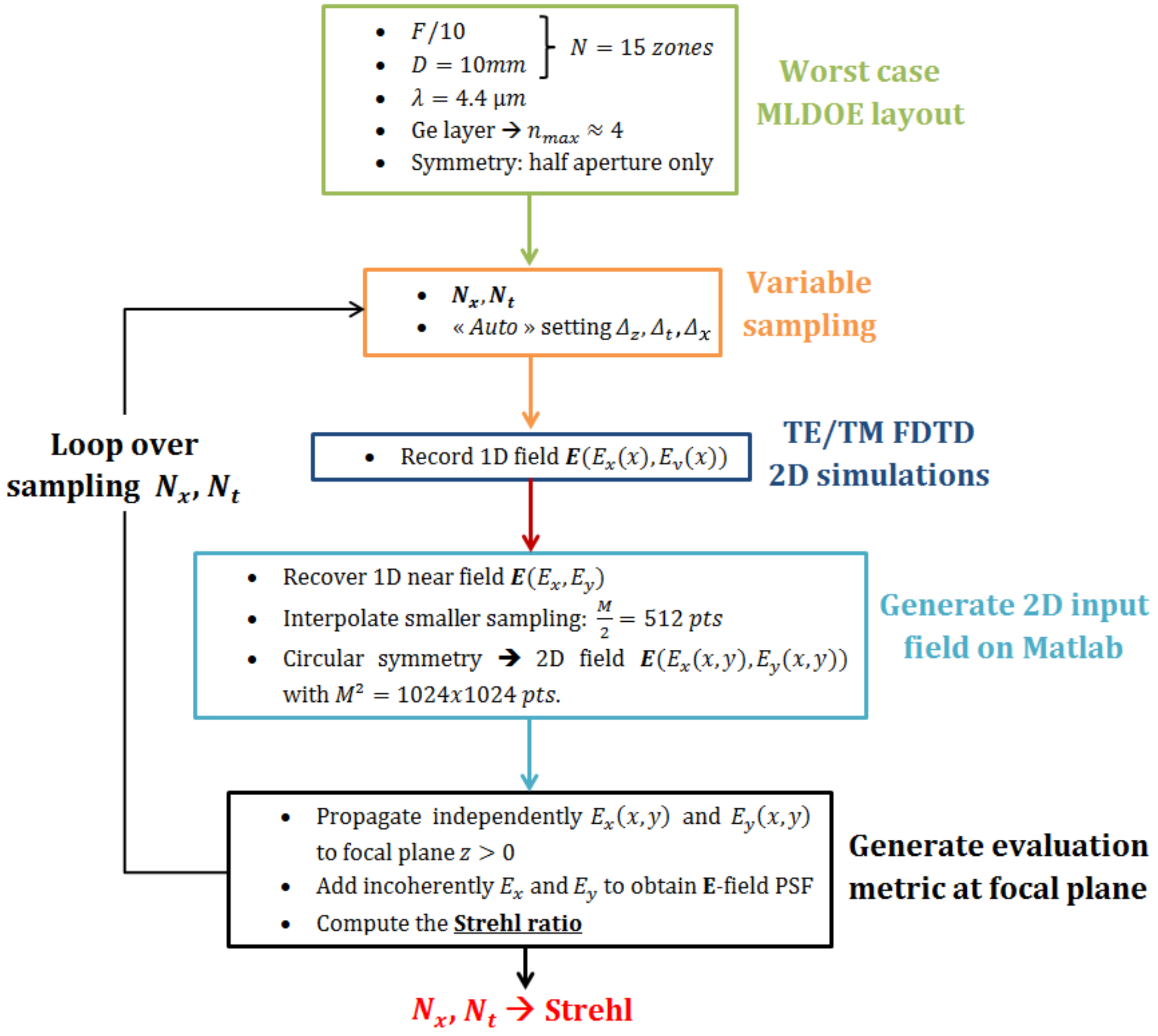


Figure 6.9: FDTD convergence study. A worst-case MLDOE layout is created (low F-number and wavelength) with a Ge layer (highest refractive index). Only half the aperture is needed by symmetry. The number of spatial and time steps N_x, and N_t are variables. The steps Δ_z, Δ_x and Δ_t are set as "*auto*", as defined in Equations 6.4 and 6.5. The one-dimensional near-field is then computed (TE gives $E_y(x)$ and TM gives $E_x(x)$). the E_z component is negligible along the propagation direction (evanescent). Matlab is then used to propagate $(E_y(x), E_x(x))$ to the focal plane (Fourier optics) and generate the Strehl ratio, used as a convergence metric. Since Fourier optics only operates with two-dimensional fields, we use the circular symmetry to extend $(E_y(x), E_x(x))$ in two dimensions.

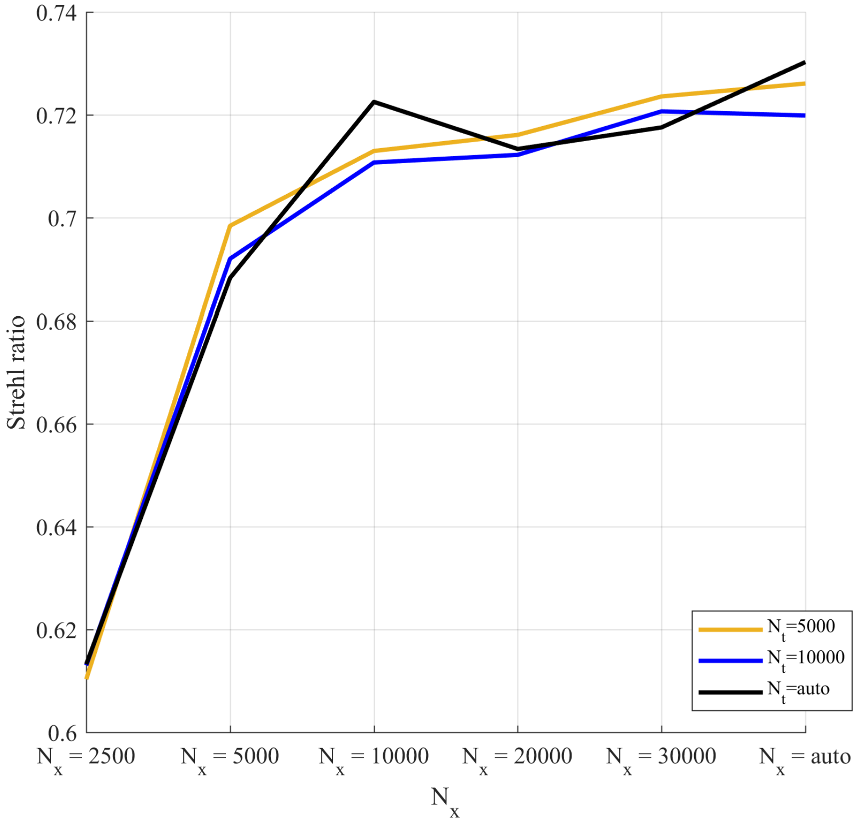
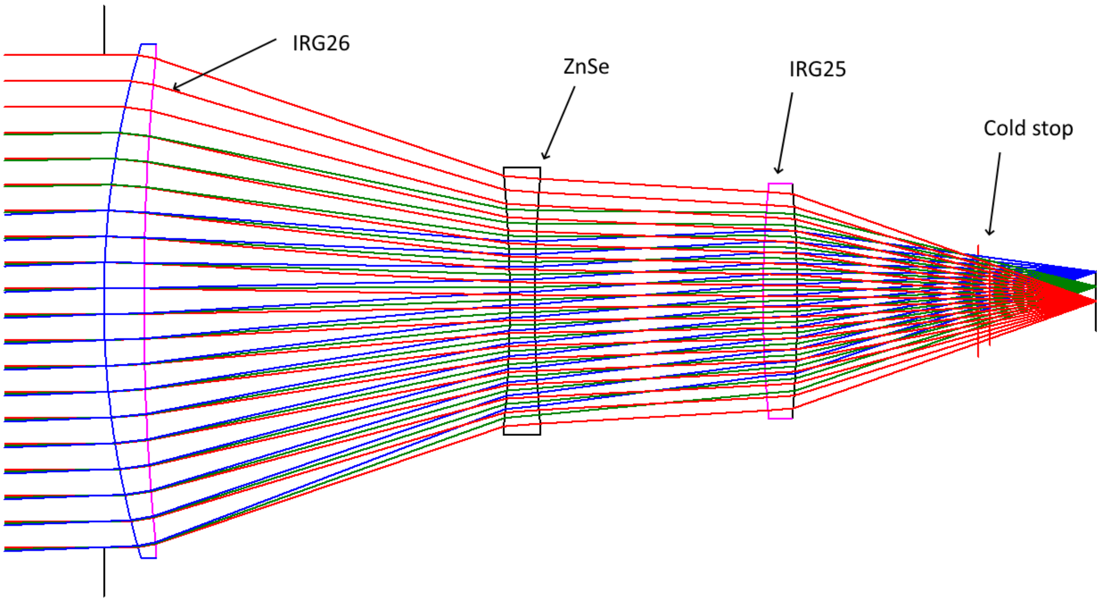


Figure 6.10: FDTD sampling convergence curve for a worst-case ZnS-Air-Ge MLDOE (see Figure 6.9). The metric associated with each sampling is the Strehl ratio, computed at the "best" focal plane. The number of sampling steps N_z along the optical axis and the sampling steps Δ_z, $Delta_x$ and Δ_t are automatically set by OptiFDTD (CFL condition). N_x and N_t are the number of samples in the radial dimension and in time.

Part III

Part 3: Designing innovative dual-band camera


IRG26
ZnSe
IRG25
Cold stop

Chapter 7

Athermal design of a MWIR camera

Following the optical requirements drawn in Chapter 3, this chapter analyses the designing process of an athermal MWIR camera with diffraction-limited performance. We present the optimised MWIR optical design and raise the critical issue of extending this camera to the LWIR band. In the following chapters, this question leads to studying innovative dual-band diffractive lenses, such as the multilayer diffractive optical element. In this Chapter, We synthesise some literature references to build a passive athermalisation algorithm for air-spaced thin lenses (diffractive or refractive). The so-called "achromat athermal system" is impractical to solve directly since it heavily impacts the design performance and does not provide realistic housing materials. Therefore, we perform an iterative study based on the "equivalent lens" principle to minimise image quality loss and ensure the existence of appropriate housing material. The athermal algorithm results in a hybrid MWIR design passively withstanding $\pm 50^{\circ} K$ without any image quality losses.

Contents

In this chapter, we propose to build and optimise an optical system following the requirements established in Chapter 3. The professional optimisation software CodeV guarantees the design specification and minimises the amount of optical aberrations (defocus, spherical, coma, astigmatism, petzval, distortion). In addition to these shape-related aberrations, optical elements intrinsically suffer from chromatic and thermal defocuses due to wavelength and temperature changes. For instance, lens thickness, curvature and refractive index vary as the surrounding temperature is modified. The variation of these optical parameters induces a focal shift called "thermal defocus". The latter significantly degrades the image quality if the detector does not actively move (piezoelectric motor) to compensate for it. Consequently, a passive infrared refractive system, composed of lenses held by a housing material, should follow an optimisation procedure that considers thermal and chromatic variations.

This chapter describes how to build such an athermal optimisation algorithm using various literature references. Then we apply this algorithm using a CodeV handmade macro to design a MWIR athermal achromatic performing hybrid design.

7.1 Thermo-optical parameters

In this section, we introduce the thermo-optical parameters for refractive and diffractive lenses.

We define α_L and α_H as the thermal coefficients of expansion (TCE) of a lens and its housing. When a temperature variation dT occurs, a lens's main optical parameters vary following [45, 46]:

- Radii of curvature R $\rightarrow R + dR = R(1 + \alpha_L dT)$
- Thicknesses D $\rightarrow D + dD = D(1 + \alpha_L dT)$
- Air spaces L $\rightarrow L + dL = L(1 + \alpha_H dT)$
- Refraction index n $\rightarrow n + dn = n + (\frac{dn}{dT} dT)$

This can be rewritten into:

$$\begin{cases} \dfrac{dR}{dT} = R\alpha_L \\ \dfrac{dD}{dT} = D\alpha_L \\ \dfrac{dL}{dT} = L\alpha_H \end{cases} \tag{7.1}$$

α_L, α_H and dn/dT are tabulated values for a lot of materials in a certain range of temperature.

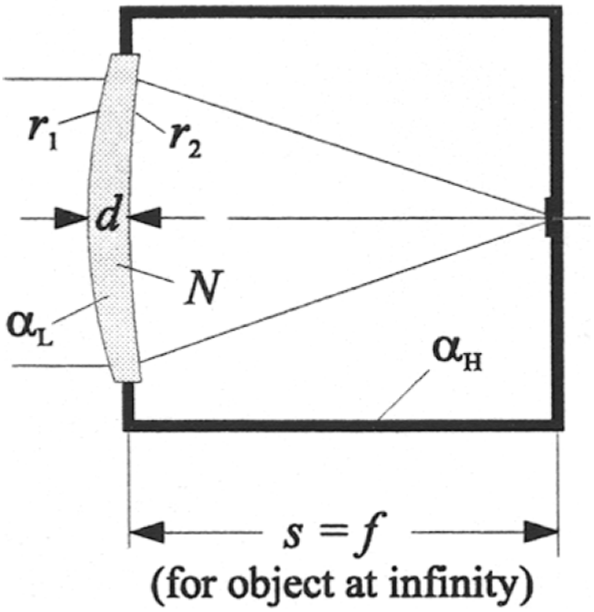


Figure 7.1: Definition of thermally varying parameters for a lens mounted in a housing.

7.1.1 Refractive lens

The optical power Φ of a refractive lens is defined as the inverse of its focal length f:

$$\Phi = \frac{1}{f} \tag{7.2}$$

A trivial derivative relation is given here for any variable X, as it will be useful in the following:

$$\frac{1}{\phi}\frac{\partial \phi}{\partial X} = -\frac{1}{f}\frac{\partial f}{\partial X} \tag{7.3}$$

For a thin refractive lens, the optical power is related to the refractive index n and radii of curvatures R_1 and R_2 by:

$$\Phi = (n-1)(\frac{1}{R_1} - \frac{1}{R_2}) \tag{7.4}$$

The thermal dispersive power γ, sometimes called optomechanical coefficient, is defined as [47, 48]:

$$\gamma = \frac{1}{\phi}\frac{\partial \phi}{\partial T} \tag{7.5}$$

Using Equations 7.3 and 7.4, we obtain:

$$\gamma = \frac{1}{n-1}\frac{\partial n}{\partial T} + \frac{1}{(\frac{1}{R_1} - \frac{1}{R_2})}(-\frac{1}{R_1^2}\frac{dR_1}{dT} + \frac{1}{R_2^2}\frac{dR_2}{dT}) \tag{7.6}$$

Introducing Equation 7.1 in Equation 7.6 leads to the expression of γ:

$$\gamma = \frac{1}{n-1}\frac{\partial n}{\partial T} - \alpha_L \tag{7.7}$$

The chromatic dispersive power ω is defined as [47, 48]:

$$\omega = -\frac{\Delta\phi}{\phi} = -\frac{1}{\Phi}\Delta\lambda\frac{\partial\Phi}{\partial\lambda} \tag{7.8}$$

Using the thin lens model (Equation 7.4) into Equation 7.8 provides the expression of ω:

$$\omega = -\Delta\lambda\frac{\partial n}{\partial\lambda}\frac{1}{n(\lambda)-1} > 0 \tag{7.9}$$

In the following, for all the considered IR materials, the chromatic dispersive power ω is always positive in the case of a thin lens (See Annex 16 and Table 7.1).

7.1.2 Diffractive lens

A diffractive optical element (DOE) is represented in Figure 7.2:

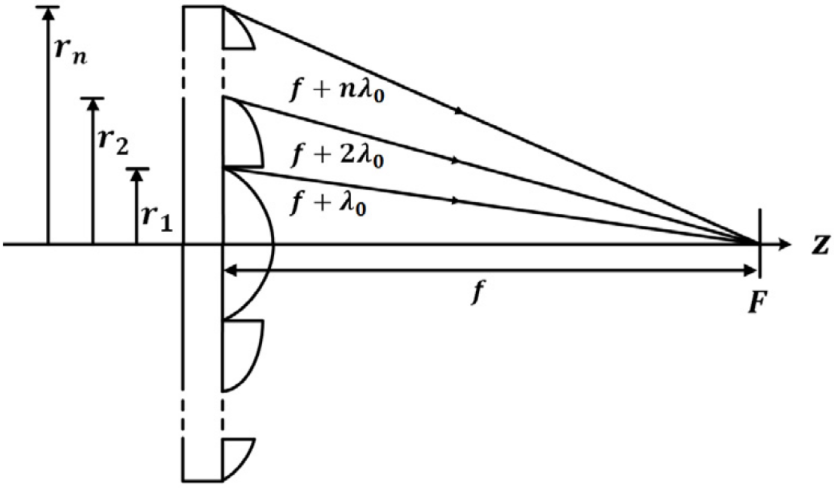


Figure 7.2: Schematic diagram of a diffractive optical element (DOE) [49]. See Chapter 8.

DOEs are deeply studied in Chapter 8. In short, a DOE is made of circular zones, each having a different optical path difference (OPD) as displayed in Figure 7.2. To impose constructive interferences, the OPD between each zone must be increased by λ_0, the "design wavelength". The *nth* zone OPD is given by: $OPD_n = f(\lambda_0) + n\lambda_0$. Applying Pythagorean theorem allows to approximate the *nth* zone radius r_n:

$$r_n^2 = 2n\lambda_0 f(\lambda_0) \tag{7.10}$$

When a DOE is illuminated at a wavelength λ, its focal distance $f(\lambda)$ is modified following [50]:

$$f(\lambda) = f(\lambda_0)\frac{\lambda_0}{\lambda} \tag{7.11}$$

Introducing Equation 7.11 in the definition of the chromatic dispersive power ω (Equation 7.8) leads to:

$$\omega = -\frac{\Delta\Phi}{\Phi} = -\frac{\Phi_0\Delta\lambda}{\lambda_0\Phi_\lambda} = -\frac{\Delta\lambda}{\lambda} < 0 \tag{7.12}$$

$\Delta\lambda = \lambda_L - \lambda_S$, with λ_L and λ_S the longest and the shortest wavelengths of the considered waveband. ω is always negative for a DOE, contrary to refractive thin lenses. This allows very good chromatic correction when DOEs and refractive lenses are combined together (hybrid lens).

Each zone radius r_n varies with the temperature as described in Equation 7.1. Using the definition of the thermal dispersive power γ (Equation 7.5) as well as the nth zone radius expression (Equation 7.10) leads to:

$$\gamma = \frac{1}{\phi}\frac{\partial\phi}{\partial T} = -\frac{1}{f}\frac{\partial f}{\partial T} = \frac{1}{r_n^2} * \frac{d}{dT}[r_n^2] = \frac{-2r_n}{r_n^2} * \frac{d}{dT}[r_n] = -2\alpha_L < 0 \tag{7.13}$$

A DOE thermal power is always negative and entirely defined by the material choice. Since the majority of infrared materials have positive thermal powers (see Table 7.1), Equation 7.13 is extremely useful for passive athermalisation using hybrid lenses.

7.2 IR materials

This section describes a handful of IR materials. Theoretically, the chromatic and thermal powers ω and γ, introduced in Section 7.1, are functions of the temperature and the wavelength. Compared to the visible, the number of IR materials (transparent between 3 µm and 12 µm) is very small. These materials are often expensive semiconductors with different transmission ranges, as shown in Figure 7.3:

Some IR materials such as NaCl, KCl are not considered since they are soluble in water, rendering them difficult to polish. In addition, materials that are too hard and cannot be diamond turned, such as Sapphire, are discarded. Figure 7.3 transmission ranges come from CRYSTRAN website [52] for regular materials and from SCHOTT datasheets [53] for IRG chalcogenide family. The following material descriptions come from [54].

7.2.1 IR materials description

Models of refractive index and transmission over the MWIR and LWIR wavebands are described in Annex 16. In this section, we provide a short description of the selected IR materials.

- Ge (Germanium) has the highest refractive index among all the IR materials and also the highest dn/dT, which makes athermalisation a difficult issue. Its optical transmission starts degrading at 100°C, which is not a problem for space applications. It is usually less expensive than ZnSe and ZnS. It is diamond turnable and can be manufactured as an aspheric or diffractive lens. Its average transmission is only 47% over the MWIR-LWIR wavebands (uncoated).

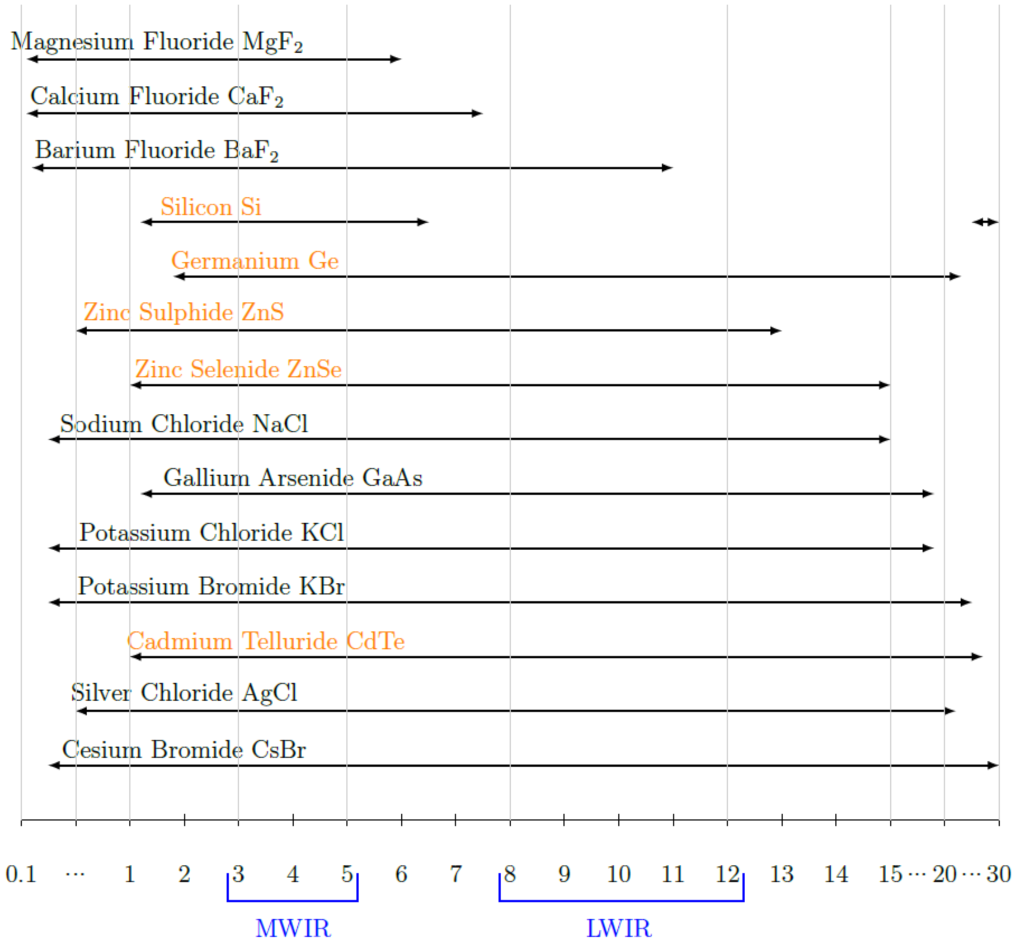


Figure 7.3: Transmission range for IR optical materials. [51].

- ZnSe (zinc Selenide) has a high resistance to thermal shock. Since its hardness is relatively low, coatings can be applied to protect it. ZnSe is very expensive. It is diamond turnable and can be manufactured as an aspheric or a diffractive lens. Its average transmission is 70% over the LWIR-LWIR wavebands.

- ZnS (Zinc Sulphide) is used in two forms: clear grade and regular grade. The clear grade form is purified to transmit homogeneously from the VIS to the LWIR waveband. Therefore, we will use clear grade ZnS, which also has low absorption in MWIR. ZnS clear grade is one-third harder than ZnSe. It is diamond turnable and can be manufactured as an aspheric or a diffractive lens. The clear grade average transmission is 75% until 10 µm and 65% afterwards.

- GaAs (Galium Arsenide) has a very similar hardness to Ge but is generally more expensive. It is commonly used in applications where hardness and durability are essential. It is diamond turnable and can be manufactured as an aspheric or a diffractive lens. It has an average transmission of only 55% over the MWIR-LWIR wavebands.

- CdTe (cadmium Telluride): Extra handling and safety precautions are required when machining CdTe due to its toxicity. Thus, only a few companies provide CdTe lens manufacturing. It has one of the highest densities among the IR materials and is also more expensive than Ge and ZnSe. It is diamond turnable and can be manufactured as an aspheric or a diffractive lens. It has an average transmission of 65% in the MWIR-LWIR wavebands.

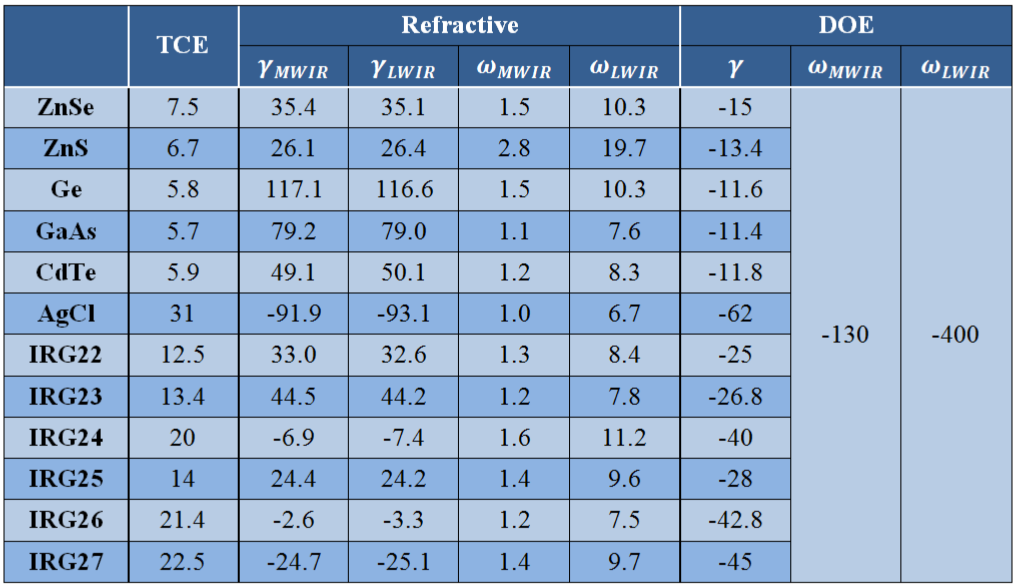
- AgCl (Silver Chloride) can easily be polished due to its low hardness but can also be deformed under heat and pressure. Its average transmission is 80% in the MWIR-LWIR wavebands. It darkens when exposed to sunlight which does not affect its IR transmission.

- Chalcogenides: These materials are composed of Ge, As, Se and Te in various proportions. We only consider chalcogenides from SCHOTT because they are very similar to the six other chalcogenide families, and only SCHOTT datasheets provide their dispersion formulas. Chalcogenides from UMICORE (gasir), VITRON (IG), AMTIR, LightPAth (BD) are equivalent:
    - $Ge_{33}As_{12}Se_{55}$: IRG22, IG2, AMTIR1, GASIR1.
    - $Ge_{30}As_{13}Se_{32}Te_{25}$: IRG23, IG3.
    - $Ge_{10}As_{40}Se_{50}$: IRG24, IG4.
    - $Ge_{28}Sb_{12}Se_{60}$: IRG25, IG5, AMTIR3, BD2.
    - $As_{40}Se_{60}$: IRG26, IG6, AMTIR2, GASIR5.
    - $As_{40}S_{60}$: IRG27, AMTIR6.

  Compared to other IR materials, Chalcogenides have low refractive indices, low dn/dT and ω. They are polishable, diamond turnable and moldable, potentially making them "cheaper" than other IR materials for mass production applications. Chalcogenides have $\sim$ 100% average transmission until 11µm, which decreases to 40% afterwards (except for IRG26).

7.2.2 Thermal and chromatic powers approximation

The refractive index of a material depends on the wavelength and, especially for IR materials, on the temperature. Consequently, we provide dispersion formulas $n(\lambda, T)$ for all IR material in Annex 16. The resulting refractive indices have been cross-checked for many λ and T with the reference site *refractiveindex.info*, and the values matched until the thousandth. Consequently, ω and γ also depend on λ and T. Their variations are presented in Annex 16 for all materials. This analysis shows that the chromatic power ω is nearly independent of the temperature. For each IR material, γ varies very little with (λ, T). Consequently, we can provide approximate values of γ and ω for each material in MWIR and LWIR wavebands. The results are synthesized in Table 7.1:



	TCE	Refractive				DOE		
		γ_{MWIR}	γ_{LWIR}	ω_{MWIR}	ω_{LWIR}	γ	ω_{MWIR}	ω_{LWIR}
ZnSe	7.5	35.4	35.1	1.5	10.3	-15	-130	-400
ZnS	6.7	26.1	26.4	2.8	19.7	-13.4		
Ge	5.8	117.1	116.6	1.5	10.3	-11.6		
GaAs	5.7	79.2	79.0	1.1	7.6	-11.4		
CdTe	5.9	49.1	50.1	1.2	8.3	-11.8		
AgCl	31	-91.9	-93.1	1.0	6.7	-62		
IRG22	12.5	33.0	32.6	1.3	8.4	-25		
IRG23	13.4	44.5	44.2	1.2	7.8	-26.8		
IRG24	20	-6.9	-7.4	1.6	11.2	-40		
IRG25	14	24.4	24.2	1.4	9.6	-28		
IRG26	21.4	-2.6	-3.3	1.2	7.5	-42.8		
IRG27	22.5	-24.7	-25.1	1.4	9.7	-45		

Table 7.1: Approximated γ and ω values for all IR materials in the MWIR and LWIR wavebands. Refractive and diffractive cases are displayed (Equations 7.7, 7.9, 7.13 and 7.12). TEC and γ values are expressed in $[10^{-6}.K^{-1}]$ while ω is given in 10^{-3}.

Table 7.1 shows that thermal power γ is nearly constant in both wavebands for each material. On the contrary, the dispersive power ω is higher in LWIR than in MWIR. Thereby, the larger the waveband, the higher the dispersive power. Consequently, the waveband choice only impacts the achromatisation of the design, which is harder to perform in LWIR. Note that diffractive elements have a very high negative chromatic power that does not depend on the material. Therefore, to compensate for much lower refractive positive ω, DOEs must be designed with low optical powers. As for the diffractive thermal power, it only depends on a material's TCE, according to Equation 7.13.

7.3 Design of thin lenses in contact

We have introduced three major parameters in Section 7.1. The optical power Φ defines the "converging ability" of a lens. Under a temperature and wavelength variation, the resulting defocus is characterized by the chromatic and thermal powers (ω and γ). An achromat optical design should provide the required effective focal length with minimal aberrations while keeping its effective ω and γ as low as possible. We have demonstrated that ω and γ are impacted mainly by lens material and focal length.

7.3.1 Single lens case

The dispersive ability of a lens is characterized by its dispersive power ω, defined in Equation 7.9. Note that this definition applies to refractive and diffractive lenses. Only the developed expression depends on the type of lens (see Equations 7.9 and 7.12). The chromatic defocus $\Delta_\lambda f$ can be expressed over the waveband $\Delta\lambda$ for any lens, using Equation 7.8:

$$\begin{cases} \omega = -\dfrac{\Delta\lambda}{\Phi}\dfrac{\partial\Phi}{\partial\lambda} \\ = -f\Delta\lambda\dfrac{\partial 1/f}{\partial\lambda} \\ = \dfrac{\Delta\lambda}{f}\dfrac{\partial f}{\partial\lambda} \\ \Leftrightarrow \dfrac{\partial f}{\partial\lambda} = \dfrac{f\omega}{\Delta\lambda} \\ \Rightarrow \Delta_\lambda f = \omega f \end{cases} \tag{7.14}$$

Similarly for the thermal defocus $\Delta_T f$, using the definition of γ (Equation 7.5):

$$\begin{cases} \gamma = \dfrac{1}{\Phi}\dfrac{\partial\Phi}{\partial T} \\ = \dfrac{1}{f}\dfrac{\partial f}{\partial T} \\ \Leftrightarrow \dfrac{\partial f}{\partial T} = -f\gamma \\ \Rightarrow \Delta_T f = -\gamma f \Delta T \end{cases} \tag{7.15}$$

The thermal variation ΔT will cause a total defocus Δz coming from the intrinsic thermal lens focal shift $\Delta_T f$ and from the housing deformation ΔL. Using Equations 7.1 and 7.15, the total amount of thermal defocus is expressed as [55]:

$$\Delta z = |(\Delta L - \Delta_T f)| = |\Delta T(L\alpha_H + f\gamma)| \tag{7.16}$$

For a single lens and object at infinity, the housing length L is equal to the focal length f as shown in Figure 7.1. We get the total thermal defocus [56]:

$$\Delta z = |f\Delta T(\gamma + \alpha_H)| \tag{7.17}$$

From aberration theory, the depth of focus for a diffraction-limited imaging system (OPD $\lambda/4$) is:

$$\Delta z = |2\lambda(F/\#)^2| \tag{7.18}$$

As a result, the maximal temperature variation ΔT that a single lens (refractive or diffractive) can withstand to keep the same optical performance is given by:

$$\Delta T_{max} = \left|\frac{2\lambda(F/\#)}{D(\gamma + \alpha_H)}\right| \tag{7.19}$$

where $F/\#$ is the F-number of the optical system, defined as $F/\# = f/D$ with D the pupil diameter. As an example, the maximal temperature variations that a Ge lens can withstand at 10 µm are shown in Figure 7.4:

For instance, we consider a 200 mm F/2 Germanium lens in an aluminium housing operating at 10 microns. With $\gamma = 130 \times 10^{-6}$, and $\alpha_H = 23 \times 10^{-6}$, the tolerable temperature difference until the thermal defocus becomes unacceptable is only 2.6°C.

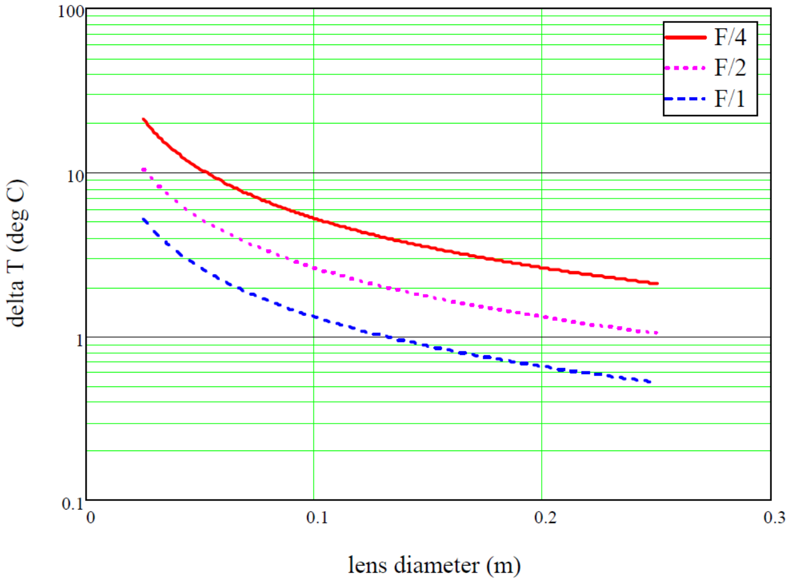


Figure 7.4: Tolerable temperature variations for a thin refractive Ge lens at 10 µm [55]

7.3.2 Case of multiple thin lenses in contact

In this section, we explain how to correct the chromatic and thermal defocuses of j thin refractive or diffractive lenses in contact. Starting with two separated lenses, the Gullstrand formula, demonstrated in [57], links the total optical power ϕ_{tot} to the powers of the two lenses composing the system:

$$\phi_{tot} = \Phi_1 + \Phi_2 - \frac{e}{n}\Phi_1\Phi_2 \tag{7.20}$$

with n the refractive index of the medium between the two lenses, whose principal planes are separated by a distance e. In the case of thin lenses in contact, the formula reduces to $\phi_{tot} = \Phi_1 + \Phi_2$. More generally, for j thin refractive/diffractive lenses in contact, we have:

$$\phi_{tot} = \sum_{i=1}^{j} \Phi_i \tag{7.21}$$

Using Relation 7.21, we can apply the definition of ω (Equation 7.8) to any set of j lenses in contact:

$$\begin{cases} \displaystyle\sum_{i=1}^{j} \frac{\partial \Phi_i}{\partial \lambda} = \sum_{i=1}^{j} (-\Delta\lambda \omega_i \Phi_i) \\ \displaystyle\Leftrightarrow \frac{\partial}{\partial \lambda}\left[\sum_{i=1}^{j} \Phi_i\right] = -\Delta\lambda \sum_{i=1}^{j} \omega_i \Phi_i \\ \displaystyle\Leftrightarrow \frac{\partial \Phi_{tot}}{\partial \lambda}(\lambda) = -\Delta\lambda \sum_{i=1}^{j} \omega_i(\lambda, \Delta\lambda)\Phi_i(\lambda) \end{cases} \tag{7.22}$$

We assume that the lens system is designed at a wavelength λ_1, resulting in an effective optical power $\Phi_{tot}(\lambda_1)$. For another wavelength λ_i, we define the notations $\Delta\lambda_k \mathrel{\widehat{=}} \lambda_k - \lambda_1$

and $\Delta_{\lambda_k}\Phi_{tot} \mathrel{\widehat{=}} \Phi_{tot}(\lambda_k)-\Phi_{tot}(\lambda_1)$. The chromatic aberration can be corrected for $k \in [2, j]$ distinct wavelengths $\lambda_{2..k}$ using Equation 7.22:

$$\begin{cases} \Delta_{\lambda_2}\Phi_{tot} = 0 \Leftrightarrow \sum_{i=1}^{j} \omega_i(\lambda_1, \Delta\lambda_2)\Phi_i(\lambda_1) = 0 \\ \dots \\ \Delta_{\lambda_k}\Phi_{tot} = 0 \Leftrightarrow \sum_{i=1}^{j} \omega_i(\lambda_1, \Delta\lambda_k)\Phi_i(\lambda_1) = 0 \end{cases} \tag{7.23}$$

When $k = 2$, System 7.23 is called an achromat system and allows for correcting the chromatic focal shift for two distinct wavelengths. Note that System 7.23 is valid as long as the lenses are in contact, regardless of the type of lenses. The difference between refractive and diffractive lenses arise from the expression of $\omega_i(\lambda_1, \Delta\lambda)$, given in Equations 7.9 and 7.12. In the case of an achromat doublet ($j = k = 2$), one lens must be diverging ($\Phi_1 < 0$) while the second must be converging ($\Phi_2 > 0$). Both lenses must be made of different materials, preferably with low and high dispersion.

A similar process can be applied to express the thermal defocus of j thin lenses in contact:

$$\begin{cases} \sum_{i=1}^{j} \frac{\partial \Phi_i}{\partial T} = \sum_{i=1}^{j} \gamma_i \Phi_i \\ \Leftrightarrow \frac{\partial \Phi_{tot}}{\partial T} = \sum_{i=1}^{j} \gamma_i(T)\Phi_i \\ \Leftrightarrow \frac{\partial f_{tot}}{\partial T} = -f^2 \frac{\partial \Phi_{tot}}{\partial T} = -f^2 \sum_{i=1}^{j} \gamma_i(T)\Phi_i \end{cases} \tag{7.24}$$

The total thermal defocus Δz expressed in Equation 7.17 can be generalized for j thin lenses in contact [56]:

$$\Delta z = \left| f \left[\sum_{i=1}^{j} (\gamma_i \Phi_i) + \alpha_H \right] \Delta T \right| \tag{7.25}$$

7.3.3 Designing an achromat athermal triplet

Using a triplet lens system is the simplest way to correct both thermal and chromatic aberrations. Remind that each IR material can be defined by two coefficients γ (thermal) and ω (chromatic). For j thin lenses in contact, the conditions for power conservation, achromatisation and athermalisation are provided by Equations 7.23 and 7.25 [56]:

$$\begin{cases} \displaystyle\sum_{i=1}^{j} \Phi_i = \Phi \\ \Delta_\lambda f = 0 \Leftrightarrow \displaystyle\sum_{i=1}^{j} \omega_i \Phi_i = 0 \\ \Delta z = 0 \Leftrightarrow \displaystyle\sum_{i=1}^{j} \gamma_i \Phi_i = -\alpha_H \Phi \end{cases} \tag{7.26}$$

Recall that System 7.26 is valid regardless of the type of thin lenses. The refractive or diffractive nature of lens i only impacts the expression of γ_i and ω_i, provided in Section 7.1. System 7.26 has three equations, meaning that at least three distinct materials must be used to have a unique design solution (Φ_1, Φ_2, Φ_3). However, it is possible to find a combination of only two materials that is achromatic and athermal. Considering a lens doublet, we solve System 7.26, leading to the athermal condition:

$$\frac{1}{\omega_1}(\gamma_1 + \alpha_H) = \frac{1}{\omega_2}(\gamma_2 + \alpha_H) \tag{7.27}$$

In the case of a thin lens triplet with three distinct materials, System 7.26 is written in matricial form and solved by multiplying the second member with the inverted matrix given by:

$$\begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \end{pmatrix} = \frac{1}{det} \begin{pmatrix} \gamma_3\omega_2 - \gamma_2\omega_3 & \gamma_2 - \gamma_3 & \omega_3 - \omega_2 \\ \gamma_1\omega_3 - \gamma_3\omega_1 & \gamma_3 - \gamma_1 & \omega_1 - \omega_3 \\ \gamma_2\omega_1 - \gamma_1\omega_2 & \gamma_1 - \gamma_2 & \omega_2 - \omega_1 \end{pmatrix} \begin{pmatrix} \Phi \\ 0 \\ -\Phi\alpha_H \end{pmatrix} \tag{7.28}$$

det is the determinant of the system, defined by :

$$det = \omega_1(\gamma_2 - \gamma_3) + \omega_3(\gamma_1 - \gamma_2) + \omega_2(\gamma_3 - \gamma_1) \tag{7.29}$$

7.4 Design of thin air-spaced lenses

In this section, we study the design of refractive systems with any number of air-spaced thin lenses. The effect of a temperature variation on a refractive system is illustrated in figure 7.5:

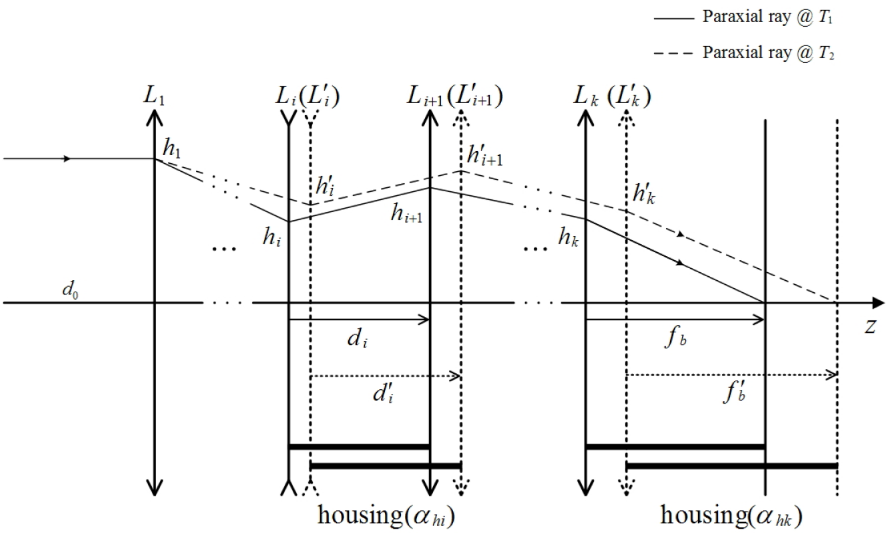


Figure 7.5: Configuration of an optical system containing k thin lenses separated by k-1 spaces. All design parameters vary when the temperature changes from T_1 to T_2. h_i represents the paraxial ray height of lens i, and f_b is the back focal length of the system.

The relation between the total power and the power of each lens (Equation 7.21) must be adapted to take air spaces into account. Figure 7.6 describes the layout of two air-spaced thin lenses:

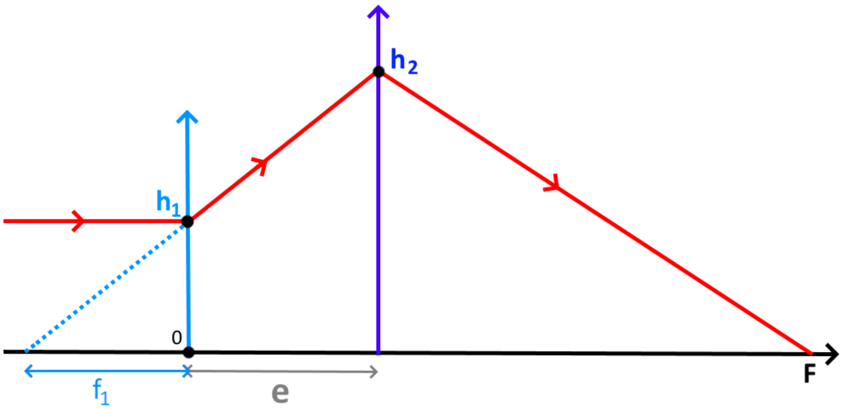


Figure 7.6: Optical system containing 2 thin lenses.

From Figure 7.6 layout, we deduce:

$$\frac{h_2}{h_1} = \frac{-f_1 + e}{-f1} = 1 - e\Phi_1 \tag{7.30}$$

Using the Gullstrand formula (Equation 7.20) for two air-spaced thin lenses:

$$\Phi = \Phi_1 + \frac{h_2}{h_1}\Phi_2 \tag{7.31}$$

This relation generalizes to any number of thin air-spaced lenses. We weight the power Φ_i of lens i by the ratio h_i/h_1, where h_i and h_1 are respectively the ith and first lens paraxial ray height. For an optical system containing k thin lenses separated by $k-1$ spaces, we re-define the optical power of each element i as [58]:

$$\Phi_i' \triangleq \frac{h_i}{h_1}\Phi_i \tag{7.32}$$

We can re-write the athermal and achromat equations for an air-spaced optical system as [59]:

$$\text{Total power: } \frac{1}{f} = \Phi = \sum_{i=1}^{k} \Phi_i' \tag{7.33}$$

$$\text{Achromatism: } \Delta_\lambda f = \frac{1}{\Phi^2}\sum_{i=1}^{k} \omega_i \Phi_i' = 0 \tag{7.34}$$

The total defocus amount is computed as the difference between the housing expansion and the thermal focal shift. We also make the approximation that the optical powers vary much more than the paraxial ray heights [59]:

$$\frac{\partial \Phi_i'}{\partial T} \simeq \frac{h_i}{h_1}\frac{\partial \Phi_i}{\partial T} \tag{7.35}$$

The athermal equation for k air-spaced thin lenses (Equation 7.25) becomes:

$$\text{Athermalism : } \frac{\Delta_T f}{\Delta T} = -f\left(f\sum_{i=1}^{k} \gamma_i \Phi_i' + L\alpha_H\right) = 0 \tag{7.36}$$

For an object at infinity, the housing length L equals the effective focal length f. The final athermal achromat system for k thin lenses separated by $k-1$ spaces is given by:

$$\begin{cases} \sum_{i=1}^{k} \Phi_i' = \Phi \\ \sum_{i=1}^{k} \omega_i \Phi_i' = 0 \\ \sum_{i=1}^{k} \gamma_i \Phi_i' = -\alpha_H \Phi \end{cases} \tag{7.37}$$

7.4.1 Equivalent lens model

System 7.37 can be solved with a unique solution $(\Phi_1, ..., \Phi_k)$ given k materials, a housing material TEC and the required effective focal length. However, since System 7.37 is not meant for aberration correction, this approach is unlikely to result in a performing design. In addition, System 7.37 relies on lenses material but does not provide any material selection method. Due to an important number of potential glasses and housing materials, combining all the possible choices to find the best one (which is not guaranteed to perform well) is extremely time-consuming and very challenging. In this section, we describe the "equivalent lens model" [59]. It allows determining suitable material combinations that do not degrade the design merit functions. It involves a graphical representation of the lens system.

We consider a refractive system made of K thin lenses separated by $K-1$ spaces. It is possible to build an equivalent system where $k-1$ lenses are replaced by an "equivalent lens" called L_e. The only remaining lens is called L_j. Hence, the system has now three equations for only two lenses, L_e and L_j. The equivalent lens L_e is defined by the optical power Φ_e, the chromatic power ω_e and the thermal power γ_e, defined as [59]:

$$\begin{cases} \Phi_e = \sum_{i=1}^{k}(\Phi'_i) - \Phi'_j \\ \gamma_e = \frac{1}{\Phi_e}\left[\sum_{i=1}^{k}(\omega_i\Phi'_i) - \omega_j\Phi'_j\right] \\ \omega_e = \frac{1}{\Phi_e}\left[\sum_{i=1}^{k}(\gamma_i\Phi'_i) - \gamma_j\Phi'_j\right] \end{cases} \tag{7.38}$$

We now try to athermalize and achromatize the doublet composed of L_e and L_j. System (7.37) is rewritten using the equivalent doublet L_e and L_j: [59] :

$$\begin{cases} \Phi'_j + \Phi_e = \Phi \\ \omega_j\Phi'_j + \omega_e\Phi_e = 0 \\ \gamma_j\Phi'_j + \gamma_e\Phi_e = -\alpha_H\Phi \end{cases} \tag{7.39}$$

Solving the first two equations of System (7.39) leads to a unique achromatic solution for the equivalent doublet:

$$\begin{cases} \Phi'_j = -\omega_e\Phi/(\omega_j - \omega_e) \\ \Phi'_e = +\omega_j\Phi/(\omega_j - \omega_e) \end{cases} \tag{7.40}$$

The third equation of System (7.39) gives the so-called "athermal condition" [60]:

$$\gamma_j = \frac{\gamma_e + \alpha_H}{\omega_e}\omega_j - \alpha_H \tag{7.41}$$

This condition geometrically represents a line equation. It corresponds to the line passing by the two points (ω_e,γ_e) and (0,$-\alpha_H$). It provides a graphical necessary (but not sufficient) condition to have an achromat athermal system. It results in a graphical tool named "athermal glass map" and is depicted in Section 7.4.2.

7.4.2 Athermal chart design tool

The "athermal glass map" or "athermal chart" tool is a visual aid to find suitable material combinations to athermalize a design quickly. A lens is represented by three parameters (Φ, ω, γ) where Φ is the optical power (linked to the lens material, shape, etc.), ω is the chromatic power and γ the thermal power. Note that the couple (ω, γ) completely defines the optical behaviour of any material. (ω, γ) values, although temperature and wavelength dependent, has been approximated for all IR materials in Section 7.2.2.

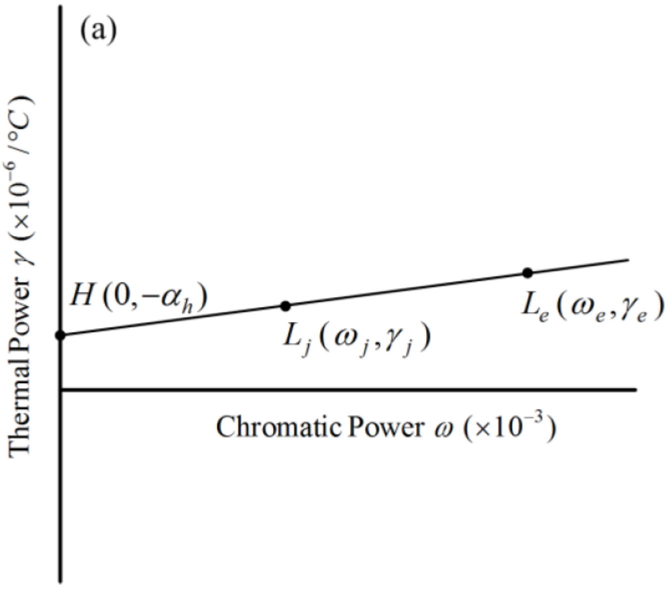


Figure 7.7: Athermal condition on an athermal glass map [59].

The equivalent doublet (L_e, L_j), is graphically represented on a 2D chart by the points (ω_e,γ_e), (ω_j,γ_j) while the housing material corresponds to the point H$(0,-\alpha_H)$ [47, 59], like in Figure 7.7.

In Figure 7.7, the equivalent doublet satisfies the athermal condition considering the housing material α_H since the line passing by $L_e(\omega_e,\gamma_e)$ and $L_j(\omega_j,\gamma_j)$ intercepts the point H$(0,-\alpha_H)$. Any couple (L_e, L_j) defines a specific housing material α_H on the $\omega = 0$ axis. However, this "fictitious" housing material is unlikely to correspond to existing material. Nonetheless, the athermal glass map tool is very helpful in quickly identifying a suitable housing material, i.e. the existing material with the closest TEC compared to the "fictitious" one.

The design is optimised in the next step so that the athermal line intercepts the newly identified existing housing material. This is called "redistributing the powers" [59] because the equivalent lens L_e, which contains all the lenses except L_j, is brought to a new position L'_e as depicted in Figure 7.8.

We use an optical optimisation software (such as Code V) to bring L_e to L'_e (Figure 7.8), using the constraints defined in Equations 7.40 and 7.41. Nevertheless, this method is impractical since it modifies nearly all the lenses at once. It is extremely likely to drastically change the design layout (to a point where no more optimisation is possible) and degrade its optical quality. Consequently, a "smoother" multi-step method is described in the following. Its goal is to minimize the impact of the change $L_e \rightarrow L'_e$.

The only degree of freedom left with the equivalent lens model is lens L_j material. The latter can be changed with fewer risks of degrading the design's optical quality. The result of modifying L_j first and then L_e is depicted in Figure 7.9:

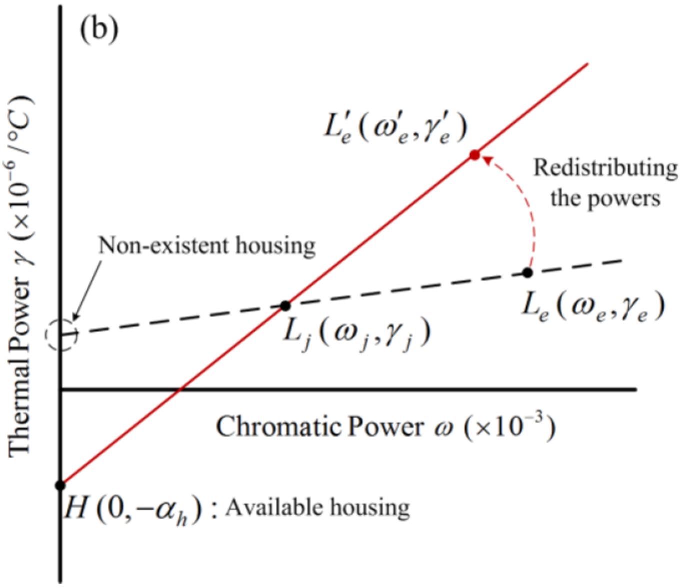
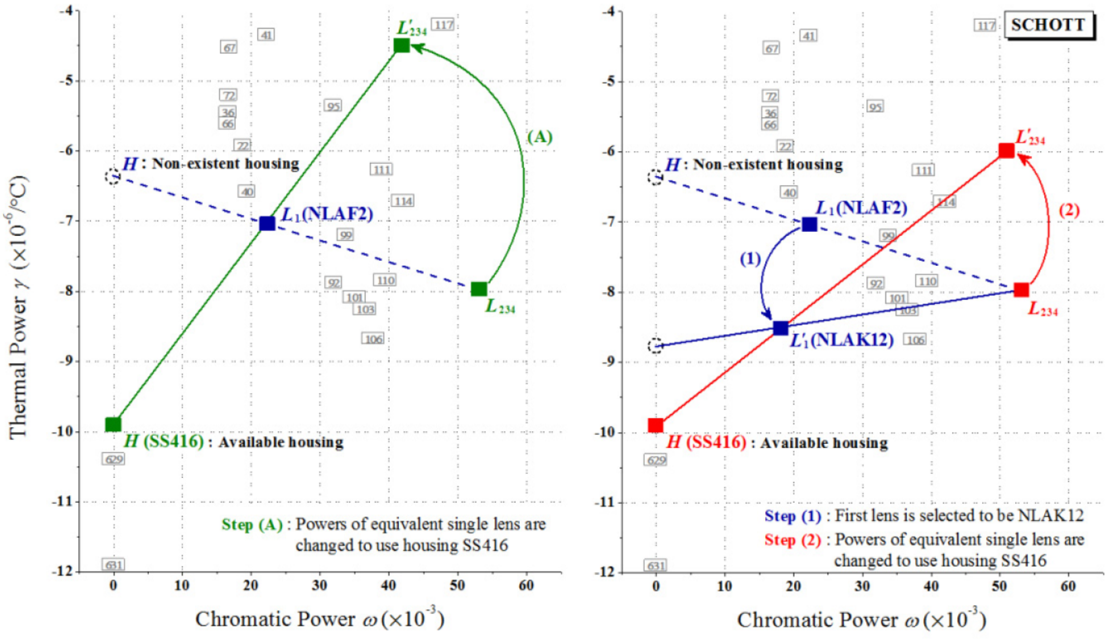


Figure 7.8: Principle of achromatic and athermal design by redistributing the powers [59].



(a)

(b)

Figure 7.9: (a) Directly redistributing the powers, (b) Changing lens j material before redistributing the powers [59].

Figure 7.9 shows how it is possible to reduce the distance (i.e. the optimisation constraints) between L_e and L'_e by modifying L_j first. The optical performance are conserved if $\omega'_j \simeq \omega_j$.

7.4.3 Selecting the substitute lens L_j

To achieve athermalisation, care must be taken regarding the choice of the substitute lens L_j. The number of lenses replacement, i.e. the efficiency of the algorithm, depends on the choice of L_j. Before applying the equivalent lens model, we define a chromatic and thermal weight for each lens i [60]:

$$\text{Chromatic weight } M_i = |\omega_i \Phi'_i| / (\sum_{i=1}^{k} |\omega_i \Phi'_i|) \tag{7.42}$$

$$\text{Thermal weight } N_i = |\gamma_i \Phi'_i| / (\sum_{i=1}^{k} |\gamma_i \Phi'_i|) \tag{7.43}$$

Then we define a global weight W_i that models the influence of each lens on the total optical dispersion and the total thermal defocus [60]:

$$\text{Total weight } W_i = M_i + N_i \tag{7.44}$$

We consider the lens with the highest weight to have the most significant influence on the design. It constitutes the substitute lens L_j while the remaining lenses constitute the equivalent lens L_e.

7.4.4 Athermal optimisation algorithm

The presented achromat athermal algorithm is illustrated in Figure 7.10, adapted from [60].

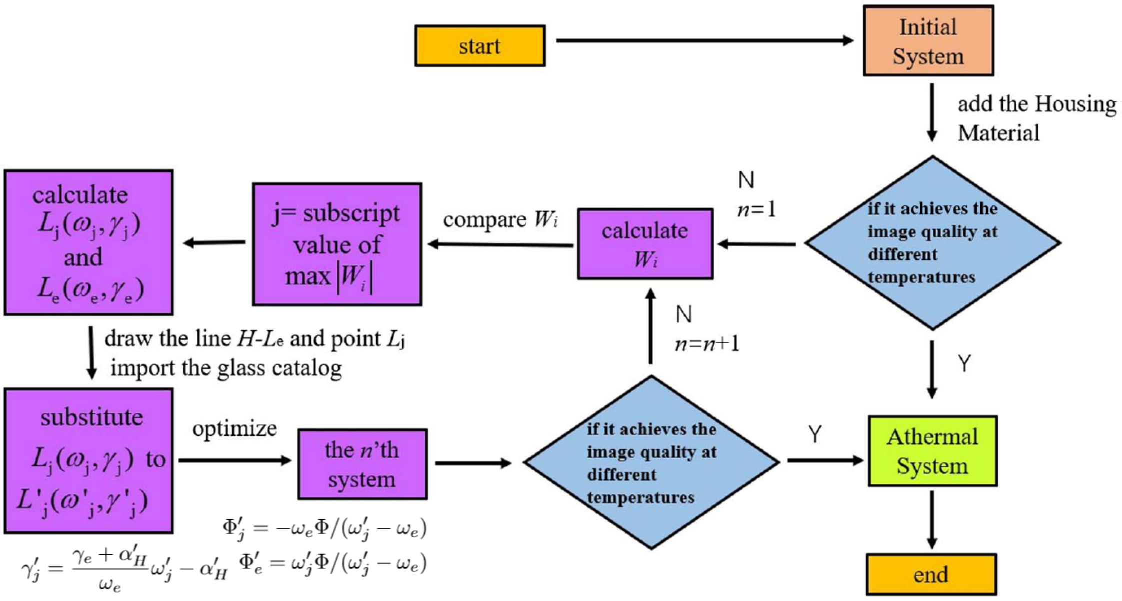


Figure 7.10: Design athermalisation algorithm [60].

The athermal algorithm is implemented in a Code V macro, along with an environmental macro that modifies the design for various temperatures and housing materials and computes the optical quality. We assume that an initial design has achieved sufficient image quality.

Step 1: Recover current lens information and material characteristics

This step initializes the algorithm. We create an "IR library" with all the couples (γ,ω) and each housing materials TEC α_H. Using an optical software (Code V), we recover all the lens parameters (f_i, h_i, Φ_i,γ_i, ω_i, W_i, surface number, lens type, etc.). Finally, we compute Φ'_i and the weight W_i for each lens i.

Step 2: substitute lens L_j and L_e

We search for the best substitute lens L_j by comparing the weights W_i of each element. We then compute the initial equivalent lens L_e.

Step 3: Replace L_j by L'_j

This step consists in changing lens L_j material ($L_j \rightarrow L'_j$). To do so, we compute the fictitious α_H obtained by changing $L_j \rightarrow L'_j$ and keeping L_e. The material of L_j is replaced by the material bringing the fictitious housing the closest to the target housing, like in Figure 7.9. In addition, the selected replacement material should have nearly the same chromatic power as L_j to degrade the optical quality as little as possible. As possible replacement solutions, we search all the materials (ω_{mat},γ_{mat}) that minimize:

$$\begin{cases} |\alpha_H - \alpha_{target}| = |\dfrac{\gamma_e \omega_{mat} - \gamma_{mat}\omega_e}{\omega_e - \omega_{mat}} - \alpha_{target}| \\ |\omega_{mat} - \omega_j| \end{cases} \tag{7.45}$$

Finally, we replace lens L_j material with the chosen one L'_j, and we auto-focus the design, using air spaces as variables.

Step 4: Redistribute the powers $L_e \rightarrow L'_e$

This step consists in "redistributing the powers". The equivalent lens L_e is changed to L'_e to satisfy both the athermal condition (Equation 7.41) and the achromatic equations 7.40. This step is performed using an optimisation software such as Code V:

- Define all the variable parameters (curvatures, conic cte, spaces,...)
- Define the merit function settings
- Create user-defined constraints and functions (ω_e,γ_e,Φ_e) that implements Equations (7.40,7.41).
- Define a flexible weight for those constraints to control the optimisation output (ensure convergence)
- Run Code V optimisation.

Step 5: Environmental check

Code V *ENV* macro allows simulating a temperature variation and recovering the optical quality for a given housing material. We use the RMS spot diagram as the figure of merit. The iterative optimisation process (steps 1 to 5) ends when the RMS spot size is below the pixel size for the required temperature range.

7.5 Application to a MWIR optical design

7.5.1 Initial design: application of athermal algorithm

In this section, we apply the athermal algorithm to build a MWIR design using CodeV, following the requirements established in Chapter 3 (see Table 3.2). We consider a LEO orbit at 400 km altitude, a 90 mm pupil diameter and an F/1.5 working F-number. The detector, similar to the one illustrated in Figure 2.2, have squared pixels of 4 µm pitch. It is composed of a Dewar tube with a cold stop to isolate the FPA at a very low temperature (∼77 °K). The diffraction Airy spot has a diameter:

$$D_{Airy} = 2.44 \lambda F/\sharp \simeq 17 \text{ µm} \tag{7.46}$$

The initial optical design contains two refractive lenses and one hybrid (refractive-diffractive) lens. This design is depicted in Figure 7.11 and the RMS spot diagram is shown in Figure 7.12.

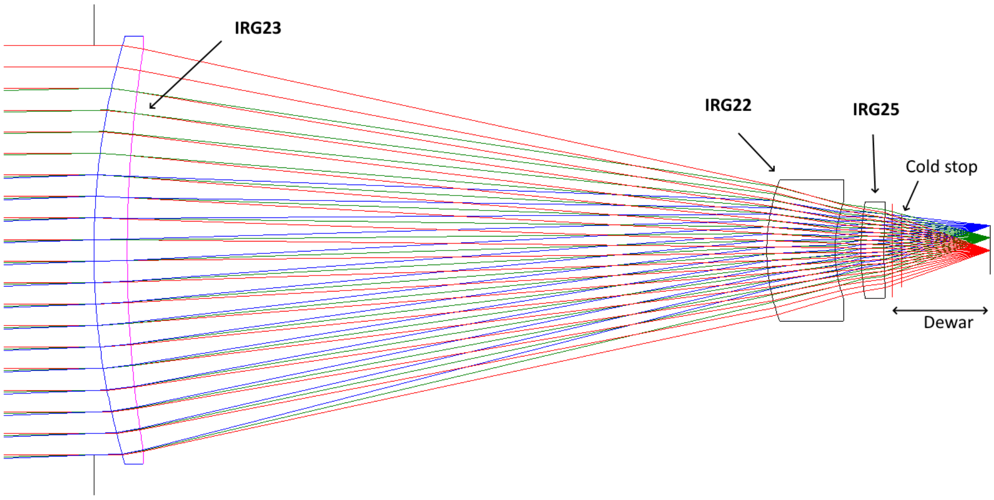


Figure 7.11: Optimised MWIR design with three lenses. The conic and diffractive surfaces are respectively displayed in purple and blue. The red vertical lines show the dewar and cold stop entrances.

The initial design spot sizes are ideal, with 100% of the energy collected inside a 6 µm diameter circle. Note that as long as optical aberrations remain smaller than the 17 µm Airy blur spot, they do not modify the PSF shape. Since the diffraction PSF is well sampled by 4.88 pixels (see Section 3.4.1), it can be deconvoluted to retrieve the exact pixel information. The RMS spot size is small enough to fit inside a 4 µm pixel.

This section aims to analyse the thermal variation of the spot size and apply the athermal algorithm if needed. Using the environment tool of Code V (*ENV*), the RMS spot size is depicted for various temperatures in Figure 7.13.

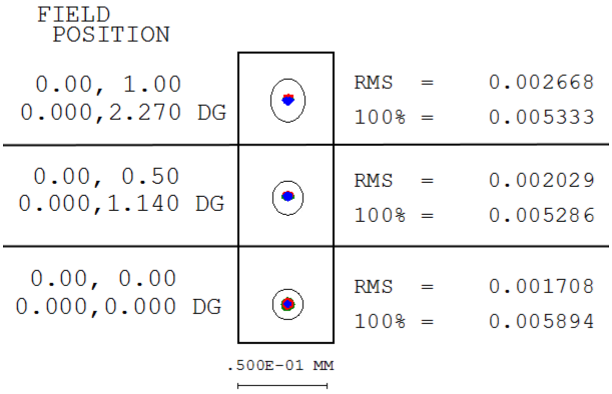


Figure 7.12: RMS spot diameter (mm) for three FoVs considering the initial MWIR design. The Airy disk is displayed with a circle or an ellipse. The off-axis FoVs has vignetting due to the cold stop, explaining the deformed Airy pattern.

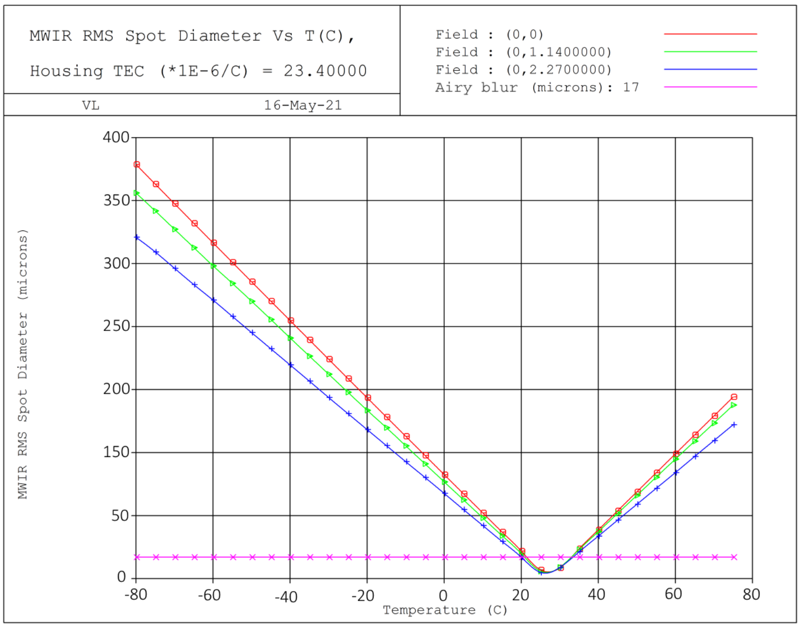


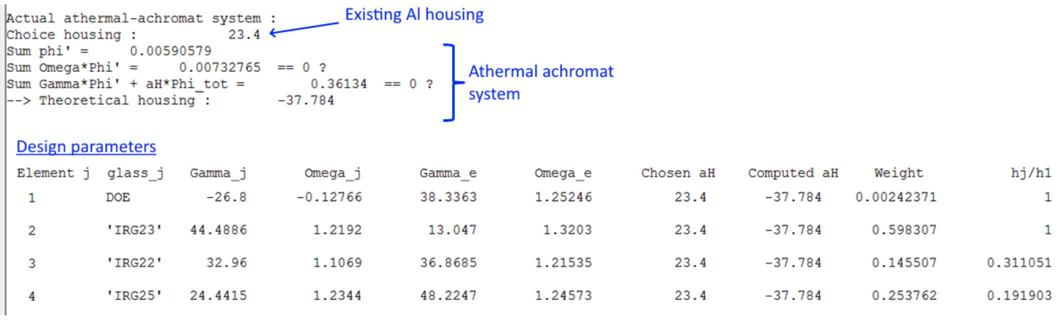
Figure 7.13: Code V environment simulation of the initial design, with Al housing. The reference temperature is 27°C (300°K). The purple line represents the Airy blur spot threshold of 17 µm.

In this section, we only consider classical aluminium housings, with a $TEC \simeq 23 \times 10^{-6}/^{\circ}K$. The reference temperature is set to 27°C. The constant purple line displays the selected pixel size. According to Figure 7.13 the image quality remains sufficient inside the thermal range $[20-35]^{\circ}C^{\circ}$. This range is small and needs to be extended, especially for space applications.

The initial design thermal parameters are computed according to steps 1 and 2 (See Section 7.4.4), and the text output of the Code V macro is displayed in Figure 7.14:

Figure 7.14 shows that the optimal housing material is fictitious and that the refractive part of the first lens has the highest weight. Figure 7.15 displays an environmental simulation of the initial design, considering the fictitious housing computed in Figure 7.14:

According to Figure 7.15, if a non-existing housing resulting from the application of the



```
Actual athermal-achromat system :
Choice housing :              23.4
Sum phi' =      0.00590579
Sum Omega*Phi' =      0.00732765   == 0 ?
Sum Gamma*Phi' + aH*Phi_tot =        0.36134   == 0 ?
--> Theoretical housing :        -37.784

Element j  glass_j   Gamma_j      Omega_j     Gamma_e     Omega_e     Chosen aH   Computed aH   Weight         hj/h1
 1         DOE         -26.8     -0.12766     38.3363     1.25246         23.4       -37.784   0.00242371         1

 2         'IRG23'   44.4886       1.2192      13.047      1.3203         23.4       -37.784     0.598307         1

 3         'IRG22'     32.96       1.1069     36.8685     1.21535         23.4       -37.784     0.145507  0.311051

 4         'IRG25'   24.4415       1.2344     48.2247     1.24573         23.4       -37.784     0.253762  0.191903
```

Figure 7.14: Code V macro text output: Initial design thermal parameters. The chosen housing material is Al with a TEC of 23.4×10^{-6}/°K. The equivalent lens model leads to a different "fictitious" housing with a TEC of -37.78×10^{-6}/°K. The first refractive lens (without the front diffractive surface) has the highest thermal weight. The equivalent lens model leads to different parameter outputs depending on the choice of L_j.

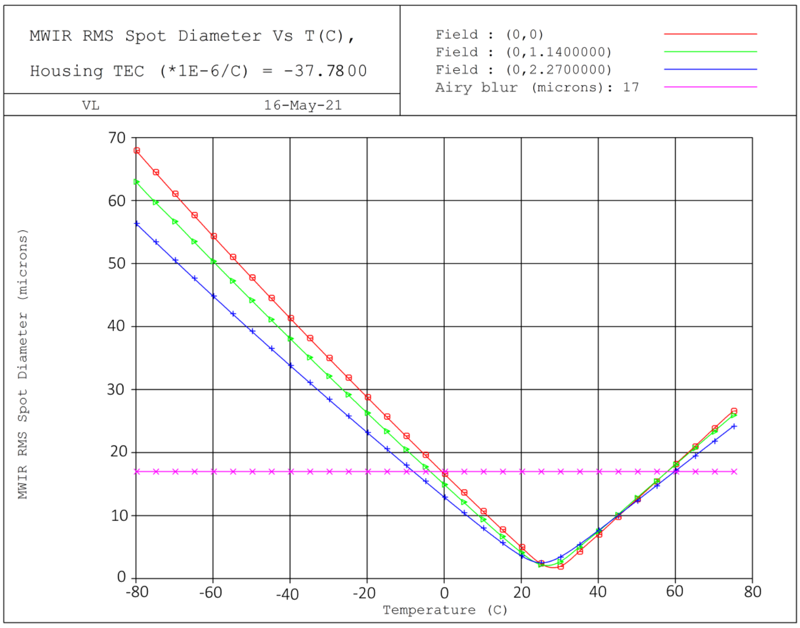


Figure 7.15: Code V environment simulation of the initial design, with non-realistic but optimal housing. The purple line represents the Airy blur spot threshold of 17 µm.

athermal equations is used in CodeV Env macro, the acceptable thermal range can be widened from 0 °C to 60 °C. This numerical simulation, though unrealistic, shows the validity of these equations. We now use the athermal algorithm to modify the design and converge to the desired Al housing.

7.5.2 Practical adjustments to the athermal algorithm

As can be seen in Figure 7.14, the achromatic equation $\sum_{i=1}^{k} \omega_i \Phi_i' = 0$ is not verified since we have, after Code V optimisation:

$$\sum_{i=1}^{k} \omega_i \Phi_i' = 0.0073 \mathrel{\widehat{=}} \epsilon \tag{7.47}$$

The residual ϵ comes from the thin lens approximation used in the achromatic equation that is not as accurate as Code V ray tracing. Considering this residual, the solution of the athermal achromat system 7.37 becomes:

$$\begin{cases} \Phi_j' = (\epsilon - \omega_e \Phi)/(\omega_j - \omega_e) \\ \Phi_e' = (\epsilon + \omega_j \Phi)/(\omega_j - \omega_e) \end{cases} \tag{7.48}$$

$$\gamma_j = \frac{\gamma_e + \alpha_H}{\omega_e - \dfrac{\epsilon}{\Phi}} \omega_j - \frac{\alpha_H \omega_e + \gamma_e \dfrac{\epsilon}{\Phi}}{\omega_e - \dfrac{\epsilon}{\Phi}} \equiv \frac{\gamma_e + \alpha_H}{\omega_e - \dfrac{\epsilon}{\Phi}} \omega_j - B(\alpha_H, \omega_e, \gamma_e) \tag{7.49}$$

The athermal line equation now depends on the ratio ϵ/Φ. When $\epsilon = 0$, Equations 7.48 and 7.49 reduce to the achromat athermal solution (Equations 7.40 and 7.41). Using Code V, we deduce that $\epsilon/\Phi \simeq 1$, meaning that an important mistake arises if the residual ϵ is neglected. The line Equation 7.49 still pass by the point (L_e, L_j). However, the line now intersects the $\omega = 0$ axis at the point $H(0, -B(\alpha_H, \omega_e, \gamma_e))$. This intersection does not directly provide the optimised housing TEC α_H anymore and depends on the equivalent lens parameters (ω_e, γ_e).

7.5.3 Athermal optimisation and material replacements

In the following graphical representations (Figures 7.16, 7.17 and 7.18), the resulting fictitious housing TEC α_H have been displayed instead of the coordinate origin $-B(\alpha_H, \omega_e, \gamma_e)$ for clarity. Therefore, the α_H values are not necessarily ordered in the γ axis.

The initial MWIR design has been described in Figures 7.11 and 7.13. The first iteration of the athermal algorithm is graphically shown in Figure 7.16.

Since the fictitious TEC resulting from the first iteration is not close enough to Al, a second iteration is made. Thermal and chromatic weights are computed again with the newly obtained design, resulting in a potentially new substitute lens L_j and equivalent lens L_e (same process as in Figure 7.14). During the redistribution of powers (i.e. Code V optimisation), the constraints weights are kept low to ensure that the design performance is not degraded too much. The second iteration is illustrated in Figure 7.17.

Similarly to Figure 7.16, Figure 7.17 shows how the algorithm converges to fictitious housings that are closer to Al. The substitute lens L_j has been replaced by IRG26 during this iteration. Since the redistribution of powers moves L_e to L_e' by a large distance on the map, the Code V constraint weights are kept low. Figure 7.18 describes the last iteration of the athermal algorithm.

In the last iteration depicted by Figure 7.18, the second IRG22 lens has been selected as substitute lens L_j and replaced by ZnSe, making a massive step towards the final goal $TEC = 23.4$. The final housing TEC value is 26.5, close to the target aluminium TEC. The constraint weights have been increased by a vast amount from 0.01 to 5, allowing the final optimisation to reach the targeted value.

From this optimisation example, the step $L_j \Rightarrow L_j'$ allows quick and efficient convergence since the last "redistribution of powers" optimisation could recover the required image quality. The final design layout as well as the environmental simulation are depicted in

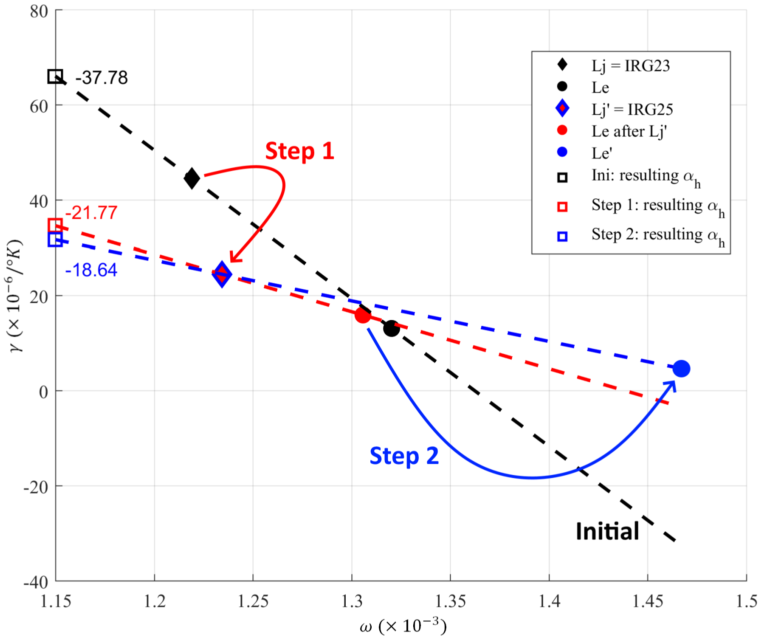


Figure 7.16: First iteration. Initially, L_j (IRG23, black diamond) and L_e (black dot) provide a fictitious material α_H, $TEC = -37.8 \times 10^{-6}/K$. Step 1: $L_j \leftarrow L'_j$ (IRG25, red and blue diamond) and the system is auto-focused (red dot L_e), providing a second fictitious material α_H, $TEC = -21.8 \times 10^{-6}/K$. Note that $\omega_j - \omega'_j \simeq 0$. Step 2: Powers are redistributed $L_e \leftarrow L'_e$ (blue dot) leading to the final fictitious material α_H, $TEC = -18.6 \times 10^{-6}/K$.

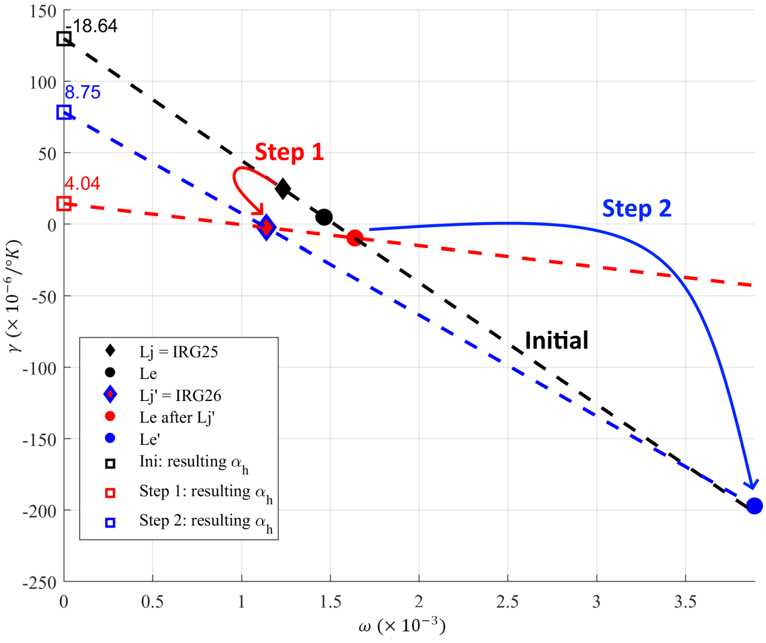


Figure 7.17: Second iteration. Initialy, L_j (IRG25, black diamond) and L_e (black dot) provide a fictitious material α_H, $TEC = -18.6 \times 10^{-6}/K$. Step 1: $L_j \leftarrow L'_j$ (IRG26, red and blue diamond) and the system is auto-focused (red dot L_e), providing a second fictitious material α_H, $TEC = 4.04 \times 10^{-6}/K$. Note that $\omega_j - \omega'_j \simeq 0$. Step 2: Powers are redistributed $L_e \leftarrow L'_e$ (blue dot) leading to the final fictitious material α_H, $TEC = 8.75 \times 10^{-6}/K$.

Figures 7.19 and 7.20.

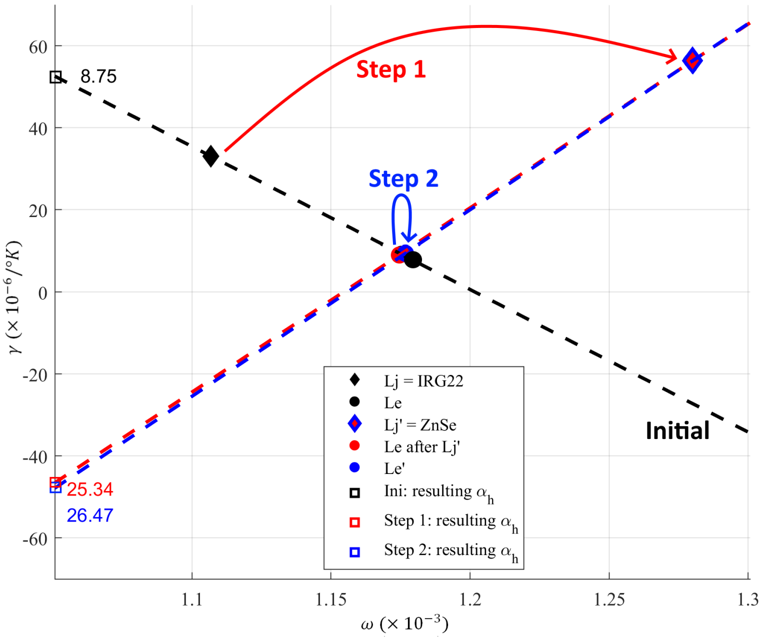


Figure 7.18: Third and last iteration. Initialy, L_j (IRG22, black diamond) and L_e (black dot) provide a fictitious material α_H, $TEC = 8.75 \times 10^{-6}/K$. Step 1: $L_j \leftarrow L'_j$ (ZnSe, red and blue diamond) and the system is auto-focused (red dot L_e), providing a second fictitious material α_H, $TEC = 24.5 \times 10^{-6}/K$. This time, the difference $\omega_j - \omega'_j$ allows to reduce the impact of redistribution of powers. Step 2: Powers are redistributed $L_e \leftarrow L'_e$ (blue dot) leading to the final fictitious material α_H, $TEC = 26.47 \times 10^{-6}/K$.

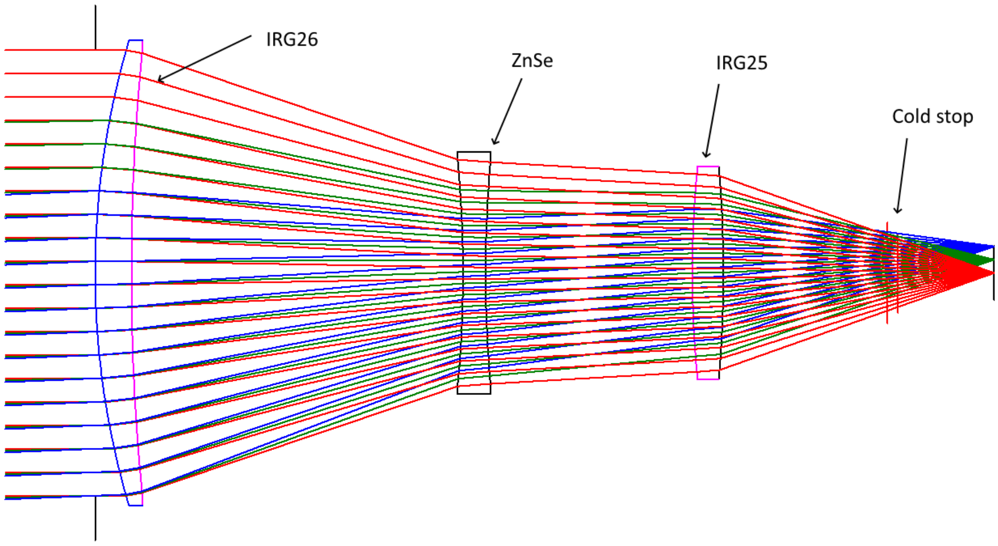


Figure 7.19: optimised athermal MWIR design with three lenses. The conic and diffractive surfaces respectively appear in purple and blue. The red vertical lines show the dewar and cold stop entrance.

In conclusion, the application of the athermal achromat algorithm enables a diffraction-limited optical performance within a large temperature range ±50°K. Recall that the initial design in Figure 7.11 had different lens materials and powers and could only withstand ±5°K (Figure 7.13). The optimised MWIR camera follows every requirements defined in Part I apart from the dual-band capability. The dual-band issue occurs for the diffractive lens used in the design since it cannot operate efficiently inside a broad wave-

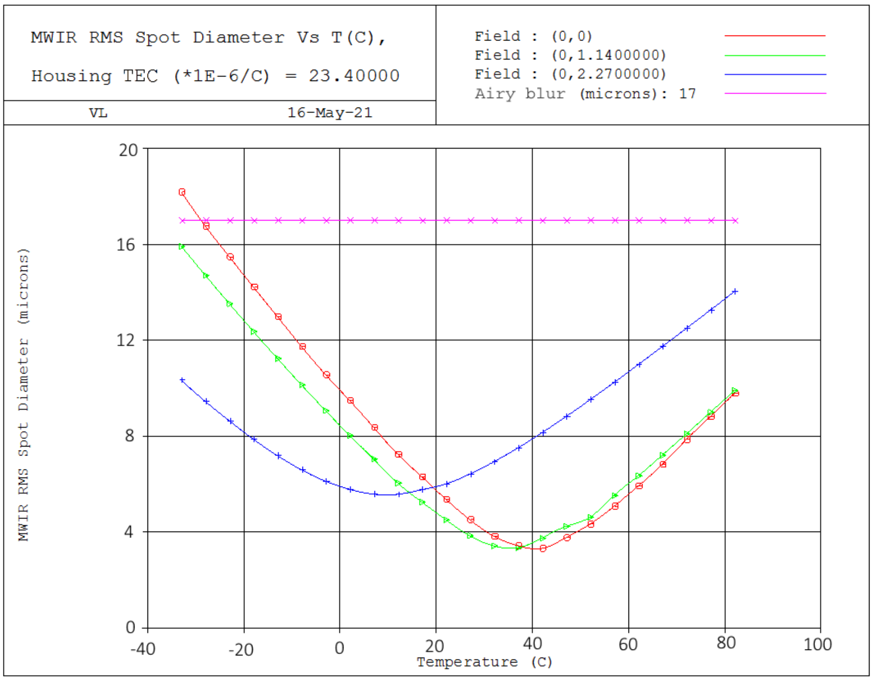
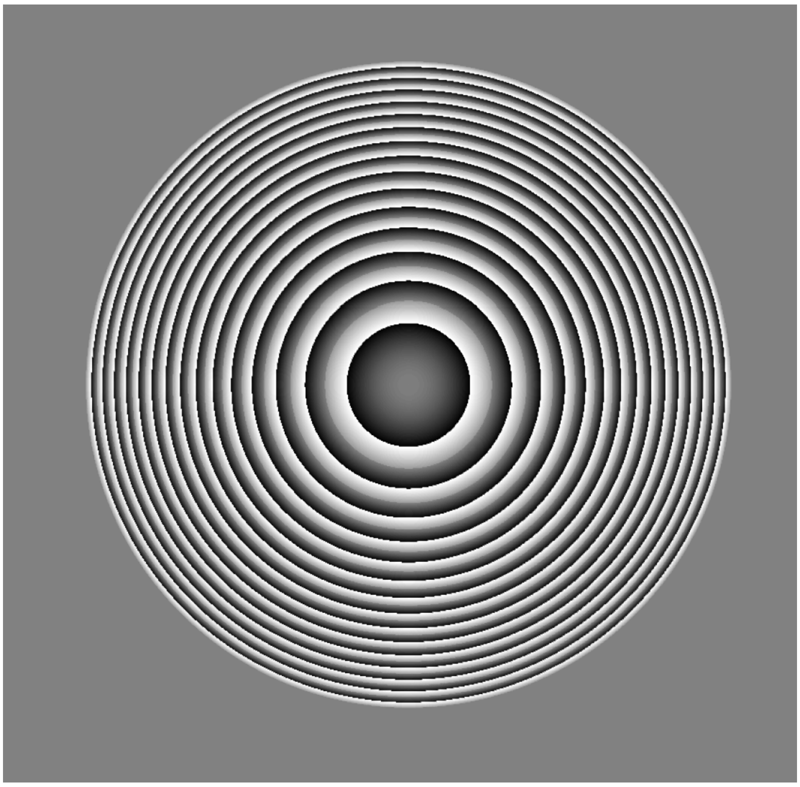


Figure 7.20: Code V ENV simulation with Al housing TEC. The RMS spot size is the figure of merit, and the reference temperature is 27°C. The purple line displays the 17 µm Airy blue spot diameter.

band as explained in Chapter 8. The extension of this camera to the LWIR bandwidth becomes the major research question of the thesis, and will lead to many innovations in the modelling of multilyaer diffractive optical elements, introduces in Chapter 9.



Chapter 8

Diffractive lenses design

We aim to extend Chapter 7 athermal MWIR design to LWIR by researching efficient dual-band diffractive optical elements solutions. This chapter presents diffractive optical elements (DOEs) principles and design equations. These extremely thin imaging elements are composed of parabolic discontinuous zones and produce multiple foci, called "diffractive orders". Their design originates from the thin element approximation (TEA) and the requirement of constructive interferences at the focal plane. We derive the analytic zone profile, phase and diffraction efficiency of DOEs and study the energy distribution across multiple orders. For a specific order, this energy can only be maximized for one wavelength, called "design wavelength", and drops for any other wavelengths. It is, however, possible to consider other orders to obtain the maximum diffraction efficiency for multiple wavelengths. The study of DOE chromatic focal shift is a significant point of this chapter, showing that multi-order DOEs are unsuitable for our dual-band optical design.

Contents

8.1 Introduction to the scalar theory of diffraction (SDT)

Refractive or reflective imaging optical elements such as lenses or mirrors are very accurately studied using ray-tracing (ZEMAX, CODEV or ASAP software). Ray-tracing is a scalar approximation of Maxwell's equations when the characteristic interface size is much larger than the wavelength. Diffractive optical elements (DOEs), in this case diffractive Fresnel lenses, are quasi-periodic with an aperture divided into discontinuous zones of period T. The ratio T/λ defines the limit between the refractive and diffractive regime. This ratio also establishes the validity of the analytical scalar theory of diffraction (SDT). Following [61], for periods $T > 14\lambda$, the SDT is considered accurate and provides an analytical phase model. The SDT allows retrieving the un-aberrated wavefront, resulting from the phase delay introduced by a DOE. This phase delay is expressed using the concept of optical path difference (OPD):

$$\begin{cases} U(x,y) = U_0 \exp\left[i\Phi(x,y)\right] \\ \Phi(x,y) \equiv \dfrac{2\pi}{\lambda} OPD(x,y;\lambda) \end{cases} \tag{8.1}$$

$U(x,y)$ is the scalar complex phasor corresponding to the DOE (see Chapter 4 for more details). It involves a phase delay Φ, computed using the optical path difference (OPD). The optical path length (OPL) of a ray travelling inside an optical component composed of N different refractive indexes n_i is defined as:

$$OPL(\lambda) = \sum_{i=1}^{N} n_i(\lambda) L_i \tag{8.2}$$

Where n_i is the refractive index of the medium i and L_i is the Cartesian length of the ray inside the medium i. Note that the N media in Equation 8.2 compose the optical component and does not account for the ray propagation in free space. The length L_i can be obtained using ray-tracing but is often computed analytically using the thin element approximation (TEA). The TEA is valid when the optical component is extremely thin (i.e. ratio thickness/period$<<$ 1, see Section 9.4 for the details). Under the TEA assumption, the length L_i can be computed as if the rays were not deviated by the optical media (no refraction). This approximation is only valid for paraxial rays parallel to the optical axis. Figure 8.1 displays the OPL calculation in the case of a thin parabolic biconvex refractive lens:

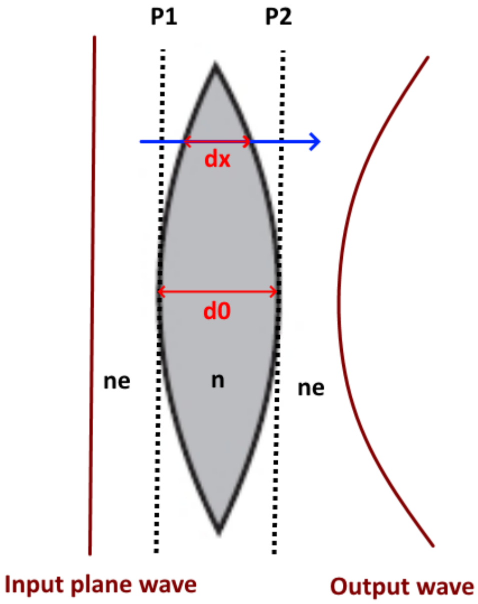


Figure 8.1: Optical path length calculation example for a thin refractive lens

Figure 8.1 displays the OPL calculation for a thin lens layout when the TEA is considered. The lens has a refractive index n while the external media has an index of n_e. At a certain radius x, the blue ray travels a distance d_x inside the lens, with no deviation at the interfaces. The optical path length of the blue ray is defined as:

$$OPL(x) = nd_x + n_e(d_0 - d_x) = d_x(n - n_e) + n_e d_0 \tag{8.3}$$

The distance d_x depends on the lens's curvature radius and conic constant. For instance, in Figure 8.1, we consider a parabolic biconvex lens with radius of curvature $\mathcal{R}$ and focal length f. We have:

$$\begin{cases} d_x = n_e d_0 - 2\dfrac{x^2}{2\mathcal{R}} = n_e d_0 - \dfrac{x^2}{2(n - n_e)f} \\ OPL(x) = -\dfrac{x^2}{2f} \end{cases} \tag{8.4}$$

Replacing x by a radius r in two dimensions, the phasor or complex transmittance associated with this biconvex parabolic lens is:

$$U(r, \lambda) = \exp\left[i\frac{2\pi}{\lambda}OPL(r)\right] = \exp\left[-ik\frac{r^2}{2f}\right] \tag{8.5}$$

Equation 8.5 describes the converging behaviour of a parabolic lens according to the TEA.

8.2 Diffractive Fresnel lenses

In the case of this thesis, where we focus only on imaging components, the general appellation diffractive optical element (DOE) always refers to diffractive imaging lenses, known as Fresnel diffractive lenses. The shape patterns of these elements are called Fresnel or kinoform zones because they are similar to the annular zones of refractive Fresnel lenses. Diffractive Fresnel lenses are very thin elements ($H \simeq 10$ µm). Such a thin imaging element drastically reduces the weight and the size of optical systems while keeping an excellent image quality, as shown hereafter.

8.2.1 Kinoform lens imaging properties

Diffractive lenses are to Fresnel lenses what gratings are to prisms. A DOE results from segmenting the parabolic profile of a refractive lens into parabolic zones, with two neighbouring zones having a 2π phase shift. Hereafter we show why the zone shapes are parabolic (Equation 8.11). The process is similar to phase wrapping, where a continuous phase is segmented into modulo 2π zones. The representation of such diffractive lens is depicted in Figure 8.2.

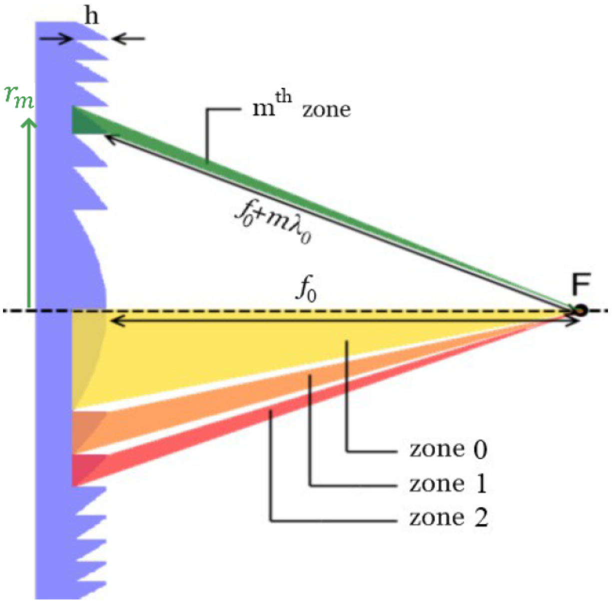


Figure 8.2: Layout definition of a kinoform diffractive lens [62]. The diffractive zones are annuli with parabolic profiles. The zone heights (h) and radii (r_m) are designed so that constructive interferences occur at the focal plane between each neighbouring zone (green, yellow, orange and red "beams") for the "design" wavelength λ_0.

Figure 8.2 describes the layout of a diffractive lens, with parabolic Fresnel zones. The primary design requirement is that constructive interferences occur between neighbouring zones at the focal plane for a specific wavelength λ_0. As expressed in Figure 8.2, the OPD between each diffractive zone is increased by λ_0, which corresponds to a 2π phase shift. For instance, the OPL associated with the mth zone is $OPL_m = f_0 + m\lambda_0$. The specific wavelength λ_0 is called the "design wavelength". Based on Figure 8.2 layout, the zones radii are calculated using the Pythagorean theorem:

$$r_m^2 = 2m\lambda_0 f_0 + (m\lambda_0)^2 \simeq 2m\lambda_0 f_0 \qquad (8.6)$$

The radii's design ensures that the diffractive zones interfere constructively at the focal point, but their profiles and heights remain unknown. Fermat's principle states: "All the rays coming together in a determined zone, even though they travel on distinct paths, result in having an equal optical path length to a fixed point F, placed on the axis at a distance f_0 from the DOE" [50]. This principle is illustrated in Figure 8.3 and serves as the basis for the DOE profile determination.

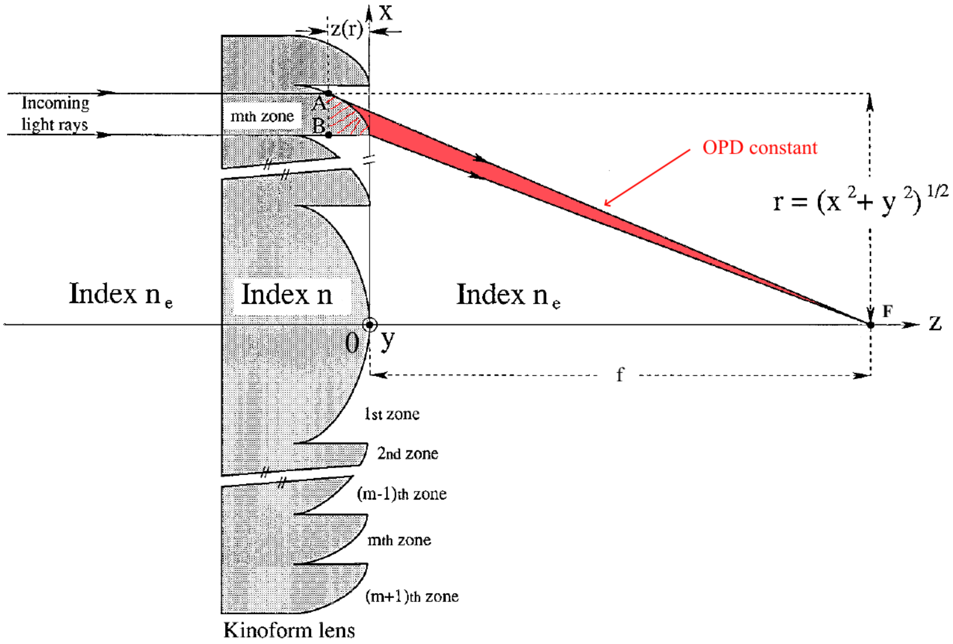


Figure 8.3: Kinoform lens profile [50] serving as a reference geometrical layout. The OPL at any radius r inside the mth zone is $\overline{\mathrm{AF}}$. $\overline{\mathrm{BF}}$ is a particular OPL, defined at the edge of zone m ($r = r_m$). The shape of the mth zone is obtained by imposing $\overline{\mathrm{AF}} = \overline{\mathrm{BF}}$ for any radius r inside zone m.

We illuminate the kinoform with a plane wave, all the incoming light rays parallel to the axis. We consider two of these incoming rays as in Figure 8.3, one of them goes through the mth zone in A, and the other goes through the zone edge in B. The rays cover the same optical path up to points A and B. According to Fermat's principle, the optical path between points A and F has to be the same as the optical path between B and F. Equating the two OPLs ($\overline{\mathrm{AF}} = \overline{\mathrm{BF}}$) leads to the following equation [50]:

$$n_e\sqrt{r^2 + (f0 - z)^2} = n_e(f0 + m\lambda_0) - nz \tag{8.7}$$

where the expression of the OPL $\overline{\mathrm{BF}}$ comes from the constructive interferences requirement expressed in Figure 8.2. Note that the profile height $z(r)$ is counted negatively in Figure 8.3. n and n_e are the refractive indices of the lens and exterior medium. We impose $n_e = 1$ for the sake of clarity.

The calculations of zone profile $z(r)$, phase $\Phi(r, \lambda)$, complex transmittance $U(r, \lambda)$ and diffraction efficiency $\eta(j, \lambda)$ are entirely based on the work of V. Moreno *et al* [50]. The main calculation steps are detailed hereafter to understand the origin of DOE shapes and figures of merit.

Reordering terms in Equation 8.7 lead to the equation of an hyperboloid of revolution:

$$\frac{[z - z_0]^2}{a^2} - \frac{r^2}{b^2} = 1 \tag{8.8}$$

The hyperboloid is centred around the point $(z_0, 0)$. The parameters z_0, a and b are defined as:

$$\begin{cases} z_0 \equiv \dfrac{(n-1)+nm\lambda_0}{n^2-1} \\ a \equiv \dfrac{(n-1)-m\lambda_0}{n^2-1} \\ b \equiv \dfrac{f(n-1)-m\lambda_0}{\sqrt{n^2-1}} \end{cases} \tag{8.9}$$

Equation 8.8 can be rewritten to obtain $z(r)$, considering an hyperboloid placed at the left hand side of the origin like in Figure 8.3:

$$\begin{cases} z(r) = \dfrac{n}{n^2-1}\left(m\lambda_0 + \zeta - [\zeta - m\lambda_0]\sqrt{1 + r^2 + \left[\dfrac{\delta}{\zeta - m\lambda_0}\right]^2}\right) \\ \zeta \equiv \dfrac{1}{n}(n-1)f \\ \delta \equiv \dfrac{1}{n}\sqrt{n^2-1} \end{cases} \tag{8.10}$$

This expression can be simplified using a Taylor expansion to approximate the square root, since $[\delta/(\zeta - m\lambda_0)]^2 << 1$. The higher terms of this expansion are negligible, and we make the approximation $\zeta - m\lambda_0 \simeq \zeta$ without losing accuracy. The final expression for the zone height $z(r)$ results in:

$$\begin{cases} z(r) = Z_{max}\left(m - \dfrac{r^2}{r_1^2}\right) \\ Z_{max} \equiv \dfrac{\lambda_0}{n - n_e} \\ r_1^2 \equiv 2\lambda_0 f_0 \end{cases} \tag{8.11}$$

where Z_{max} is the maximal zone height, n and n_e respectively being the refractive index of the DOE and the outside media, for the wavelength λ_0. The radius r is taken inside the mth zone: $r \in [r_m, r_{m+1}[$. Equation 8.11 is an accurate parabolic approximation of the rigorous hyperbolic shape (Equation 8.10).

In summary, the interference requirement (for the specific wavelength λ_0) combined with Fermat's principle allows retrieving the zone radii and profile expressions. The Thin Element Approximation (TEA) enables calculating optical path differences (OPDs). The complex phasor $U(r, \lambda)$ associated with the mth zone of the kinoform diffractive lens depends on the expression of the phase delay introduced by the DOE:

$$\begin{cases} \Phi(r,\lambda) = \dfrac{2\pi}{\lambda} OPD(r,\lambda) = \dfrac{2\pi}{\lambda} z(r)(n_\lambda - ne_\lambda) \\ U_m(r,\lambda) \equiv \exp\left[i\Phi(r,\lambda)\right] = \exp\left[i2\pi\alpha\left(m - \dfrac{r^2}{r_1^2}\right)\right] \\ \alpha \equiv \dfrac{\lambda_0(n_\lambda - ne_\lambda)}{\lambda(n - n_e)} \end{cases} \tag{8.12}$$

where the radius $r \in [r_m, r_{m+1}[$, $r_1^2 = 2f_0\lambda_0$. The refractive index n_λ and ne_λ are defined for any wavelength λ whereas n and n_e are defined at the specific "design" wavelength λ_0. The parameter α is the fraction of 2π phase delay introduced for wavelengths other than the design wavelength λ_0.

The phasor U_m is a complex exponential function from which an approximate analytical expression can be retrieved using a Fourier expansion:

$$\begin{cases} U_m(r,\lambda) = \sum\limits_{j=-\inf}^{\inf} C_j \exp\left[i\frac{2\pi j}{r_1^2}r^2\right] \\ C_j \equiv \frac{1}{r_1^2}\int\limits_0^{r_1^2} \exp\left[i2\pi\alpha\left(m - \frac{r^2}{r_1^2}\right)\right] \exp\left[i\frac{2\pi j}{r_1^2}r^2\right] \mathrm{d}r^2 \end{cases} \tag{8.13}$$

The energy content (C_j) and its location depend on the integer j, called "diffractive order". The resulting field is divided into an infinite number of spatially separated orders. Rearranging Equation 8.13 and solving the integral lead to:

$$\begin{cases} U_m(r,\lambda) = \sum\limits_{j=-\inf}^{\inf} C_j \exp\left[\frac{-ikr^2}{2\lambda_0 f_0/(j\lambda)}\right] \\ C_j = \exp\left[i2\pi\alpha m\right]\frac{\exp\left[2\pi(\alpha+j)\right]-1}{2\pi(\alpha+j)} \end{cases} \tag{8.14}$$

The phasor (or complex transmittance) expression is interesting since it encompasses the DOE behaviour. Each order j is associated with a focal length f_j and with an energy content η_j, called "diffraction efficiency", expressed as [50]:

$$\begin{cases} \eta(j,\lambda) \equiv |C_j|^2 = \mathrm{sinc}^2(\alpha - j) \\ f_j(\lambda) = \frac{\lambda_0}{\lambda}\frac{f_0}{j} \end{cases} \tag{8.15}$$

To have positive values of j corresponding to converging orders, we change j to $-j$ in Equation 8.14 and reverse the order of summation. If we compare the complex transmittance obtained for the DOE and the thin parabolic refractive lens (Equation 8.5) displayed in Figure 8.1. We observe that the DOE acts like a refractive lens with an infinite number of foci f_j, each associated with a certain amount of energy. The parameter α, called the "design parameter", defines the energy content for each order. The plot of diffraction efficiency in the IR band (4.4 - 12 µm) is shown in Figure 8.4

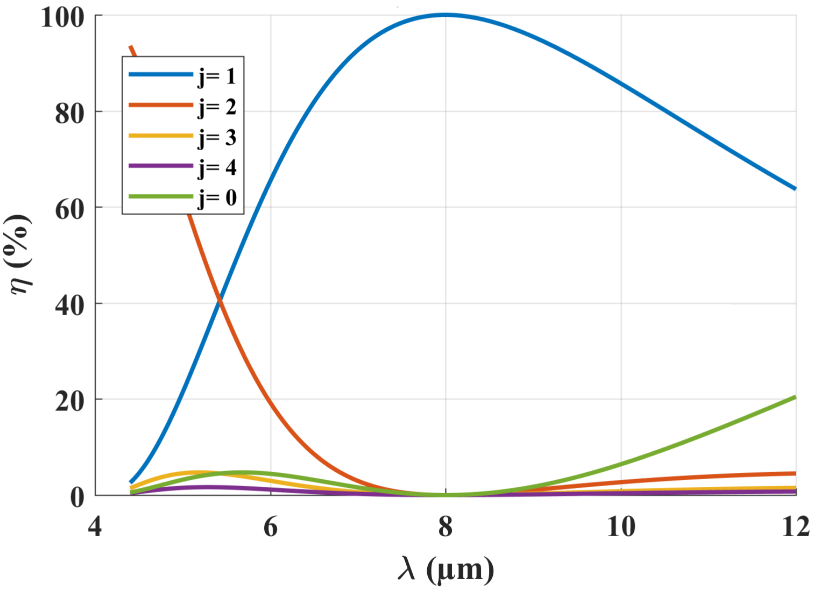


Figure 8.4: Diffraction efficiency for a ZnS kinoform diffractive lens in the IR band. The design wavelength is $\lambda_0 = 8$ μm, for which the diffraction efficiency reaches 100% for the first order (j=1). The second order is dominant in MWIR.

According to Figure 8.4, a kinoform DOE is unsuitable for a dual-band MWIR-LWIR design. Its diffraction efficiency can be optimized for a specific wavelength and order but quickly drops for other wavelengths. In this example, The diffraction efficiency in only 60% at 12 μm and less than 20% between 4 and 5 μm. It has been shown [63] that the modulation transfer function depends on the diffraction efficiency, meaning that bad imaging performances are expected when the diffraction efficiency is low. In addition, a bad efficiency for order 1 also means an increased stray light caused by the other diffractive orders (especially orders 0 and 2).

8.2.2 Harmonic diffractive optical element

A harmonic diffractive lens (HDOE) introduces a phase shift of $2p\pi$, $p \in \mathbb{Z}$, between adjacent zones. The integer p is called the harmonic parameter of the DOE. This generalized version of the 2π modulo diffractive lens follows the same reasoning as in Section 8.2 with slightly different results. HDOEs are detailed in [64, 65]. As shown in Figure 8.5, the main difference is that their profile heights are p times larger than the classical kinoform lens.

Following the generalization to a $2p\pi$ phase shift, the zone radii (Equation 8.6) becomes:

$$r_m^2 = 2pm\lambda_0 f_0 \tag{8.16}$$

The optical phase introduced by the harmonic element is written:

$$\Phi(r,\lambda) = 2\pi\alpha p(m - \frac{r^2}{r_1^2}) \tag{8.17}$$

where α is the "design" parameter defined in the previous section as $\alpha = \frac{\lambda_0}{\lambda}\left(\frac{n_\lambda - ne_\lambda}{n_0 - ne_0}\right)$.
As already said, the maximal height of the HDOE is increased:

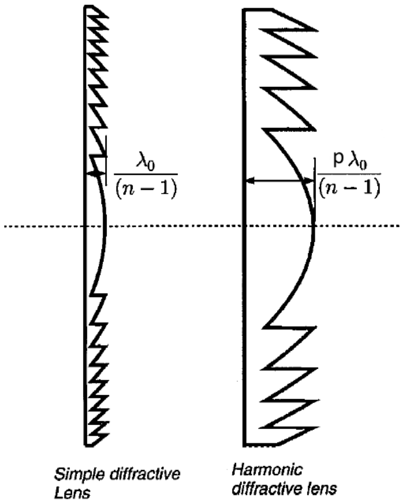


Figure 8.5: Harmonic diffractive lens profile (right) [65]. The HDOE is p times larger than the diffractive lens (left) because it introduces a $2p\pi$ phase shift between adjacent zones.

$$H_{max} = p\frac{\lambda_0}{n_0 - ne_0} \tag{8.18}$$

With the same reasoning as in the previous section, the focal length becomes:

$$f_j(\lambda) = p\frac{\lambda_0}{\lambda}\frac{f(\lambda_0)}{j} \tag{8.19}$$

With j the considered diffractive order. Finally, HDOEs diffraction efficiency is:

$$\eta(j, \lambda, p) = \text{sinc}^2(p\alpha - j) \tag{8.20}$$

The diffraction efficiency maxima now occur at wavelengths defined by:

$$\lambda_{peak} = \frac{p\lambda_0}{j} \tag{8.21}$$

The diffraction efficiency can now be tuned for higher working orders, meaning that the optical power of the HDOE can be much higher. However, the higher the operating order, and the thinner the operating waveband, as shown in Figure 8.6.

Figure 8.6 shows that the diffraction efficiency keeps its peak value of 100% for the design wavelength, even at high orders. However, the operating waveband progressively reduces with the diffractive order. The denomination "operating waveband" represents a waveband where the diffraction efficiency keeps a relatively high value.

Interestingly, for a single harmonic parameter, multiple wavelengths can reach 100% efficiency (Equation 8.21) when the considered order varies, as illustrated in Figure 8.7.

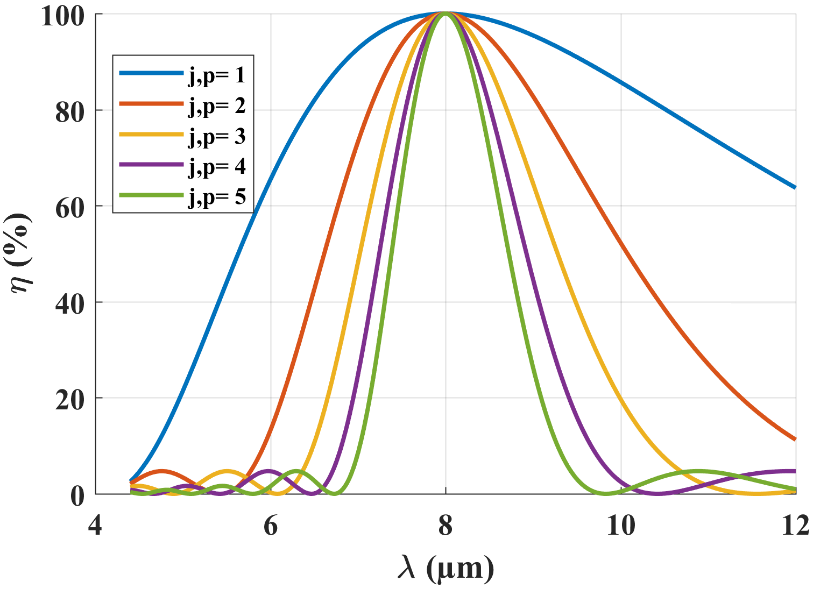


Figure 8.6: Diffraction efficiency of a ZnS harmonic kinoform diffractive lens (HDOE) working in the IR waveband. The design wavelength is $\lambda_0 = 8$ µm, for which the diffraction efficiency can reach 100% for various orders ($j = p$). p is the harmonic parameter.

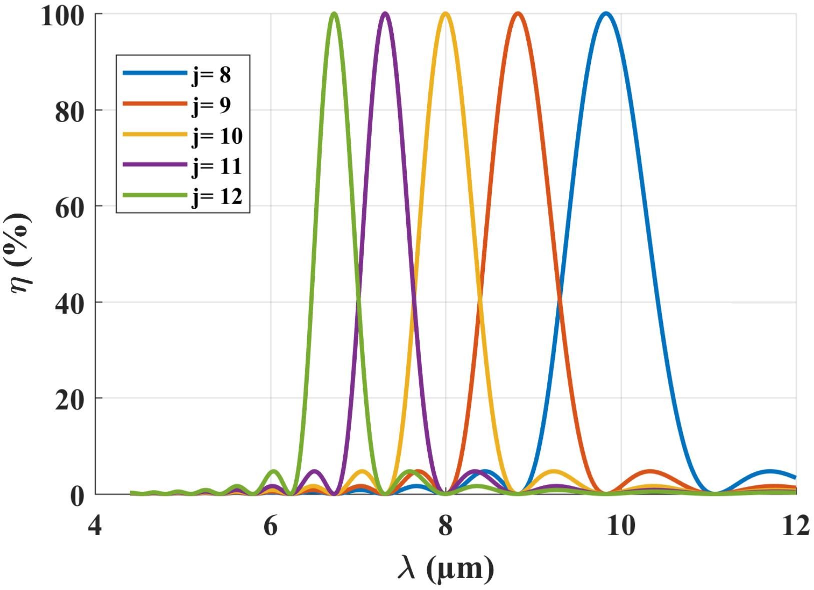


Figure 8.7: Diffraction efficiency for a ZnS harmonic kinoform diffractive lens (HDOE) working in the IR waveband. The design wavelength is $\lambda_0 = 8$ µm and the harmonic parameter $p = 10$. The diffraction efficiency can reach 100% for various other wavelengths when the order changes due to Equation 8.21.

This result is more interesting than the previous one in our case because the number of efficient wavelengths increases. However, in between those wavelengths, the diffraction efficiency drops drastically because the harmonic parameter and the orders are high. Yet, let's recall that in our case, the atmospheric gap between 5 µm and 8 µm is opaque; thus, this waveband does not require high diffraction efficiency. In addition, if the order and harmonic parameter are kept sufficiently small, the diffraction efficiency drop softens. Figure 8.8 describes a setup providing an efficient dual-band imaging solution:

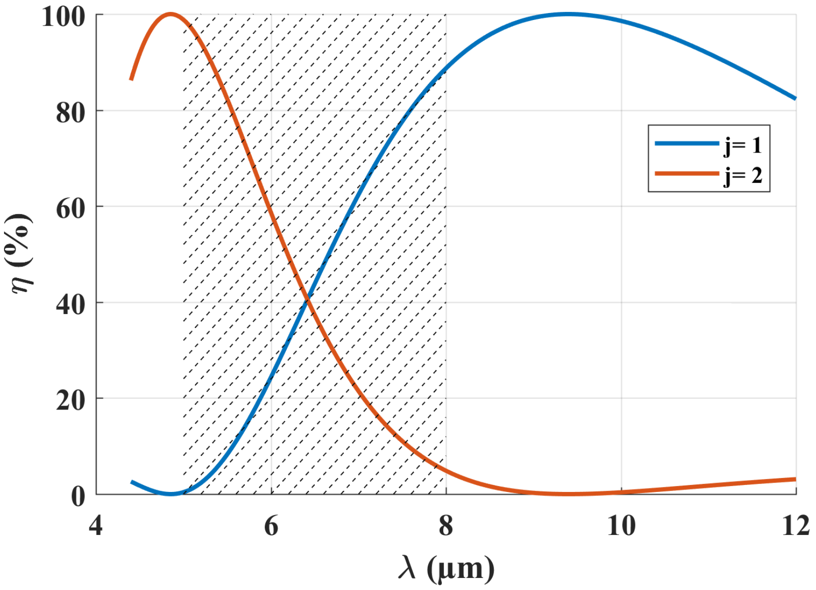


Figure 8.8: Diffraction efficiency for a ZnS harmonic kinoform diffractive lens (HDOE) working in the IR waveband. The dashed lines represent the opaque atmospheric window. The diffraction efficiency keeps a high value in both interesting bands (LWIR and MWIR). $\lambda_0 = 9.4$ µm and $p = 1$.

Figure 8.8 setup allows to obtain the maximal diffraction efficiency and the same focal point for two distinct wavelengths: The first order is designed with $\lambda_1 = 9.4$ µm, therefore the second order keeps the same maximal diffraction efficiency and focal point at $\lambda_2 = 9.4/2 = 4.7$ µm (Equation 8.21). Section 8.3 further studies this interesting "multi-order" design in the frame of hybrid refractive-diffractive doublets. Multi-order diffractive optical elements are called MODOEs in the following.

8.3 MODOE designs

In this section, we will study the imaging capabilities of multi-order diffractive optical elements (MODOEs). We have previously described the theoretical model of MODOEs and showed in Figure 8.8 that their diffraction efficiency remained high in both MWIR (4.4-5 µm) and LWIR (8-12 µm). Achieving high dual-band diffraction efficiency is possible using a particular setup: The first order is selected in LWIR, while the second is selected in MWIR since MWIR wavelengths are half the LWIR wavelength. We recall the focal law of such lens (Equation 8.19):

$$f_j(\lambda) = p\frac{\lambda_0}{\lambda}\frac{f(\lambda_0)}{j}$$

For the following we choose $\lambda_M = 4.7$ µm and $\lambda_L = 9.4$ µm. To have a closer look at the behaviour of the MODOE along the optical axis, a Fourier optics simulation (see Chapter 4) is performed with an F/10 Ge MODOE with an aperture diameter $D = 10$ mm (focal length of 100 mm at $\lambda = 8$ µm). Figure 8.9 displays the irradiance along the optical axis for multiple wavelengths:

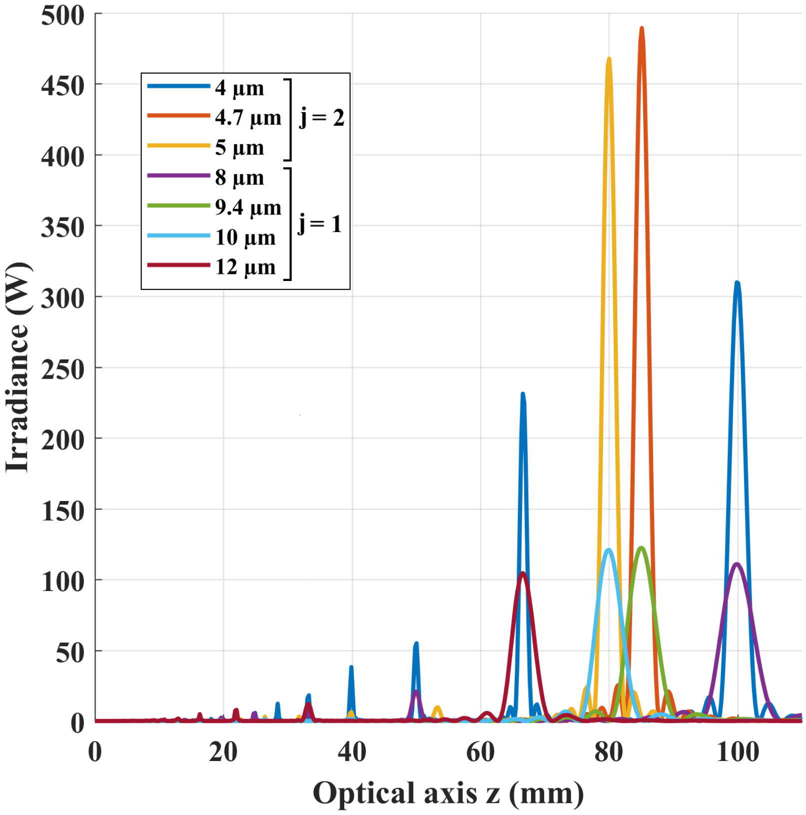


Figure 8.9: Irradiance along the optical axis for various wavelengths. The focal length is $f = 100$ mm for $\lambda = 8$ µm meaning that MWIR wavelengths ($\lambda < 8$ µm) have no visible focus in this Figure, for the order $j = 1$. Therefore, the curves for these wavelengths start with the order $j = 2$.

Figure 8.9 depicts the evolution of the irradiance along the optical axis using Fourier optics. The MWIR wavelengths appear only for orders $j > 1$ since the graph does not go beyond the focal length $f = 100$mm, defined for the LWIR wavelength $\lambda = 8$ µm. For instance, the blue curve on the right ($\lambda = 4$ µm, $j = 2$) shares the focus of the purple curve ($\lambda = 8$ µm, $j = 1$). The second blue curve from the right ($\lambda = 4$ µm, $j = 3$), is located at $z = 66.6$ mm, at the same position than the first order of the $\lambda = 12$ µm wavelength.

The other point of interest for this design lies in the "focalization efficiency". This measure is similar to the diffraction efficiency and uses the Strehl ratio for various wavelengths. Since each wavelength has a distinct focal plane, the "best" focal plane is considered for each wavelength when computing the Strehl ratio. The result is shown in Figure 8.10.

As expected, Figure 8.10 illustrates the drastic image quality drop inside the opaque atmospheric window waveband. On the contrary, the Strehl ratio (also called "focalization efficiency") keeps a high value (above 85% and even 90%) in both MWIR and LWIR wavebands. It means that foci corresponding to unwanted diffractive orders receive little energy. However, for a HDOE, the focal length variation is very strong due to the $1/\lambda$ "focal law" (Equation 8.19). Figure 8.11 studies the MODOE chromatic focal shift, also called the longitudinal chromatic aberration (LCA).

Figure 8.11 shows that the LCA is discontinuous, with a sharp shift at $\lambda = 6$ µm. We could anticipate this result since diffractive orders are discrete integers. This discontinuity is a huge drawback for this design since it will result in hybrid refractive-diffractive combi-

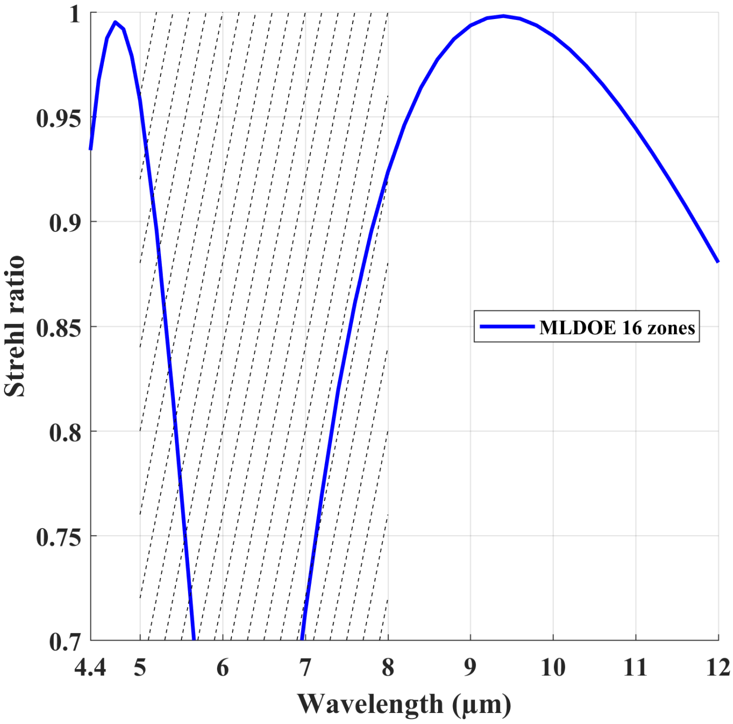


Figure 8.10: Strehl ratio at the best focus for various wavelengths. The second diffractive order is considered for $\lambda < 6$ µm. The grey dashed zone corresponds to the atmospheric opaque window, where no imaging is needed. The best focus is computed for each wavelength, so the chromatic aberration is not depicted in this figure.

nations with chromatic and thermal defocuses that cannot be compensated (see Chapter 7).

In conclusion, the research for a compact dual-band infrared component has led to investigating diffractive optical elements. Their intrinsic diffractive nature combines exceptionally well with refractive lenses, providing efficient achromat doublets (i.e. a high $F/\sharp$ DOE can compensate for the LCA of a refractive element). However, diffractive lenses classically operate in a relatively short waveband. We have therefore leaned towards MODOEs, which, in the case of a dual MWIR-LWIR waveband, operate efficiently thanks to their multiple orders. In exchange, their discontinuous LCA might be tough to couple with other optics, limiting their use. In the following chapter, we further investigate the multilayer diffractive optical element, which is a diffractive dual-band solution, extending the desired behaviour of diffractive lenses to multiple wavebands.

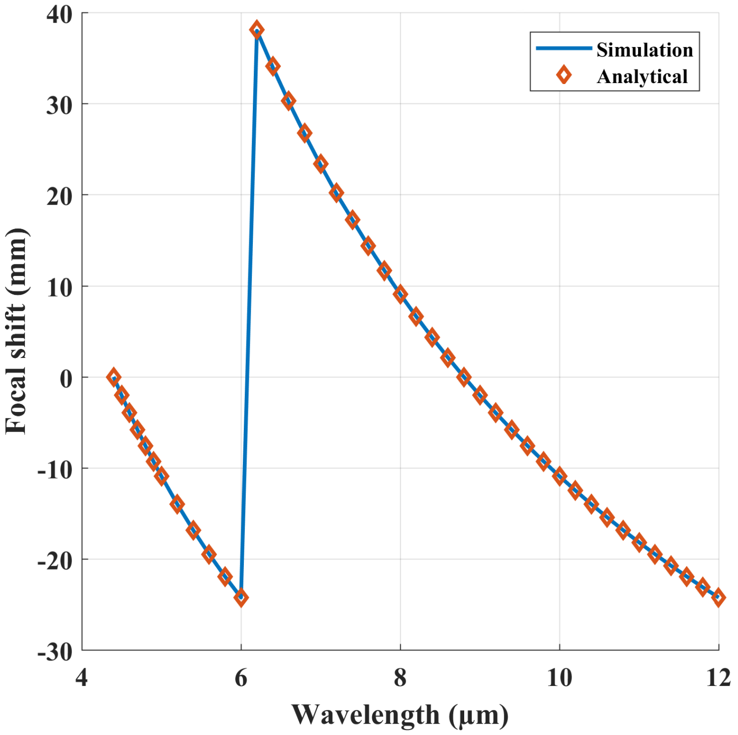
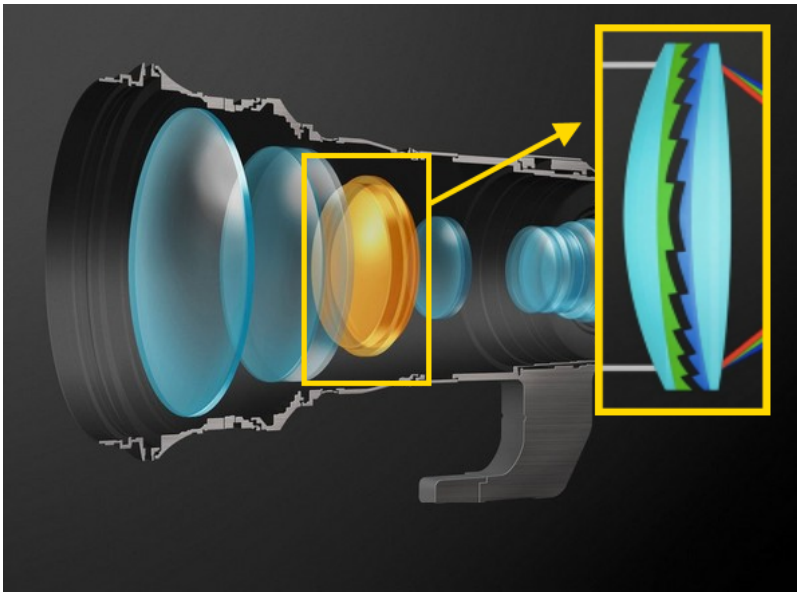


Figure 8.11: Longitudinal chromatic aberration (LCA) obtained with a MODOE lens. The sharp discontinuity is due to the diffraction order discontinuity ($j = 2$ for $\lambda < 6$ µm and $j = 1$ for $\lambda \geq 6$ µm). The analytical points are calculated with Equation 8.19.



Chapter 9

MLDOE design based on the TEA

This chapter follows the DOE designing method described in Chapter 8 and applies it in the case of multilayer diffractive optical element (MLDOE). Contrary to Chapter 8 designs, an MLDOE is theoretically an ideal dual-band diffractive solution for hybrid systems, like in Chapter 7. The thin element approximation (TEA) provides an MLDOE shape and phase expressions. Increasing the number of layers and materials increases the number of design wavelengths. As a result, the diffraction efficiency can reach its maximum for two distinct wavebands instead of one for classic DOEs. This analytical result complies with the multiple performance studies made using Fourier optics. This formalism allows the evaluation of the irradiance, Strehl ratio and chromatic focal shift at any plane perpendicular to the optical axis. Finally, the validity of the thin element approximation is discussed based on bibliographic references. Considering thick MLDOEs, the TEA applicability is limited to very high F-numbers and particular materials. The MLDOE design and results of this chapter have been published in [66]. The TEA validity study has been published in [67].

Contents

9.1 Introduction

Diffractive lenses have been studied in Chapter 8 under the scalar theory of diffraction (SDT) based on the Thin Element Approximation (TEA). This theory is analytical and very accurate for thin lenses. We concluded that diffractive lenses (DOEs) could reach high diffraction efficiencies but only for small wavebands. However, considering the first and second diffractive orders, these types of lenses, called multi-order diffractive optical elements (MODOEs), could achieve high diffraction efficiency for both MWIR and LWIR wavebands. Nevertheless, using multiple orders also generates a discontinuous chromatic aberration, which is challenging to correct with refractive elements. Therefore, MODOEs are unsuitable for wide-band hybrid designs, raising the need for another dual-band diffractive solution.

This chapter focuses on developing a wide-band diffractive solution with high diffraction efficiency while maintaining a continuous chromatic focal shift, easily corrected by hybrid doublets. The constructive interferences requirement is imposed for a single "design" wavelength causing a steep DOE diffraction efficiency drop. The diffraction efficiency (for the 1^{st} order) can only be maximal for this wavelength. We aim to increase the degree of freedom by employing a second "design" wavelength. Doing so will undoubtedly reshape the diffraction efficiency for two distinct wavelengths. As seen in Chapter 8, one diffractive layer can only operate (constructive interferences) for only one wavelength, so the solution is to add a diffractive layer to the element. This idea of extending a DOE to a multilayer diffractive optical element (MLDOE) was proposed by Y. Arieli [68, 69], who put two binary DOEs side by side and studied the resulting diffraction efficiency. The same idea is used in this chapter, using two harmonic DOEs instead of binary DOEs. In short, the MLDOE is an extension of the DOE where multiple "design" wavelengths are considered.

9.2 MLDOE design based on the TEA

The designing steps and equations of this Section are inspired by the work of V. Moreno *et al* [50] and have been published in [66].

9.2.1 Diffractive zone profile determination

MLDOEs are DOE extensions, utilizing two "design" wavelengths instead of one. Therefore, the diffractive zone profile of each layers must be optimised for two wavelengths. In this Section, we extend the work of V. Moreno *et al* [50], supposing that constructive interferences occur for two wavelengths. The diffractive zone shape design relies on three requirement/principle/approximation:

- **Constructive interferences**: The diffractive element must provide constructive interferences at the focal plane for two distinct "design" wavelengths $\lambda_d = \lambda_1 | \lambda_2$.
- **Fermat's principle**: All the rays coming together in a determined zone, even though they travel on distinct paths, result in having an equal optical path length (OPL) to a fixed point F, placed on the axis at the focal plane of the MLDOE.
- **Thin element approximation (TEA)**: The OPLs are computed assuming an infinitely thin element. The rays travel through the MLDOE with no deviation, parallel to the optical axis. This approximation's validity is discussed in Section 9.4.

Figure 9.1 shows an MLDOE's geometrical layout, assuming a converging-diverging layer configuration. The calculations that follow, while done with this particular setup, can be effortlessly adapted to other configurations (converging-converging, diverging-diverging, diverging-converging, see Section 9.5).

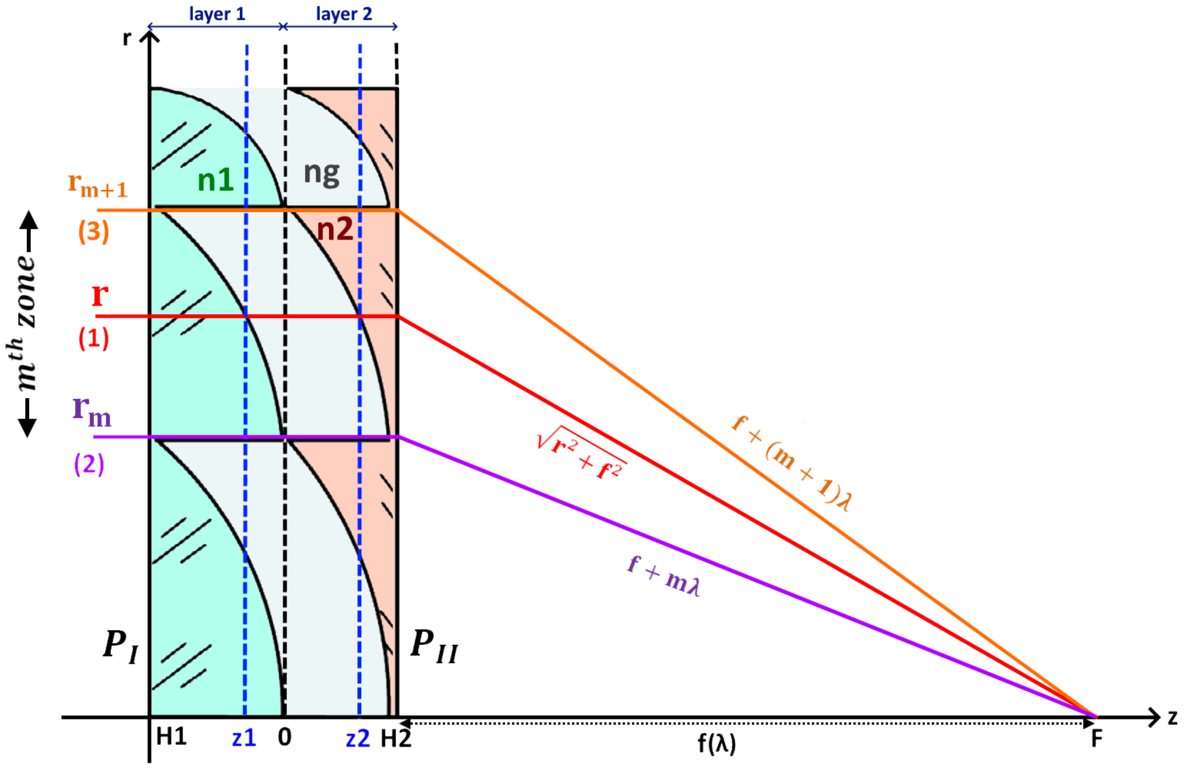


Figure 9.1: Extended model of DOE to multilayer DOE. The Thin Element Approximation (TEA) implies a paraxial incident beam. The shape of layer $i = 1, 2$ along the optical axis z at a radius r is described by $z_i(r)$. H_i is the maximal height of the layer i, with the following convention for a converging diverging MLDOE: $H_1 < 0$ and $H_2 > 0$. $f(\lambda)$ is the focal distance of the MLDOE at the wavelength λ and m is the considered diffractive zone, starting at a radius r_m and finishing at r_{m+1}. Each layer i is made of a distinct material with refractive index $n_i(\lambda)$ and the gap is filled with a material $n_g(\lambda)$. Interferences between each zone are constructive at the focal plane.

Both layers must be made of distinct materials to guarantee that two wavelengths can interfere constructively at the focal plane. The gap between each layer is potentially filled with air or a third IR material. We are currently unable to discuss which materials should be combined to make a "good" MLDOE. This selection material topic is treated for the visible and infrared bands in [70] and will be detailed in Chapter 11.

For obvious geometrical reasons, we only consider the case of both kinoform layers having the same number of zones. Therefore, the mth diffractive zone has the same period (radial dimension) regardless of the layer but a profile height that can vary depending on the considered layer.

In order to have constructive interferences at the focal plane at a design wavelength $\lambda_d = \lambda_1|\lambda_2$, the optical path difference (OPD) between two consecutive zones must be λ_d. This recursive constraint gives the radius of the m^{th} zone, using the Pythagorean theorem in Figure 9.1:

$$r_m^2 = 2mf_d\lambda_d + (m\lambda_d)^2 \simeq 2mf_d\lambda_d \tag{9.1}$$

where f_d is the focal length of the MLDOE defined at one of the design wavelengths λ_d. This expression of the zone radius is the same as for standard DOE (Equation 8.6). In Figure 9.1, the origin is taken between both layers; therefore, the optical path length (OPL) is counted negatively inside the first layer and positively inside the second one. The general OPL associated with a ray incident at a radius $r \in [r_m, r_{m+1}[$ (ray (1) in Figure 9.1) can be expressed as:

$$OPL(r) = -n_1\left[H_1 - z_1(r)\right] + z_1(r)n_g + z_2(r)n_g + n_2\left[H_2 - z_2(r)\right] + \sqrt{f_d^2 + r^2} \tag{9.2}$$

z_i, H_i, n_i are respectively the profile height at radius r, the maximal profile height and the refractive index of layer i, as shown in Figure 9.1. n_g is the refractive index of the gap-filling material. Using the ray at the start of the m^{th} zone (ray (2)) provides a particular case of Equation 9.2:

$$\begin{cases} z_1(r_m) = 0 \\ z_2(r_m) = H2 \\ OPL(r = r_m) = -n_1H_1 + n_gH_2 + f_d + m\lambda_d \end{cases} \tag{9.3}$$

Therefore, the optical path difference (OPD) at any radius $r \in [r_m, r_{m+1}[$ is expressed by:

$$OPL(r) - OPL(r = r_m) = (n_g - n_1)z_1 + (n_2 - n_g)(H_2 - z_2) + \sqrt{f_d^2 + r^2} - f_d - m\lambda_d \tag{9.4}$$

The square root is approximated by its second limited expansion term:

$$\sqrt{f_d^2 + r^2} = f_d\left(1 + \frac{r^2}{2f_d^2}\right) = f_d + \frac{r^2}{2f_d} = f_d + \frac{r^2}{r_1^2}\lambda_d \tag{9.5}$$

r_1 is the starting radius of the first zone, defined by Equation 9.1: $r_1^2 = 2f_d\lambda_d$. Finally, we have:

$$OPL(r) - OPL(r = r_m) = (n_2 - n_g)(H_2 - z_2(r)) + z_1(r)(n_1 - n_g) - \lambda_d\left(m - \frac{r^2}{r_1^2}\right) \tag{9.6}$$

A particular case arises when $r = r_{m+1}$ (ray (3) in Figure 9.1):

$$\begin{cases} z_1(r_{m+1}) = H_1 \\ z_2(r_{m+1}) = 0 \\ OPL(r_{m+1}) - OPL(r_m) = (n_1 - n_g)H_1 + (n_2 - n_g)H_2 + \lambda_d \end{cases} \tag{9.7}$$

The optical path lengths and differences have been geometrically computed (Equations 9.6 and 9.7) as consequences of both the constructive interferences requirement and the TEA model. Applying Fermat's principle to the m^{th} zone provides a system of 2 equations:

$$\begin{cases} OPL\,(r \in [r_m, r_{m+1}[) - OPL(r_m) = 0 \\ OPL(r_{m+1}) - OPL(r_m) = 0 \end{cases} \tag{9.8}$$

System 9.8 has 2 equations and 4 unknowns: $z_1(r), z_2(r), H_1$ and H_2. However, each OPD equation in System 9.8 can be expressed for the two design wavelengths $\lambda_d = \lambda_1|\lambda_2$ providing 4 equations in total:

$$\begin{cases} OPL(r,\lambda_1) = OPL(r_m,\lambda_1) \\ OPL(r,\lambda_2) = OPL(r_m,\lambda_2) \\ OPL(r_{m+1},\lambda_1) = OPL(r_m,\lambda_1) \\ OPL(r_{m+1},\lambda_2) = OPL(r_m,\lambda_2) \end{cases} \tag{9.9}$$

The design wavelengths λ_1 and λ_2 are respectively associated with the focal lengths f_1 and f_2, and with the refractive indexes n_{11}, n_{g1}, n_{21} and n_{12}, n_{g2}, n_{22}. n_{ij} describes the index of layer i with respect to wavelength λ_j $(i, j = 1, 2)$ and n_{gj} is the filling material index at wavelength λ_j. For clarity, we define the following auxiliary parameters $(i = 1, 2)$:

$$\begin{cases} A_i = n_g(\lambda_i) - n_1(\lambda_i) \\ B_i = n_2(\lambda_i) - n_g(\lambda_i) \end{cases} \tag{9.10}$$

Solving System 9.9 provide a unique solution for $z_1(r), z_2(r), H_1$ and H_2, for all $r \in [r_m, r_{m+1}]$:

$$\begin{cases} H_1 = \dfrac{-\lambda_1 B_2 + \lambda_2 B_1}{A_1 B_2 - A_2 B_1} \\ H_2 = \dfrac{-\lambda_1 A_2 + \lambda_2 A_1}{A_1 B_2 - A_2 B_1} \\ Z_1 = -H_1 \left(m - \dfrac{r^2}{r_1^2} \right) \\ Z_2 = +H_2 \left(m + 1 - \dfrac{r^2}{r_1^2} \right) \end{cases} \tag{9.11}$$

Recall that r_1 is defined using relation 9.1: $r_1^2 = 2f_1\lambda_1 = 2f_2\lambda_2$. System 9.11 is only applicable when the configuration of the MLDOE follows the layout of Figure 9.1: converging-diverging (i.e. $H_1 < 0$ and $H_2 > 0$). Depending on the choice of materials, this configuration can change. A more general expression of the zone profiles and a configuration selection process is detailed in Section 9.5. Figure 9.2 depicts a summary of this process:

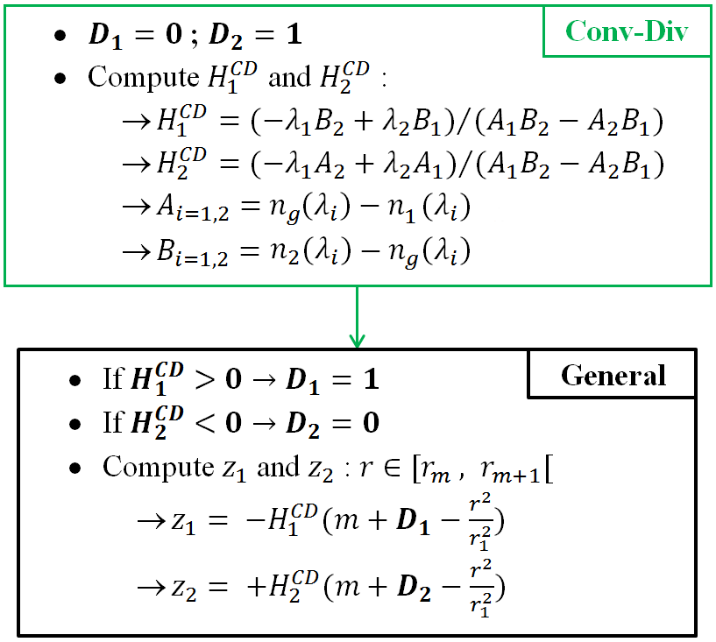


Figure 9.2: Summary of the MLDOE shape design process. The initial step consists in designing a converging-diverging configuration ($D_1 = 0$ and $D_2 = 1$): H_1^{CD} and H_2^{CD}, selecting two design wavelengths λ_1 and λ_2. If the signs of the converging-diverging zones height H_1^{CD} and H_2^{CD} are wrong, reverse the booleans ($\mathcal{D}_{1,2} \leftarrow 1 - \mathcal{D}_{1,2}$). The profile $z(r)$ is detailed in System 9.40.

Compared to standard HDOEs, we can see that the zone shapes are still parabolic and that only the maximal height expression changes. We recall that for a standard HDOE, the maximal zone height (Equation 8.18) is:

$$Z_{max} = \frac{p\lambda_d}{n(\lambda_d) - n_e(\lambda_d)} \tag{9.12}$$

p is the harmonic parameter. For comparison, we consider a ZnS HDOE and a ZnS-air-Ge MLDOE, with $\lambda_1 = 10.4$ µm in LWIR and $\lambda_2 = 4.7$ µm in MWIR. The ZnS HDOE has a height of 8.7 µm ($p = 1$), while the ZnS-air-Ge MLDOE has a total height of 173 µm. It means that the harmonic parameter p of the ZnS HDOE should be around 20 to match the height of the ZnS-air-Ge MLDOE. This simple example shows that one main characteristic of MLDOEs is their"high thickness.

9.2.2 MLDOE analytical phase and diffraction efficiency

For any wavelength λ, the optical path difference inside an MLDOE at any radius $r \in [r_m, r_{m+1}[$ is expressed using Equations 9.6 and 9.11:

$$\begin{aligned} OPD(r,\lambda) &= (n_g - n_1)z_1 + (n_2 - n_g)(H_2 - z_2) \\ &= [A(\lambda)H_1 - B(\lambda)H_2]\left(m - \frac{r^2}{r_1^2}\right) \end{aligned} \tag{9.13}$$

Recall that $A(\lambda) = n_g(\lambda) - n_1(\lambda)$ and $B(\lambda) = n_2(\lambda) - n_g(\lambda)$. The phase delay generated by an MLDOE is therefore expressed as:

$$\begin{aligned}\Phi(r,\lambda) &= \frac{2\pi}{\lambda} OPD(r,\lambda)\\ &= \frac{2\pi}{\lambda}\left(H_1A(\lambda) - H_2B(\lambda)\right)\left(m - \frac{r^2}{r_1^2}\right)\end{aligned} \tag{9.14}$$

As for any DOE, an MLDOE phase can be separated into a spatial component and a spectral component $\Phi_0(\lambda)$:

$$\Phi(r,\lambda) = \Phi_0(\lambda)\left(m - \frac{r^2}{r_1^2}\right) \tag{9.15}$$

where $\Phi_0(\lambda)$ depends only on the wavelength and is defined by:

$$\Phi_0(\lambda) \equiv \frac{2\pi}{\lambda}\left(H_1A(\lambda) - H_2B(\lambda)\right) \tag{9.16}$$

This equation is the extension of the spectral behaviour of DOEs (see Equation 8.12). The operating wavelengths are included in the expression of the profile height in both MLDOE and DOE cases. The MLDOE complex transmission function is therefore expressed as:

$$T_{MLDOE}(r,\lambda) = \exp\left[i\Phi(r,\lambda)\right] = \exp\left[i\Phi_0(\lambda)\left(m - \frac{r^2}{r_1^2}\right)\right] \tag{9.17}$$

This expression is similar to the complex transmittance of DOEs, demonstrated in Chapter 8 (Equation 8.12). The Fourier expansion of the complex transmittance function gives, for a period $r_1^2 = 2f_1\lambda_1 = 2f_2\lambda_2$:

$$T_{MLDOE}(r,\lambda) = \sum_{j=-\inf}^{\inf} C_j \exp\left[i2\pi j\frac{r^2}{r_1^2}\right] \tag{9.18}$$

We retrieve the same "focal length law" than for standard DOEs: $f(\lambda,j)\lambda j = f_0(\lambda_0)\lambda_0$. This result will be demonstrated more rigorously in the next Section. The Fourier coefficient C_j is defined as:

$$\begin{aligned}C_j &= \frac{1}{r_1^2}\int_0^{r_1^2} \exp\left[i\Phi_0\left(m - \frac{r^2}{r_1^2}\right)\right]\exp\left[-i2\pi j\frac{r^2}{r_1^2}\right]\mathrm{d}r^2\\ &= \exp[i\Phi_0 m]\frac{1}{r_1^2}\int_0^{r_1^2}\exp\left[-i\left(\Phi_0 + 2\pi j\right)\frac{r^2}{r_1^2}\right]\mathrm{d}r^2\\ &= \exp[i\Phi_0 m]\frac{\exp[-i(\Phi_0 + 2\pi j)] - 1}{-i(\Phi_0 + 2\pi j)}\end{aligned} \tag{9.19}$$

Each j^{th} term in the Fourier expansion defines a diffraction order j and is associated with some energy content. The measure of this energy is, by definition, the diffraction efficiency η_j for the order j and is defined by:

$$\begin{aligned}
\eta_j(\lambda, j) &= C_j C_j^* = |C_j|^2 \\
&= 1 * \frac{|\exp[-i(\Phi_0 + 2\pi j)] - 1|^2}{(\Phi_0 + 2\pi j)^2} \\
&= \frac{2 - 2\cos(\Phi_0 + 2\pi j)}{(\Phi_0 + 2\pi j)^2} \\
&= \frac{4\sin^2[(\Phi_0 + 2\pi j)/2]}{4[(\Phi_0 + 2\pi j)/2]^2} \\
&= \text{sinc}^2\left(\frac{\Phi_0(\lambda)}{2\pi} + j\right)
\end{aligned} \tag{9.20}$$

Where $\text{sinc}(x) = \sin \pi x / \pi x$. The MLDOE was introduced as an extension of the DOE, which appears clearly in the expression of the phase. Whereas the DOE produces constructive interferences at the focal point for a single wavelength, the MLDOE does so for two distinct wavelengths. Based on the expression of H_1, H_2 defined in Equation 9.11, the spectral term Φ_0 can be developed for a given wavelength λ:

$$\begin{aligned}
\Phi_0(\lambda) &= \frac{2\pi}{\lambda}(H_1 A(\lambda) - H_2 B(\lambda)) \\
&= \frac{2\pi}{\lambda}\frac{(-\lambda_1 B_2 + \lambda_2 B_1)A(\lambda) - (-\lambda_1 A_2 + \lambda_2 A_1)B(\lambda)}{A_1 B_2 - A_2 B_1} \\
&= \frac{2\pi}{\lambda}\left(\lambda_1 \frac{-A(\lambda)B_2 + A_2 B(\lambda)}{A_1 B_2 - A_2 B_1} + \lambda_2 \frac{-A_1 B(\lambda) + A(\lambda) B_1}{A_1 B_2 - A_2 B_1}\right)
\end{aligned} \tag{9.21}$$

The MLDOE diffraction efficiency is expressed for $\lambda = \lambda_1, \lambda_2$, for the j^{th} diffractive order:

$$\begin{cases}
\eta_j(\lambda_1, j) = \text{sinc}^2\left(\frac{1}{\lambda_1}\left[\lambda_1 \frac{-A_1 B_2 + A_2 B_1}{A_1 B_2 - A_2 B_1} + \lambda_2 \frac{-A_1 B_1 + A_1 B_1}{A_1 B_2 - A_2 B_1}\right] + j\right) = \text{sinc}^2(-1 + j) \\
\eta_j(\lambda_2, j) = \text{sinc}^2\left(\frac{1}{\lambda_2}\left[\lambda_1 \frac{-A_2 B_2 + A_2 B_2}{A_1 B_2 - A_2 B_1} + \lambda_2 \frac{-A_1 B_2 + A_2 B_1}{A_1 B_2 - A_2 B_1}\right] + j\right) = \text{sinc}^2(-1 + j)
\end{cases} \tag{9.22}$$

Therefore, the theoretical diffraction efficiency reaches 100% at the design wavelengths (order $j = 1$). So far, in the thin element approximation (TEA) frame, we have demonstrated a method to design an MLDOE with optimal efficiency given two distinct materials and two design wavelengths. The full plot of the diffraction efficiency for different orders and wavelengths is shown in Figure 9.3 based on Equation 9.20.

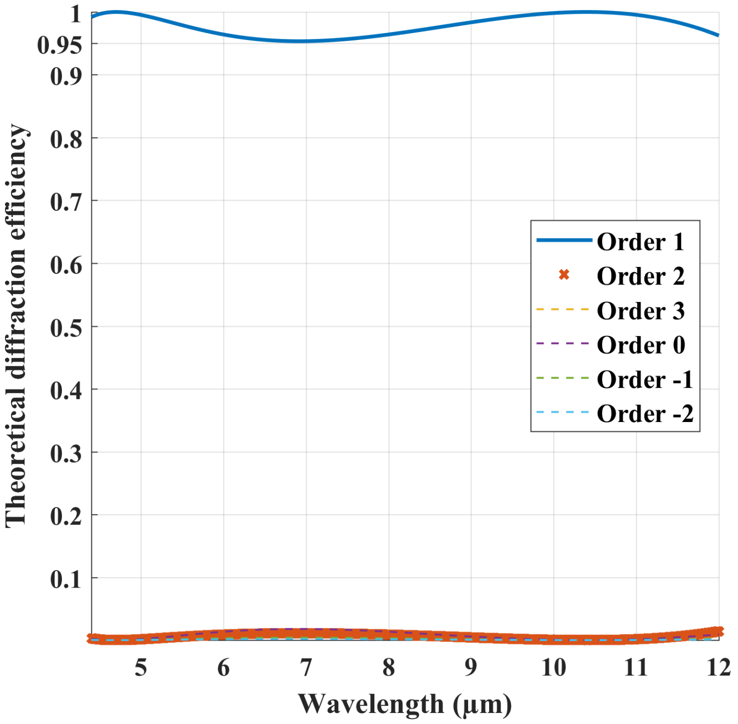


Figure 9.3: Diffraction efficiency for a ZnS-air-Ge MLDOE designed in MWIR and LWIR. The design wavelengths are at 4.7 µm and 10.4 µm. The orders refer to the global MLDOE diffractive order. The diffraction efficiency is maximal for both design wavelengths and stays above 95% in the considered infrared band (MWIR and LWIR), including the opaque atmospheric window (between 5 and 8 µm.

The MLDOE presented in Figure 9.3 is arbitrarily made of zinc selenide (ZnS) for the first medium and germanium (Ge) for the second. An air gap separates both materials. The optimal design wavelengths are 4.7 µm and 10.4 µm in MWIR and LWIR bands, respectively. The selection of these design wavelengths results from an optimisation: the integral of diffraction efficiency over all wavelengths (PIDE) is maximal for these two wavelengths. This optimization process is explained later in Section 9.2.3 and described in [71].

Figure 9.3 represents the evolution of the diffraction efficiency with the wavelengths. Only the first order operates, a strong requirement for an optical imaging component (monofocal property). Hence it is demonstrated that the MLDOE is a diffractive imaging element that could theoretically reach above 95% efficiency in the whole thermal infrared bandwidth for the first order.

9.2.3 Optimising the design wavelengths using the PIDE

It has been shown in the previous Section that an MLDOE design involves two (or three) distinct materials and two distinct design wavelengths, for which the diffraction efficiency is maximal. Nevertheless, MLDOEs aim to increase the diffraction efficiency for the whole MWIR/LWIR wavebands, not only for two wavelengths. Therefore, the design wavelength selection must follow an optimisation process that ensures that the diffraction efficiency for the first order is the highest possible, between 4.4 - 5 µm and 8 - 12 µm. No infrared imaging can be performed between 5-8 µm due to the opaque atmosphere. Inside this window, we do not require high diffraction efficiency. The most appropriate metric is the integral of the diffraction efficiency over the waveband, called PIDE (polychromatic integral diffraction efficiency).

The expression of the PIDE ($\overline{\eta}$) is obtained as follows [72]:

$$\overline{\eta_{\mathrm{j}}}(\lambda_{max}, \lambda_{min}) = \frac{1}{\lambda_{max} - \lambda_{min}} \int_{\lambda_{min}}^{\lambda_{max}} \mathrm{sinc}^2 \left(\frac{\Phi_0(\lambda)}{2\pi} + j \right) \mathrm{d}\lambda \tag{9.23}$$

According to Equation 9.23, one PIDE is computed in MWIR and one in LWIR. Both values are then weighted (50% - 50%) and added to obtain the final PIDE value. By doing so, the opaque atmospheric window is not accounted for in the optimisation process, leading to much higher diffraction efficiencies in the considered dual wavebands. For a given MLDOE material combination, the PIDE only depends on the chosen design wavelengths, and the goal is to maximise it.

We consider design wavelengths ranging from 4.4 - 5 µm and from 8 - 12 µm. For each couple of wavelengths, the PIDE is computed for the previously arbitrary selected ZnS-air-Ge MLDOE in Figure 9.4:

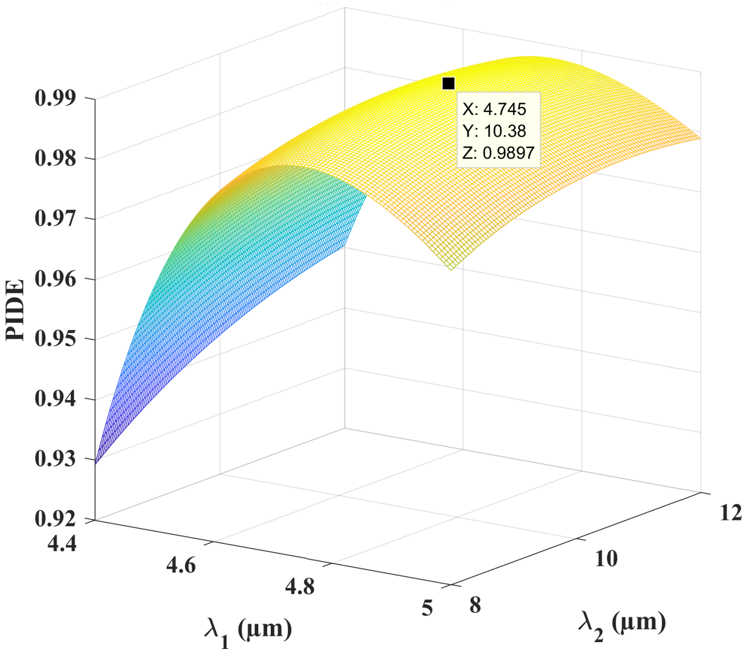


Figure 9.4: Variation of the Polychromatic integral diffraction efficiency (PIDE) with the design wavelengths for a ZnS-air-Ge MLDOE. The best design wavelengths are $\lambda_1 = 4.7$ µm and $\lambda_2 = 10.4$ µm.

The optimal design wavelengths are 4.7 µm and 10.4 µm. However, note that any design wavelengths close to the optimal ones give very close PIDE results. Therefore, this optimisation process does not need to be repeated for each MLDOE combination; the selected wavelengths can be kept regardless of the MLDOE and will remain 4.7 and 10.4 µm for the rest of the thesis.

9.2.4 Chromatic focal shift

In Section 9.2.2, the MLDOE broadband imaging capabilities have been demonstrated: the diffraction efficiency is above 95 % for the first order in both MWIR and LWIR bands. However, using different wavelengths generates a variation of the focal length called "chromatic focal shift", evaluated using the longitudinal chromatic aberration (LCA). The chromatic compensation of refractive optical elements remains the primary use of diffractive optical elements, including MLDOEs. Standard DOE focal length follows a $1/\lambda$ "focal law". Counter-intuitively, in the case of MLDOE, this relationship holds the same dependency with the wavelength. This result has been published in [66] and is demonstrated in the following. This section aims to express the MLDOE focal length $f(\lambda)$ for any given wavelength λ.

Recall that an MLDOE is composed of two harmonic diffractive optical elements (HDOEs) [64, 65], associated with the harmonic parameters p_1 and p_2, respectively. They are related to the MLDOE layers heights H_1 and H_2 [65] by:

$$\begin{cases} p_1(\lambda_d) = \dfrac{H_1}{\lambda_d}\left[n_1(\lambda_d) - n_g(\lambda_d)\right] = -\dfrac{H_1}{\lambda_d}A(\lambda_d) \\ p_2(\lambda_d) = \dfrac{H_2}{\lambda_d}\left[n_2(\lambda_d) - n_g(\lambda_d)\right] = \dfrac{H_2}{\lambda_d}B(\lambda_d) \end{cases} \tag{9.24}$$

The zone heights H_1 and H_2 have been defined in System 9.11. Recall that λ_d is a design wavelength, n_1, n_2 and n_g are the refractive indexes of the first, second and gap layers, respectively. p_1 and p_2 have a simple relation, using the same calculations as in Equations 9.21 and 9.22:

$$p_1(\lambda_d) + p_2(\lambda_d) = 1 \tag{9.25}$$

This equation confirms that the total operating MLDOE diffractive order is 1, similarly to converging DOE. We define $F_j^i(\lambda)$ the focal length of the harmonic layer $i = 1, 2$ for the diffractive order j at the wavelength λ. According to [65], $F_j^i(\lambda)$ follows the following relation:

$$\begin{cases} F_j^1(\lambda) = \dfrac{p_1 \lambda_d F_{p_1}^1(\lambda_d)}{j\lambda} = \dfrac{\lambda_d}{j\lambda}\dfrac{R^2}{2Np_1\lambda_d} \\ F_j^2(\lambda) = \dfrac{p_2 \lambda_d F_{p_2}^2(\lambda_d)}{j\lambda} = \dfrac{\lambda_d}{j\lambda}\dfrac{R^2}{2Np_2\lambda_d} \end{cases} \tag{9.26}$$

The focal length $F_{p_i}^i(\lambda_d)$ represents the operating focal length of layer i, when the diffractive order $j = p_i$ [65]. This focal length is related to the aperture radius R and the total number of zones N. Both layers can be considered thin lenses in contact. Therefore the total MLDOE optical power ($1/f_j$) is the sum of the optical powers of its harmonic layers. Combining Equations 9.25 and 9.26 leads to:

$$\frac{1}{f_j(\lambda)} = \frac{1}{F_j^1(\lambda)} + \frac{1}{F_j^2(\lambda)} = \frac{2Nj\lambda}{R^2} \tag{9.27}$$

Finally, using Equation 9.1, the MLDOE focal length for any order j is expressed as:

$$f_j(\lambda) = \frac{f_d \lambda_d}{j\lambda} \tag{9.28}$$

In conclusion, an MLDOE shares the same focal law and operating order as a standard DOE, with an increased diffraction efficiency in larger wavebands. This law benefits optical designs since shorter wavelengths are focused after longer wavelengths, contrary to refractive elements. Diffractive optics are mainly used in combination with refractive optics to correct the chromatic aberration [73]. For this reason, MLDOEs are extremely valuable for wide-band applications.

9.3 MLDOE Fourier Optics simulations

The author has published the simulation results of this section in [66]. Canon [74] has already realised a hybrid MLDOE zoom camera for the visible band. Designs involving MLDOEs have been realized [73, 75, 76] using ray tracing software such as CodeV or ZEMAX. However, these software do not accurately model the diffraction efficiency of complex diffractive elements. This section proposes a diffraction-based analysis of MLDOEs using the Fourier optics angular spectrum method, described in Chapter 4.

The Fourier Optics treatment of a monochromatic scalar field has already been presented in Figure 4.1. We recall that the angular spectrum method allows deriving the exact scalar field $U(x, y; z)$, based on a known initial field $U(x, y; 0)$, assuming that the propagating medium is "free space" (See Chapter 4). In this section, we use our Matlab Fourier optics propagator implementing the angular spectrum of plane waves, validated in Chapter 4. It allows us to accurately establish the field amplitude and phase at any position in space, based on the analytical phase shift given in Equation 9.14.

We finally use the same simulation parameters as in [66]: The simulation has $M = 1024$ uniformly distributed samples in each direction. The pupil diameter and the sampling window size are $D = 14$ mm and $S = 21$ mm, respectively. Having $D < S$ ensures that the edges of the computation window will not create and propagate numerical artefacts. The considered wavebands are: MWIR (4.4-5 µm) and LWIR (8-12 µm).

To ensure the validity of the thin element approximation (TEA), given the above parameters, the number of zones must be capped at five diffractive zones (F/43.8) [66]. Section 9.4 details the validity of this approximation. In short, this zone requirement ensures that the worst-case aspect ratio ($H/period_{min}$) is higher than 1/6, provided as the TEA validity limit in [77].

A ZnS-air-Ge MLDOE is designed at $\lambda_1 = 4.7$ µm and $\lambda_2 = 10.4$ µm. The total energy content of the field in the initial plane is normalised:

$$\iint_{aperture} |U(x, y; 0)|^2 \, \mathrm{d}x\mathrm{d}y = 1 \tag{9.29}$$

The aperture function $W(x, y)$ is a circular pupil of diameter D, leaving a constant uniform field amplitude distribution:

$$|U(x, y; 0)| = \frac{1}{\sqrt{\pi(D/2)^2}} \tag{9.30}$$

The angular spectrum method is conservative, the total energy of the field at any propagation plane remaining the same. The phase (modulo-2π) of the field in the initial plane $z = 0$ mm is shown in Figure 9.5:

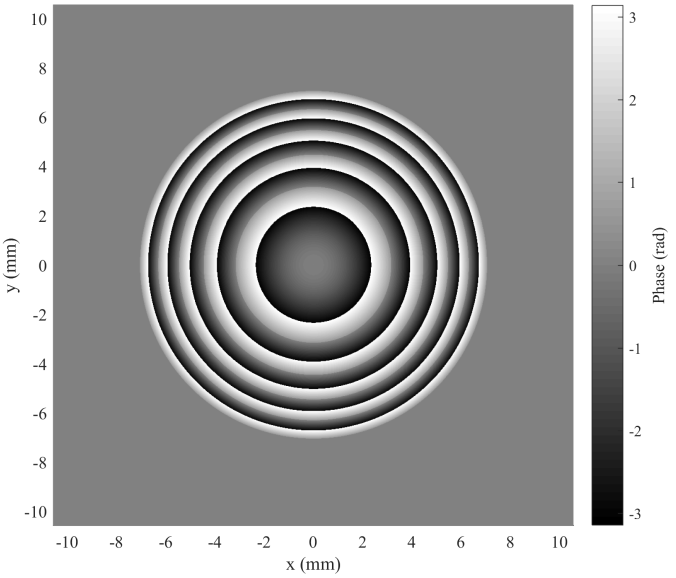


Figure 9.5: Phase modulo-2π of the input field in the initial plane $z = 0$ mm, $\lambda = 8$ µm. The phase represents the total phase shift caused by the MLDOE on a paraxial incident plane wave. The phase is calculated based on Equation 9.14 using the TEA.

The phase delay applied by the MLDOE follows the parabolic kinoform profile (Equation 9.14) inside the pupil and is null outside.

9.3.1 Image at the target plane

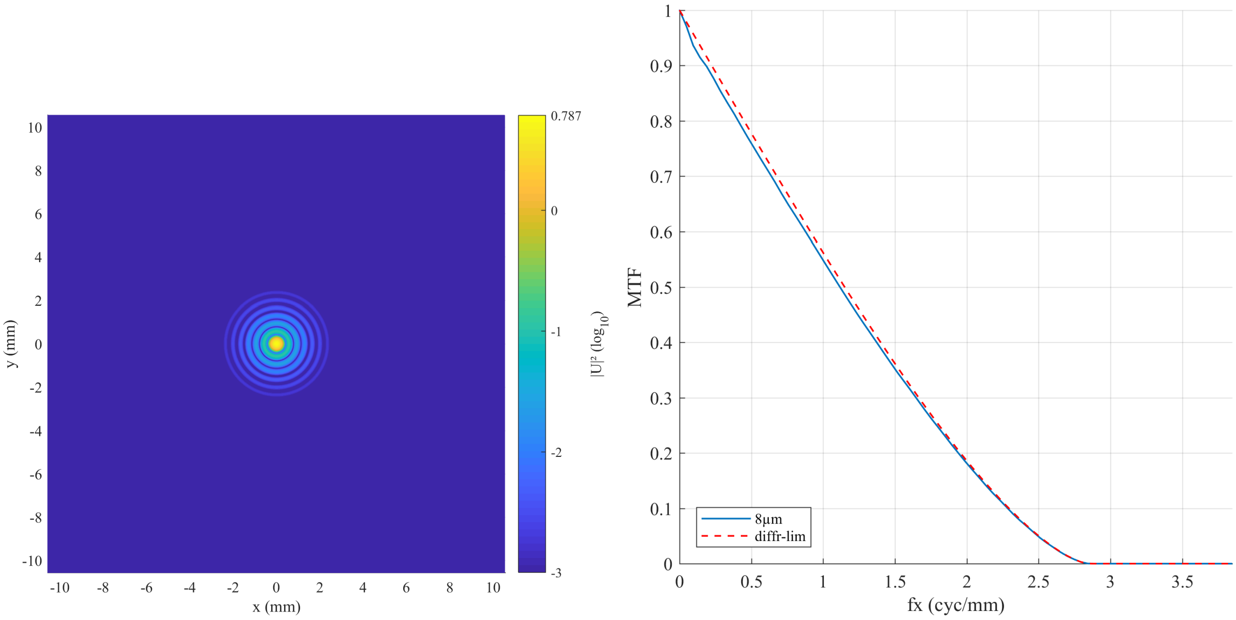
Using previously selected simulation parameters, the focal length of the selected ZnS-air-Ge MLDOE at $\lambda = 8$ µm is $f = 620$ mm: the input field is propagated to a target plane $z = 620$ mm. The irradiance and modulation transfer function (MTF) of the output field is shown in Figure 9.6.

The MLDOE approaches the diffraction limit, the spot displaying a Bessel pattern caused by the diffraction effect of the circular aperture. The focalization efficiency is measured by the Strehl ratio in Figure 9.7. It hides the diffraction effect of the pupil and leaves only the intrinsic MLDOE efficiency.

The MLDOE Strehl ratio remains above 0.96 for the whole waveband, with maximal values located at the design wavelengths. The MLDOE geometric construction could force constructive interferences at these wavelengths, resulting in higher focalisation efficiency (closer to the diffraction-limited case). In conclusion, the selected ZnS-air-Ge MLDOE has a focalisation efficiency of over 96% in the TIR bandwidth and is a diffraction-limited component, as long as the TEA remains valid. Additional optical aberrations may arise and degrade the Strehl ratio but cannot be accounted for considering the TEA and Fourier optics.

9.3.2 Results along the optical axis

Fourier optics is used in this section to analyse the light beam shape along the optical axis. In particular, the study concentrates on the energy contained in the unwanted diffractive orders. The same input field at $\lambda = 8$ µm is propagated from plane to plane along the optical axis with a propagation step d_z, between the input plane $z = 0$ mm and the image plane $z = 620$ mm (focal distance). The propagation step $d_z = 1$ mm is small enough to



(a) Irradiance $|U(x, y; z)|^2$ at $z = 620$ mm (b) MTF at $z = 620$ mm

Figure 9.6: Output field in the target plane $z = 620$ mm. a) The irradiance map is displayed with a logarithmic scale. b) Both simulated and diffraction-limited MTFs are displayed for comparison.

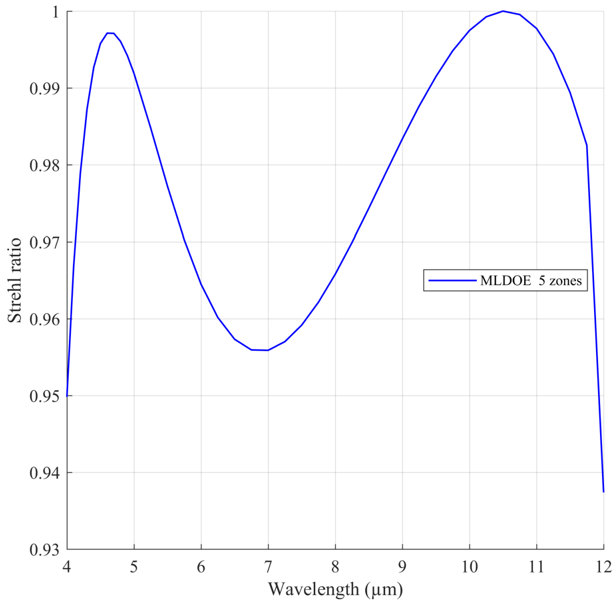


Figure 9.7: The ZnS-air-Ge MLDOE has a focalisation efficiency over 0.96 in the TIR bandwidth and nearly reaches 1 for its design wavelengths (4.7 µm and 10.4 µm).

obtain a "continuous" representation of the field along the optical axis. For each plane $id_z, i \in [1, 620]$, the 2D irradiance $|U|^2$ is recorded in a central slice $y = M/2 + 1$. Figure 9.8 shows the irradiance logarithm along the optical axis:

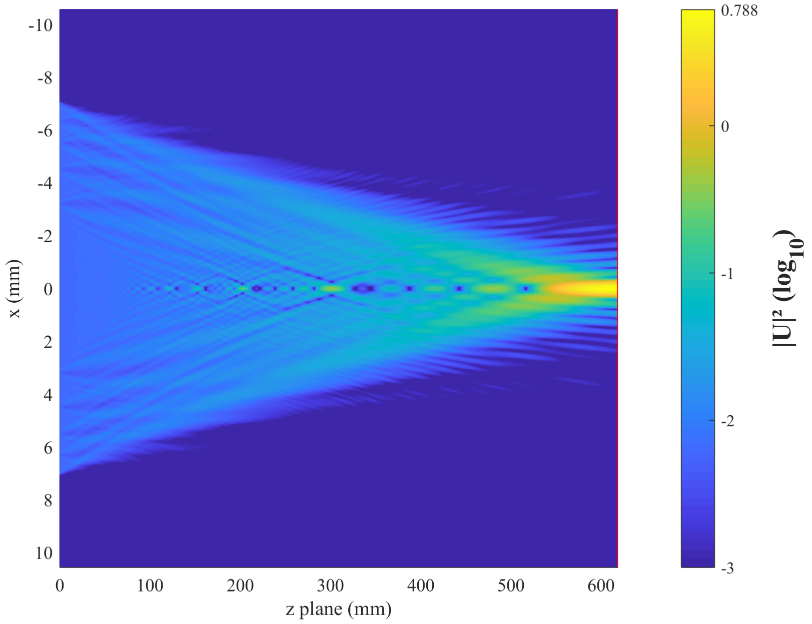


Figure 9.8: Simulation of the irradiance logarithm along the optical axis z. The input field resulting from the MLDOE phase shift is situated in the plane $z = 0$ mm, and the output field is located in the plane $z = 620$ mm. The propagation step is $d_z = 1$ mm. The x-axis represents the aperture radius of the MLDOE. Diffractive orders appear along the optical axis.

At each propagation plane id_z, the 2D Strehl ratio is recorded and gives information about the energy content in this plane. Figure 9.9 shows the Strehl ratio variation along the optical axis:

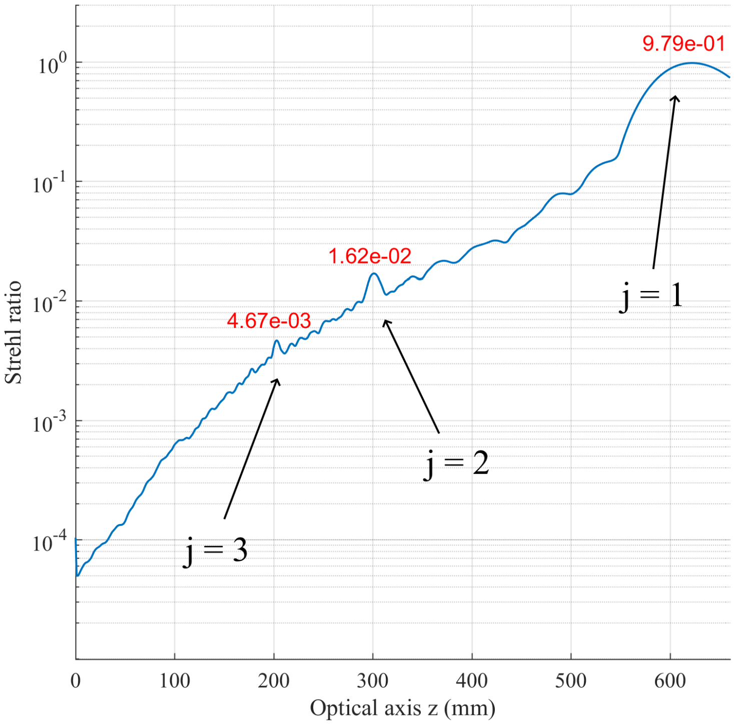


Figure 9.9: Strehl ratio variation along the optical axis. The logarithmic scale enables distinguishing other diffractive orders.

The peaks corresponding to the diffractive orders $j = +1, +2, +3$ are respectively located at $z = f$, $f/2$, $f/3$ in Figure 9.9. The unwanted orders $j = +2$ and $j = +3$ carry nearly no energy. Consequently, MLDOEs have a unique focus located in the first diffractive order. This property illustrates the imaging potential of MLDOEs for wide-band applications. However, further analysis should be conducted for lower f-number MLDOEs (i.e. higher number of zones), for which the TEA-based Fourier optics method is no longer accurate.

9.3.3 Chromatic aberration (LCA)

This section evaluates the Strehl ratio along the optical axis (similarly to Figure 9.9) for multiple wavelengths. For each wavelength, we aim to find the "best focus" plane to derive a focal law $f_{simu}(\lambda)$. The simulation domain is confined between the planes $z = f(\lambda) - 10$ mm and $z = f(\lambda) + 10$ mm, where $f(\lambda)$ is defined using Equation 9.28. We still consider 620 propagation steps. The "best focus" plane is determined with an accuracy $d_z/(620 - 1) = 32$ µm. Figure 9.10 shows the comparison between the simulated focal length and the analytical one (Equation 9.28):

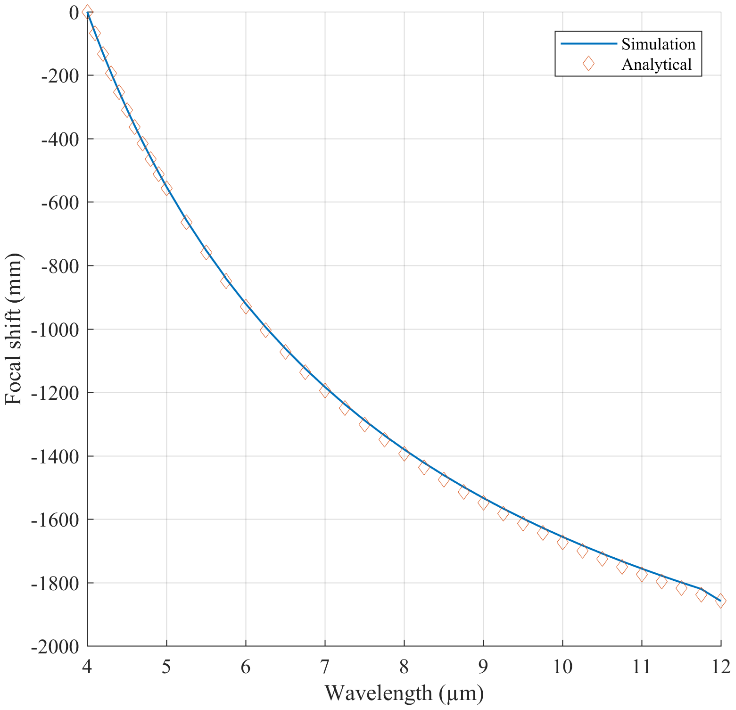


Figure 9.10: Chromatic focal shift comparison between the analytical focal plane and the determined "best focus" plane. Each point of the simulation curve represents the plane with highest Strehl ratio among 620 planes between $f(\lambda) - 10$ and $f(\lambda) + 10$ for a given wavelength λ. The dotted curve shows the expected analytical behaviour of the MLDOE (Equation 9.28).

Figure 9.10 depicts an excellent agreement between the analytical focal law and the determined best focus. This study shows that, like other DOEs, MLDOEs' LCA has a strong negative variation. Therefore, they combine very well with refractive components (positive LCA).

9.3.4 Conclusion: MLDOE performance under the TEA

This chapter has used a geometric model based on the thin element approximation (TEA) to deduce the phase function of multilayer diffractive optical elements (MLDOEs). This phase function influences a paraxial input field travelling through the MLDOE. The angular spectrum method allows us to propagate this field to any target plane using the angular spectrum method. The Strehl ratio, computed in the focal plane, showed a focalisation efficiency above 96% between 4.4 and 12 µm for the operating order +1. The other unwanted orders have negligible energy, so MLDOEs are considered mono-focal. Inside the validity domain of the TEA, an MLDOE is a wide-band diffraction-limited imaging device. The second study has shown that MLDOEs have a chromatic focal shift that matches the analytical expression. Achromat MLDOE hybrid doublets efficiently correct multiple aberrations and are easily designed using standard achromatisation equations. MLDOEs have a high potential for broadband optical design as an extension of the single-band DOE. This conclusion only holds inside the validity domain of the TEA.

9.4 Validity of the thin element approximation (TEA)

The TEA validity domain is derived based on the study of MLDOE aspect ratios. The author has published the results presented in this section in [67].

The presented TEA-based model is valid when meeting three restrictions [78]:

- The microstructure periods are much larger than the wavelength (more than 14 times [61])
- The field is paraxial
- The microstructures are thin: their periods must be at least 6 times larger than their heights ([77])

We study the first restriction through a worst-case MLDOE design. The mth zone period involves Equation 9.1 (similar for DOEs and MLDOEs) and only depends on the zone number (N) and the aperture diameter. We consider a F/10 MLDOE (at $\lambda = 8$ µm) with a high number of diffractive zones $N = 30$. The resulting aperture diameter D is:

$$N = \frac{D}{8\lambda F/\sharp} \Leftrightarrow D = 19.2 \text{ mm} \tag{9.31}$$

The resulting focal distance f is therefore $f = 192$ mm. The extreme and lowest zone period T_N is:

$$T_N = R_N - R_{N-1} = D/2 - \sqrt{2f\lambda(N-1)} = 161\text{µm} \tag{9.32}$$

According to the first restrictions, the maximal wavelength that ensures the validity of the TEA is $\lambda = 161/14 = 11.5$ µm. This study shows that a worst-case MLDOE with a very high diameter and number of zones still fulfils the first TEA restriction considering its lowest zone period. The second restriction is fulfilled as long as the incident radiation is parallel to the optical axis and the diameter remains "small".

The aspect ratio ($\mathcal{R}_m$) of the mth zone is introduced:

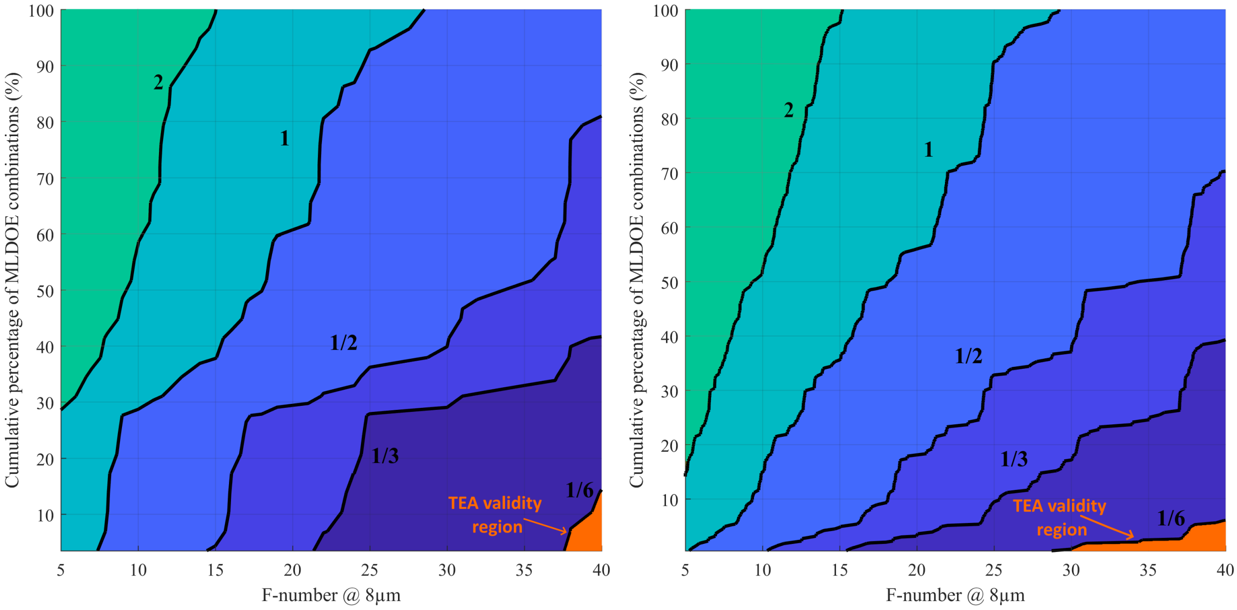
$$\mathcal{R}_m = max(H_1, H_2)/T_m \tag{9.33}$$

where T_m is the mth zone period (i.e. ring width). According to [77], the thin element approximation remains accurate when $\mathcal{R} < 1/6$. T_m only depends on optical parameters:

$F/\sharp$ and diameter D, whereas the zone heights are influenced by the material combination and the choice of design wavelengths (Section 9.2.1). Hence, the aspect ratio $\mathcal{R}$ must be analysed through a parametric study.

Keeping manufacturability concerns in mind, we fix the number of zones $N = 10$ and the aperture diameter $D = 10$ mm. We check the validity of the TEA (Figure 9.11) in the worst-case scenario by considering the lowest zone period and the thickest layer. This study is done for various $F/\sharp$ and considers all possible MLDOE material combinations. Each layer (as well as the gap) may consist of 12 potential infrared materials: Ge, ZnS, ZnSe, GaAs, AgCl, CdTe, and six chalcogenide materials, from IRG22 to IRG27. We therefore define an "MLDOE combination" as a triplet of material (m_1, m_g, m_2). The denominations "two-layer DOE" and "three-layer DOE" are employed when the gap material m_g is filled by air or another IR material.

Figure 9.11 displays the aspect ratios obtained for each MLDOE combination and $F/\sharp$. Because changing any material in a combination may lead to very different aspect ratios, the latter are sorted by thickness in Figure 9.11, where the cumulative percentage of MLDOE combinations is considered. This is the number of MLDOE combinations (in %) that display at most a specific aspect ratio for a given $F/\sharp$.



(a) Air gap only combinations, 29 in total

(b) Filled gap only combinations, 224 in total

Figure 9.11: MLDOE aspect ratios $\mathcal{R}$ (Eq. 9.33) in function of the cumulative percentage of MLDOE combinations, for various $F/\sharp$. The combinations have been sorted from the thinnest ($\mathcal{R} < 1/6$) to the thickest ($\mathcal{R} > 2$) for all $F/\sharp$ between 5 and 40. a) Two-layer DOE (air gap). b) Three-layer DOE (filled gap). The aperture diameter is fixed to $D = 10$ mm and the design wavelengths are $\lambda_1 = 4.7$ µm and $\lambda_2 = 10.4$ µm. The worst case implies the thickest layer and the lowest zone period. Only material combinations with a total height < 1 mm are kept and displayed to keep a relatively thin element. The orange region shows the validity of the TEA model [77].

An air gap is imposed in Figure 9.11a). Figure 9.11 only provides combinations with a total height lower than 1 mm, resulting in 253 combinations. In Figure 9.11a), the TEA validity requirement is fulfilled at most by 15% of the two-layer DOEs, with a minimal F-number of 37. Only 5% of the three-layer DOEs fulfil the third TEA requirement in

Figure 9.11 b).

As a result, the TEA model thickness requirement is never satisfied, except for extremely high $F/\sharp$. Thereby, the period size of MLDOEs and monolayer DOEs is the same (Eq. 9.1), but MLDOEs are 10 to 100 times thicker than DOEs. For instance, considering a monolayer DOE, we analyse the TEA model's validity in the worst possible case: the lowest refractive index material, AgCl ($n \simeq 2$), and an F-number of 5 at $\lambda = 8$ µm:

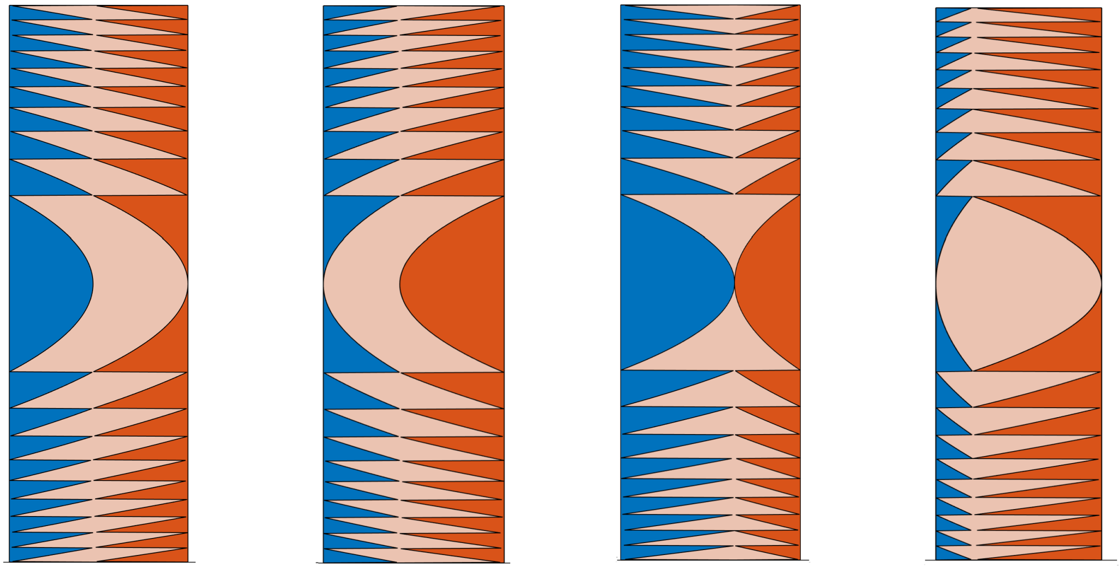
$$\mathcal{R}_{\mathcal{DOE}} = \frac{H}{T_{min}} = \frac{\lambda}{(n-1)T_{min}} \simeq 1/10 < 1/6 \tag{9.34}$$

This numerical application proves that the TEA model, while always accurate in the case of monolayer DOEs, is not suited for the study of MLDOEs.

In conclusion, we have demonstrated the need for an alternative MLDOE modelling method. MLDOEs have high potential in solving the critical issue of dual-band hybrid systems, bringing aberration correction and compactness. However, the current phase model is inacurate for thick MLDOEs and must be replaced to correclty assess there real performance. This is done in the following chapters using the innovative "ray model", combining ray-tracing (Chapter 5) and Fourier optics (Chapter 4). The ray model is dedicated to estimating the phase delay introduced by an MLDOE, and is deeply examined in Chapter 10.

9.5 Annex: design and selection of the right layer configuration

This Section does not add to the theoretical material given in Section 9.2 but exhaustively explains the calculation results when the MLDOE is not in the converging-diverging configuration (Figure 9.1). Depending on the chosen materials, the layer configuration that effectively acts like an imaging MLDOE (globally converging) is not straightforward. Figure 9.12 displays a representation of the four possible configurations to study:



(a) Converging-diverging MLDOE (b) diverging-converging MLDOE (c) Converging-converging MLDOE (d) diverging-diverging MLDOE

Figure 9.12: MLDOE possible configurations. They depend on the material choice for each layer and gap. For each material combination, only one configuration effectively corresponds to a converging MLDOE.

For each configuration, the calculation of phase and OPL follows the same steps and a similar geometry as in Figure 9.1. In fact, the optical path length for any radius r, given by Equation 9.2, is valid regardless of the configurations. Only the OPL associated with the extreme radii of each zone, namely $OPL(r_{m+1})$ and $OPL(r_m)$ for the mth zone, change depending on the configuration. Thereby, Equations 9.2, 9.3 and 9.7 become:

$$\begin{cases} OPL(r) = -n_1\left[H_1 - z_1(r)\right] + z_1(r)n_g + z_2(r)n_g + n_2\left[H_2 - z_2(r)\right] + f_d + \lambda_d \dfrac{r^2}{r_1^2} \\ OPL(r_m) = -\begin{bmatrix}\mathcal{D}_1 & (1-\mathcal{D}_1)\end{bmatrix}\begin{bmatrix}n_g \\ n_1\end{bmatrix} H_1 + \begin{bmatrix}\mathcal{D}_2 & (1-\mathcal{D}_2)\end{bmatrix}\begin{bmatrix}n_g \\ n_2\end{bmatrix} H_2 + f_d + m\lambda_d \\ OPL(r_{m+1}) = -\begin{bmatrix}\mathcal{D}_1 & (1-\mathcal{D}_1)\end{bmatrix}\begin{bmatrix}n_1 \\ n_g\end{bmatrix} H_1 + \begin{bmatrix}\mathcal{D}_2 & (1-\mathcal{D}_2)\end{bmatrix}\begin{bmatrix}n_2 \\ n_g\end{bmatrix} H_2 + f_d + (m+1)\lambda_d \end{cases} \tag{9.35}$$

where $\mathcal{D}_i$, $i = 1, 2$, is a Boolean parameter coding for the configuration: $\mathcal{D}_i = 1$ means that layer i is diverging whereas $\mathcal{D}_i = 0$ codes for a converging layer. Using the auxiliary parameters A and B defined in Equation 9.10: $A(\lambda) = n_g(\lambda) - n_1(\lambda)$ and $B(\lambda) = n_2(\lambda) - n_g(\lambda)$, the OPD expression (Equations 9.6 and 9.7), for any radius r becomes:

$$\begin{cases} OPL(r) - OPL(r_m) = A\left[\mathcal{D}_1 H_1 - z_1(r)\right] + B\left[\mathcal{D}_2 H_2 - z_2(r)\right] - \lambda_d\left(m - \dfrac{r^2}{r_1^2}\right) \\ OPL(r_{m+1}) - OPL(r_m) = A(2\mathcal{D}_1 - 1)H_1 + B(2\mathcal{D}_2 - 1)H_2 + \lambda_d \end{cases} \tag{9.36}$$

For instance, in the converging-diverging configuration depicted in Figure 9.1, System 9.36 reduces to Equations 9.6 and 9.7, using $\mathcal{D}_1 = 0$ and $\mathcal{D}_2 = 1$. The Fermat's principle (System 9.9) can be applied to System 9.36 for the two design wavelengths λ_1 and λ_2 and will lead to an analytical expression for z_1, z_2, H_1 and H_2, similarly to System 9.11. First, we solve the height equation $OPL(r_{m+1}) - OPL(r_m) = 0$ for $\lambda = \lambda_d$, $d = 1, 2$:

$$\begin{cases} H_1 = \dfrac{+\lambda_1 B_2 - \lambda_2 B_1}{A'_1 B_2 - A'_2 B_1} \\ H_2 = \dfrac{-\lambda_1 A_2 + \lambda_2 A_1}{A_1 B'_2 - A_2 B'_1} \\ A'_i = (2\mathcal{D}_1 - 1) A_i \ ; \ i = 1, 2 \\ B'_i = (2\mathcal{D}_2 - 1) B_i \ ; \ i = 1, 2 \end{cases} \tag{9.37}$$

System 9.11 is retrieved when $D_1 = 0 \to A' = -A$ and $D_2 = 1 \to B' = B$. Note that the height of layer 1 only depends on the configuration of layer 1 ($\mathcal{D}_1$) and vice versa. This means that z_1 and z_2 also only depends on $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Comparing Systems 9.37 and 9.11, it simply appears that the sign of H_1 is reversed when layer 1 is diverging instead of converging, and the same applies to H_2 with layer 2. Consequently, the expressions of H_1 and H_2 in any configuration can be very simply related to the heights expressed in System 9.11 for the converging-diverging situation, rewritten here as H_1^{CD} and H_2^{CD}:

$$\begin{cases} H_1 = (1 - 2\mathcal{D}_1) H_1^{CD} \\ H_2 = (2\mathcal{D}_2 - 1) H_2^{CD} \end{cases} \tag{9.38}$$

Noticing that $\mathcal{D}_{1,2}^2 = \mathcal{D}_{1,2}$, we have:

$$\begin{cases} \mathcal{D}_1 H_1 = -\mathcal{D}_1 H_1^{CD} \\ \mathcal{D}_2 H_1 = +\mathcal{D}_2 H_2^{CD} \end{cases} \tag{9.39}$$

Introducing Equations 9.39 into System 9.36 and solving for $\lambda = \lambda_d$, $d = 1, 2$ leads to the general expression of $z_1(r)$ and $z_2(r)$ for any configuration:

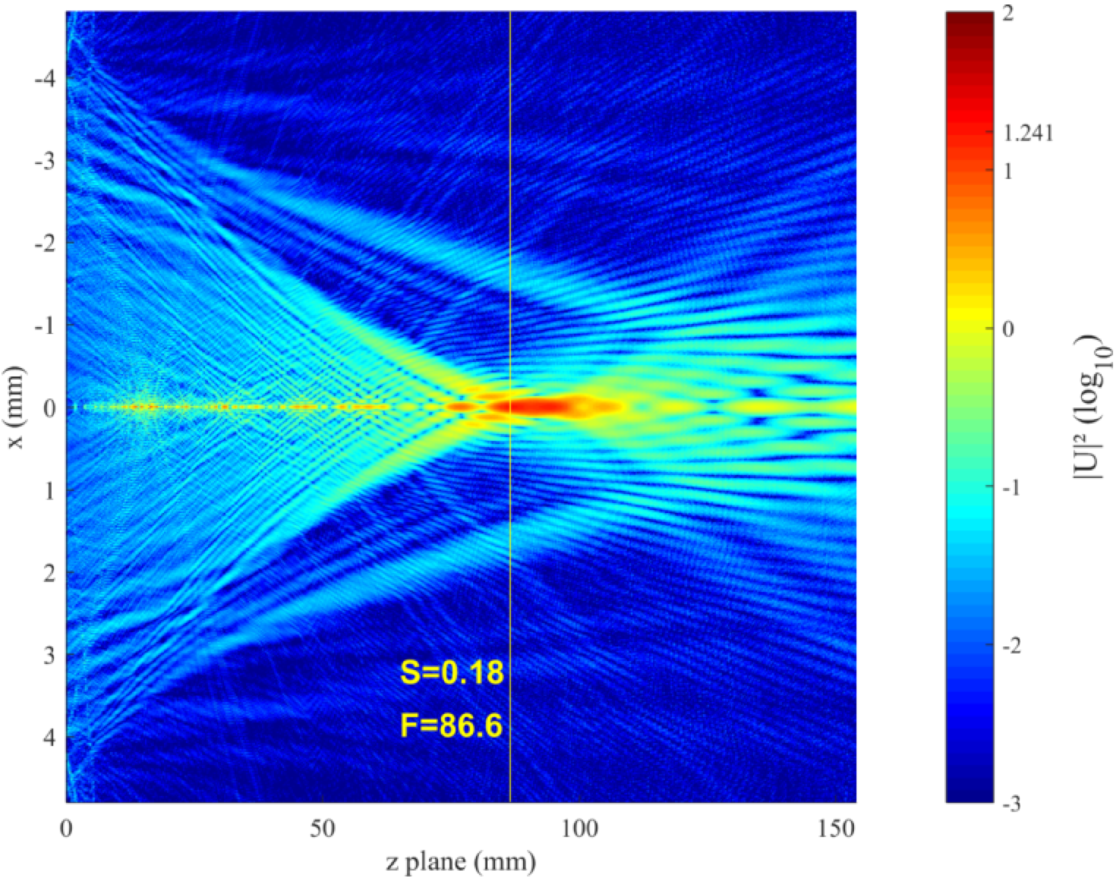
$$\begin{cases} z_1 = -H_1^{CD}(m + \mathcal{D}_1 - \dfrac{r^2}{r_1^2}) \\ z_2 = +H_2^{CD}(m + \mathcal{D}_2 - \dfrac{r^2}{r_1^2}) \end{cases} \tag{9.40}$$

Note that when $\mathcal{D}_1 = 0$ and $\mathcal{D}_2 = 1$, System 9.40 is equivalent to System 9.11. Therefore, changing the configuration from diverging to converging and vice versa is equivalent to computing the profile of the $m + 1^{th}$ parabolic zone, using $r \in [r_m; r_{m+1}[$. One issue arises with System 9.40: assuming a material combination defined by the refractive indices n_1, n_2 and n_g, it is simple to compute H_1^{CD} and H_2^{CD}. However, the value of $\mathcal{D}_1$ and $\mathcal{D}_2$ is still unknown and not *a priori* directly related to the material combination.

It has been explained in Section 9.2.1 that the converging-diverging MLDOE design equations (System 9.11) can only be used when $H_1 < 0$ and $H_2 > 0$. If the signs of H_1 or H_2 are reversed, the configuration must be changed and System 9.40 must be used. Consequently, the booleans $\mathcal{D}_1$ and $\mathcal{D}_2$ are defined by:

$$\begin{cases} \mathcal{D}_1 = H_1^{CD}(n_1, n_2, n_g) > 0 \\ \mathcal{D}_2 = H_2^{CD}(n_1, n_2, n_g) < 0 \end{cases} \tag{9.41}$$

The MLDOE design process for any configuration is depicted in Figure 9.2 in Section 9.2.1.



-4
-3
-2
-1
0
1
2
3
4
x (mm)
0
50
100
150
z plane (mm)
S=0.18
F=86.6
2
1.241
1
0
-1
-2
-3
$|U|^2$ ($\log_{10}$)

Chapter 10

Alternative phase model: Ray model

Chapter 9 has proven the inapplicability of the thin element approximation (TEA) for thick MLDOEs. An other modelling method must be used to assess the actual dual-band potential of MLDOEs for hybrid systems. We propose and describe the innovative "ray model", combining ray-tracing (Chapter 5) and Fourier optics (Chapter 4). Instead of the TEA, the ray model uses ray-tracing to compute the optical path differences inside an MLDOE, eading to a scalar phase estimation in the near-field. We generate optical figures of merit (point spread function, Strehl ratio, and chromatic focal shift) using the angular spectrum method. In parallel, we use the rigorous finite difference time domain (FDTD) method as a numerical reference, to analyse the ray model accuracy. This is done by quantifying the deviation between our ray model predictions and the exact FDTD calculations. It results from this study that the ray model is much more accurate than the TEA, but cannot precisely compute the Strehl ratio compared to FDTD. The analysed MLDOE designs' performance are much poorer than predicted by the TEA, raising the question of their optimisation.

Contents

10.1 Introduction

The scalar theory of diffraction involves the thin-element approximation (TEA), which becomes progressively inaccurate with increasing diffractive microstructure height [61, 77]. In the case of MLDOEs, the typical groove heights are a hundred times higher than for conventional monolayer DOEs. Accordingly, rigorous electromagnetic numerical approaches such as the coupled-wave analysis have been modelling the effect of microstructure depth and incident angle on the diffraction efficiency of saw-tooth MLDOEs [79, 80]. However, rigorous methods are computationally intensive, especially in the case of three-dimensional simulations. Recently, attractive techniques have involved geometric optics approximations. These approaches sit between the analytical TEA and exact electromagnetic methods. For instance, the effective area method has performed accurate diffraction efficiency computations in the case of MLDOEs with fixed period widths [81, 82]. Field Tracing [83] also involves geometric optics approximations to describe an optical component's phase delay in conjunction with Fourier optics or rigorous methods. It results in efficient and reliable field propagations even through complex optical systems.

This chapter is exclusively based on a paper published by the Author [67]. We use a scalar, ray-based method called the "ray model" instead of the TEA to compute an MLDOE phase delay. This approach previously appeared in [39] under the name "zone decomposition", applied to the design of hybrid optical systems. It is further implemented in [40] and coupled with diffraction calculation to estimate the modulation transfer function of relief-type DOEs.

We aim to demonstrate that the ray model gives more accurate and reliable results than the TEA. An MLDOE is designed with parabolic diffractive zones to provide constructive interferences at the focal plane [66] (Chapter 9). Compared to the TEA, which predicts a near-perfect imaging performance, the ray model heavily depends on refractive index variations and aspect ratios due to Snell's law of refraction. Accordingly, we show that the ray model is considerably less optimistic than the TEA. To prove that, FDTD phase calculations are conducted, providing a reference to estimate the accuracy of the ray model predictions. Each of the three presented approaches (TEA, ray and FDTD) generates a two-dimensional phase function in the MLDOE near field. A Fourier optics propagator based on the angular spectrum method is used to produce optical fields and evaluation metrics at the detector plane. We investigate various F-numbers in both MWIR and LWIR wavebands.

10.2 Fourier-based simulation process

This section describes how we combine FDTD simulations, the ray model and Fourier optics. This process uses the angular spectrum representation (similarly to Figure 4.1), recalled hereafter for clarity.

The angular spectrum of plane waves provides a straightforward method for evaluating, at any point in space, the properties (phase and amplitude) of a monochromatic electromagnetic wave propagating in a free-space medium (linear, homogeneous, non-magnetic, free of electric charges and currents). The complex field phasor $U(x, y; z)$ can be retrieved at any point in space using:

$$\begin{cases} U(x,y;z) = \mathcal{F}^{-1}\{\mathcal{F}\{U(x,y;0)\}H_z(f_x,f_y)\} \\ H_z(f_x,f_y) = \mathcal{F}\{h_z(x,y)\} = \exp\left[ikz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right] \end{cases} \tag{10.1}$$

where (f_x, f_y) are the spatial frequencies of the field, H_z is the transfer function of the wave propagation phenomenon, and h_z is the system's impulse response. Since free-space assumptions constrain the angular spectrum method, it cannot be employed to compute the optical field inside an MLDOE. Therefore, Fourier optics is bounded to propagate a complex field, originating after the MLDOE, onto a target plane. The band-limited angular spectrum method developed in [36] (Chapter 4) is used in the following (i.e. there is no aliasing due to high-frequency components).

Figure 10.1 shows the simulation process serving as a comparison tool to adjudicate between the TEA and the ray model:

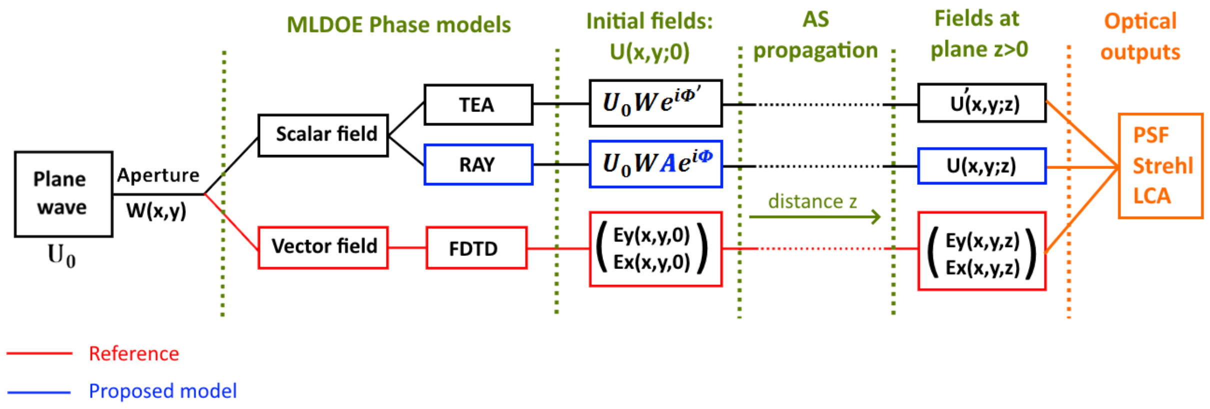


Figure 10.1: Diagram showing the numerical processes involved in this chapter to compare the TEA and the ray model, using FDTD as a reference.

As indicated in Figure 10.1, an MLDOE optical effect is mathematically represented by a scalar complex phasor $U(x,y;0) = A(x,y)\exp[i\Phi(x,y)]$ in both the TEA and ray models, considering an on-axis collimated monochromatic source. In contrast, FDTD provides an exact electromagnetic vectorial description of the MLDOE near-field $\mathbf{E} = (E_x, E_y)$. The resulting wavefront is propagated to the focal plane using the angular spectrum method, where we assess the MLDOE image quality (Strehl ratio and LCA). These evaluation metrics are computed for various F-numbers ($F/\sharp$) in MWIR and LWIR. Recall that the "best" focal plane, a priori unknown, has been defined in Chapter 9 as the plane $z > 0$ that provides the highest Strehl ratio. Consequently, it may differ from the expected analytic focal plane, whose expression is the same for standard DOEs and MLDOEs (Equation 9.28):

$$f_j(\lambda) = D\frac{F/\sharp}{j\lambda} \tag{10.2}$$

The diffractive order $j = 1$ is considered.

10.3 Alternative phase models: Ray model and FDTD

This section describes two alternatives to the TEA model. Our ray model, inspired by [39, 83], is a scalar model relying on ray-tracing to calculate optical path lengths (OPLs). The FDTD method simulates vectorial fields using an exact electromagnetic calculation and provides us with a reference.

10.3.1 Ray model

Chapter 9 has shown the inadequacy of the TEA to describe MLDOEs, except for very high $F/\sharp$ and specific material combinations. The examined ray model still belongs to the SDT but introduces a deviation of the rays at each interface according to Snell's law of refraction. We eliminate the TEA assumption of rays travelling parallel to the optical axis. Therefore, OPLs are longer in the ray model, and their lengths strongly depend on the diffractive zone shape, aspect ratio and on the transition from one material to another. The ray model accounts for undesirable optical effects such as total internal reflection, appearing for high aspect ratios or high index variations.

The MLDOE effect on a plane wave is obtained in the near-field through the calculation of a phase mask:

$$\Phi_{ray} = \frac{2\pi}{\lambda} OPL = \frac{2\pi}{\lambda} \sum_{i=1}^{3} n_{i;i+1} OPL_{i;i+1} \tag{10.3}$$

where i is an MLDOE optical interface, numbered between 1 and 3 (layer 1 grooves, layer 2 grooves and layer 2 back, see Figure 11.3). $n_{i;i+1}$ and $OPL_{i;i+1}$ are respectively the refractive index and the OPL between interface i and $i+1$. The latter is obtained using a ray-tracing engine: the professional ray-tracing software ASAP NextGen [84] or the analytical ray-tracing engine described in Chapter 5.

Fundamentally, the ray model and the TEA only differ by the OPL calculation method. The complex scalar field phasor associated with the ray model is expressed in the MLDOE near-field by [40]

$$U(x, y, \lambda) = A_{ray}(x, y) \exp\left[i\Phi_{ray}(x, y, \lambda)\right] \tag{10.4}$$

Recall that the TEA model can analytically optimise the zone heights to obtain a maximal performance at the focal plane for any selected material combination (see Section 10.4). In contrast, the more complex ray model includes refraction at the interfaces. Consequently, its performance predictions depend heavily on the material choice and groove aspect ratio.

10.3.2 FDTD phase simulation

Since the TEA and the ray model are scalar approximations, this chapter uses the vectorial FDTD method as a comparison reference. FDTD wave simulation is performed using the OptiFDTD software [43]. It propagates the electric field through the MLDOE, solving the discretized Maxwell's equations numerically (See Chapter 6). The E_z component is negligible along the propagation direction. The MLDOE circular symmetry reduces the sampling effort needed, leading to the adequate sampling study detailed in Section 6.3.

This numerical sampling investigation has resulted in the following worst-case sampling: $N_x = 20000$, $N_t = 5000$ and Δ_z, Δ_t are set to *auto* setting (CFL conditions in Equations 6.4 and 6.5).

The vectorial field obtained in the MLDOE near field is $E(x, y) = [E_x(x, y); E_y(x, y)]$. The E_x and E_y components are obtained respectively through separate transverse magnetic (TM) and transverse electric (TE) polarisation simulations. E_x and E_y are not coupled; they can be propagated independently using the angular spectrum method and incoherently summed to obtain the point spread function (PSF), modulation transfer function (MTF), and Strehl ratio at the target plane. Thereby, the simulation considers unpolarised light.

Finally, the two-dimensional FDTD layout of the selected ZnS-air-Ge MLDOE is shown in Figure 10.2.

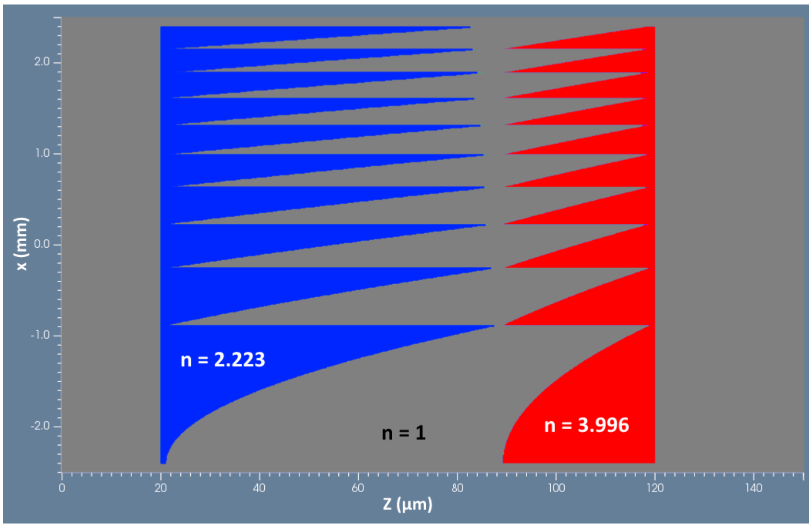


Figure 10.2: Two-dimensional FDTD layout of a ZnS-air-Ge F/15 MLDOE. The source and the detector respectively lies in the planes $z = 0$ µm and $z = 300$ µm. The blue and red layouts represent the two HDOE layers separated by an air gap.

The MLDOE layout of Figure 10.2 was obtained by superposing parabola pieces made of the specified materials, as explained in Chapter 6.

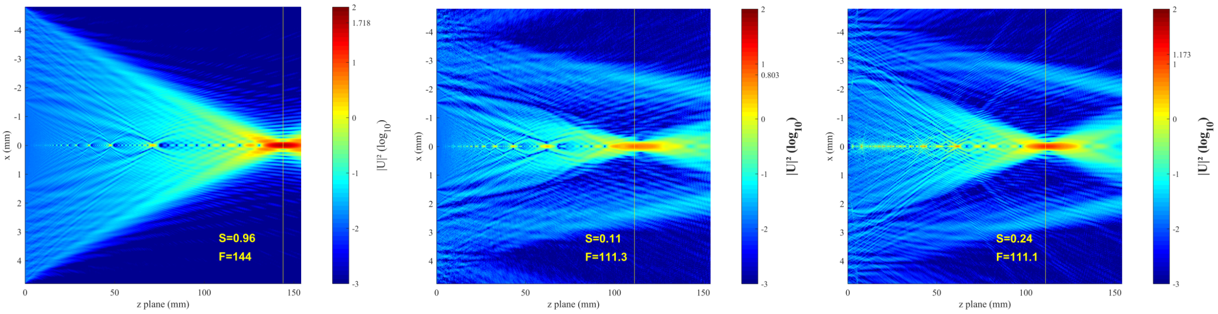
10.4 Results

Two arbitrarily selected MLDOE combinations will be considered for comparing the ray and the TEA models: a ZnS-air-Ge MLDOE (used for the sampling study) and a ZnS-air-IRG25 MLDOE. The former combination has the lowest thickness among "air-gap combinations" and has already been studied in [66]. IRG25 replaces the expensive and hard Ge layer in the second configuration. IRG25, a "soft" chalcogenide material, increases the manufacturability potential.

This section computes optical figures of merit (PSF, Strehl ratio, and LCA) according to the simulation process described in Figure 10.1. The aperture diameter is fixed at $D = 9.6$ mm, corresponding to exactly $N = 10$ diffractive zones (Equation 9.31).

10.4.1 Results along the optical axis: PSF comparison

We compare the irradiance along the optical axis, predicted by each of the three models (TEA, ray, and FDTD). From an input field at plane $z = 0$, the angular spectrum method allows us to accurately predict the resulting field at any parallel plane $z > 0$ in the context of free-space propagation. The first step for each MLDOE design and propagation wavelength is to retrieve the "best" focal plane. We propagate the near-fields (obtained via the ray model, the TEA model or FDTD) to multiple planes z along the optical axis, and for each plane, we compute the PSF for all three models. In the case of FDTD, vectorial field irradiances are incoherently summed to provide a total irradiance output in each propagation plane. Since we compare the prediction accuracy of the two scalar models, we use a single setup: $\lambda = 8$ µm and $F/\sharp = 15$, giving an analytic focal length $f = F/\sharp D = 144$ mm. The resulting irradiance patterns and "best" focal planes are described in Figures 10.3 and 10.4:

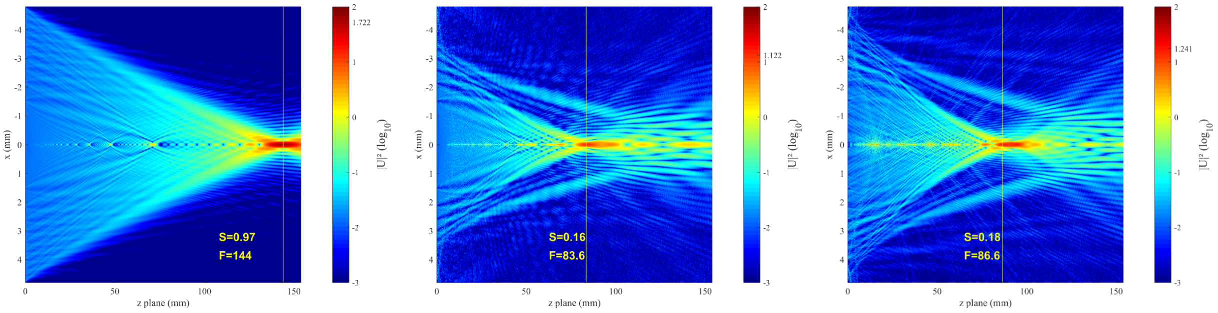


(a) TEA model. (b) Ray model. (c) FDTD.

Figure 10.3: PSF along the optical axis for a ZnS-air-Ge F/15 MLDOE considering: a) the TEA model, b) our ray model, and c) FDTD. The initial plane is in $z = 0$ mm, just after the MLDOE component. The maximum irradiance value, obtained at the focal plane, is displayed in the colour bar. The vertical yellow line shows the simulated "best" focal plane location F and its associated Strehl ratio S. The predictions of the TEA model are too optimistic, in contrast to the ray model and FDTD.

Figure 10.3 displays the ZnS-air-Ge MLDOE energy distribution along the optical axis for the three models. The TEA model predicts diffraction-limited optical performance at the focal plane (Strehl ratio of 0.98), and the latter coincides with the analytical focal length value. This result is not in accordance with the prediction of the ray and FDTD models. Both predict that only a central portion of the MLDOE contributes to the focus energy. A secondary ring of light, coming from the "intermediate zones", converges to a further focal point, creating important stray light. Finally, the ray and FDTD models predict a different "best" focal plane (around $z = 111$ mm) and different Strehl ratio values. The ray model is approximate, giving more pessimistic Strehl ratio values than FDTD (0.11 for the ray model and 0.24 for FDTD) for this design and F-number. Note that the number of zones contributing to the focus energy is higher for FDTD, explaining the lower peak irradiance value and, thus, the lower Strehl ratio for the ray model.

A one-dimensional cut of the logarithm of the PSF at the best focal plane (yellow line) is displayed in Figure 10.7a (Section 10.6). The TEA model provides a diffraction-limited Airy pattern, while the other models predict a very different PSF shape. The secondary light ring visible in Figure 10.3b and 10.3c clearly appears in Figure 10.7a, with an irradiance level of $10^{-1.5}$ [W m^{-2}]. The ray model accurately predicts the existence of this secondary ring but, as mentioned above, underestimates the peak irradiance value compared to FDTD.



(a) TEA model. (b) Ray model. (c) FDTD.

Figure 10.4: PSF along the optical axis for a ZnS-air-IRG25 F/15 MLDOE considering: a) the TEA model, b) our ray model, and c) FDTD. The initial plane is in $z = 0$ mm, just after the MLDOE component. The maximum irradiance value, obtained at the focal plane, is displayed in the colour bar. The vertical yellow line shows the simulated "best" focal plane location F and its associated Strehl ratio S. The predictions of the TEA model are too optimistic, again. In addition, these predictions are not impacted by the Ge layer change to an IRG25 layer. In contrast, this change has modified the predicted PSF, Strehl ratio, focal length and beam shape in the case of the ray model and FDTD.

Figure 10.4 displays the same optical metric as Figure 10.3, considering the selected ZnS-air-IRG25 MLDOE. The TEA model still generates a diffraction-limited performance and predicts a $f = 144$ mm focal length. In contrast, the ray model and FDTD generate a very different beam shape (Figures 10.4b and c), with a thinner outer light ring. FDTD and the ray model both predict the location of the best focal plane at $z = 86$ mm. Once again, the ray model is pessimistic with a Strehl ratio value of 0.11 against 0.18 for FDTD. Figure 10.7b in Section 10.6 details the PSF pattern difference at the "best" focal plane between the TEA and the other two models.

This section highlights the accuracy of the ray model over the TEA model for the presented MLDOE designs. Even if less precise than FDTD, the ray model can retrieve the shape of the irradiance pattern at the focal point and along the optical axis. It provides precise irradiance peak value, Strehl ratio, and focal length estimates. It is worth noting that changing the material of one or more layers strongly impacts the optical performance, beam shape, and focal length.

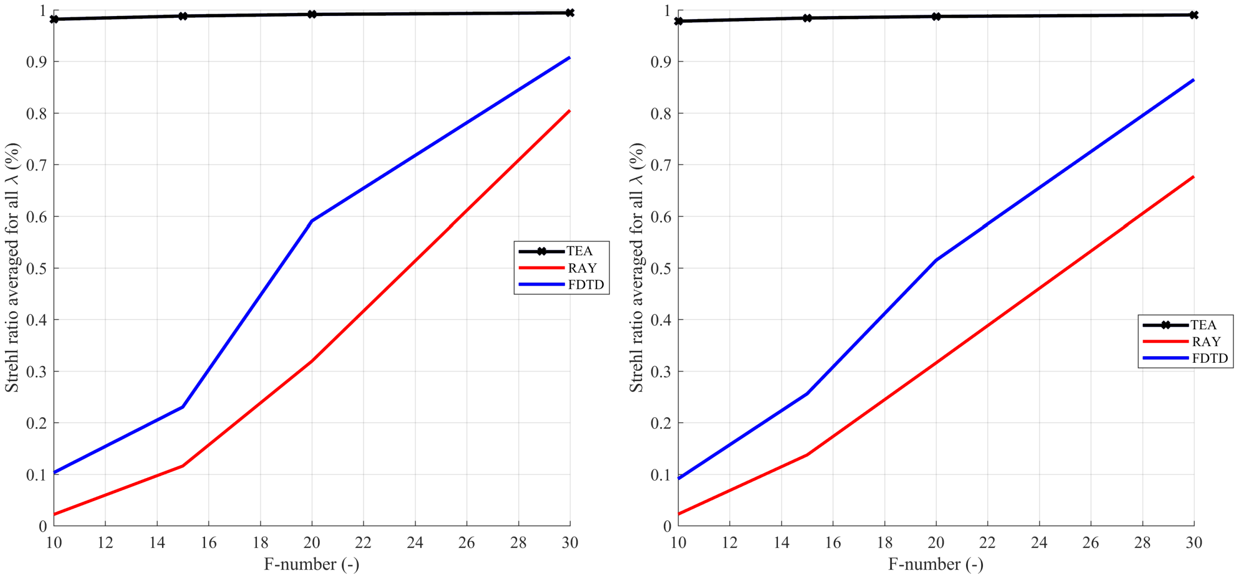
10.4.2 Results at the detector plane: Strehl ratio and LCA

This section compares the TEA and ray model using the Strehl ratio and the chromatic focal shift as optical performance metrics. Multiple wavelengths in the MWIR and LWIR bandwidths and multiple $F/\sharp$ are studied for both ZnS-air-Ge and ZnS-air-IRG25 MLDOE combinations.

We define the infrared Strehl ratio $Strehl_I$ as:

$$Strehl_I(F/\sharp) = \text{mean}_\lambda[Strehl(\lambda, F/\sharp)] \tag{10.5}$$

where the wavelength λ is in the range 4.4–12 µm. Figure 10.5 shows the evolution of the infrared Strehl ratio with the $F/\sharp$:



(a) ZnS-air-IRG25 MLDOE.

(b) ZnS-air-Ge MLDOE.

Figure 10.5: Infrared Strehl ratio at the best focus for multiple $F/\sharp$. The TEA model, ray model and FDTD are respectively displayed in black, red and blue. Each point depicts the wavelength averaged Strehl ratio in the 4.4–12 µm bandwidth. a) ZnS-air-IRG25 MLDOE. b) ZnS-air-Ge MLDOE.

Looking at the horizontal black curve, the TEA model predicts a near-perfect infrared Strehl ratio regardless of the materials and $F/\sharp$. Observing the ray model (red curve) and FDTD (blue curve) results, we can infer the behaviour of both MLDOE designs: Increasing the $F/\sharp$ lowers the aspect ratio; thus, more light passes through the MLDOE. Likewise, less stray light will be generated by the MLDOE, increasing the Strehl ratio. As seen in Figure 9.11, decreasing the aspect ratio makes the TEA progressively more valid, providing an asymptotic behaviour as displayed in Figure 10.5.

As previously noted in Figures 10.3 and 10.4, for both MLDOE designs, the ray model gives more pessimistic results than FDTD, which, as discussed in Section 10.5, is an acceptable result. Nevertheless, Figure 10.5 displays a constant divergence between the ray model and FDTD, quantitatively analysed in Section 10.4.3.

We recall that the chromatic focal shift (also known as LCA) is a significant figure of merit for diffractive optics. Thereby, diffractive lenses provide efficient achromatic hybrid solutions when combined with refractive lenses, so the accurate knowledge of their LCA is of prime importance. The LCA is therefore associated with the design performance, while the Strehl ratio gives a direct optical performance metric.

Figure 10.6 shows the evolution of the LCA with the wavelength for the ZnS-air-Ge MLDOE and various $F/\sharp$. The LCA is defined as $LCA(\lambda) = f(\lambda) - f(\lambda = 12 \text{ µm})$. A similar graph can be found in Section 10.6 for the ZnS-air-IRG25 MLDOE.

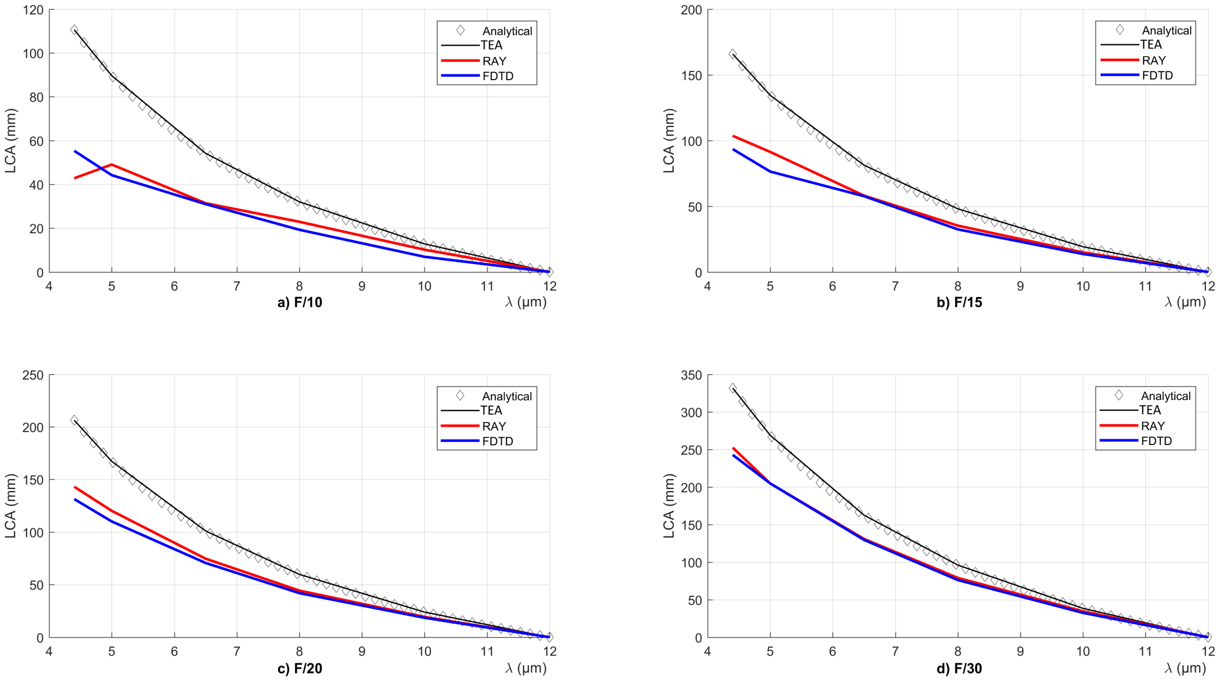
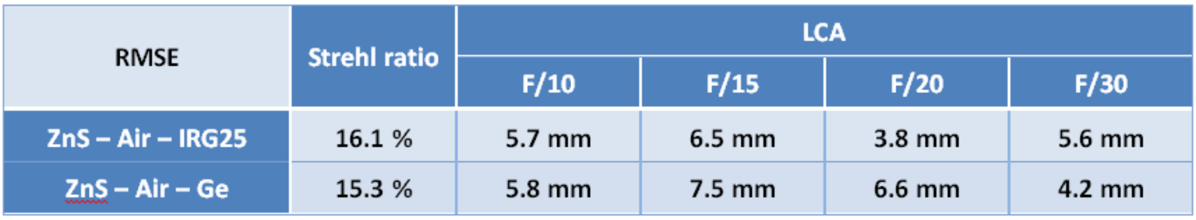


Figure 10.6: ZnS-air-Ge MLDOE LCA for various $F/\sharp$: a) F/10, b) F/15, c) F/20 and d) F/30. The "best" focal plane is defined as the plane $z > 0$ with the highest Strehl ratio. The TEA model, ray model and FDTD are respectively displayed in black, red and blue. The analytical LCA is shown with black diamonds (Eq. 9.28). FDTD is the reference curve.

Figure 10.6 confirms that the TEA model (black curve) is perfectly in agreement with the diffractive lens focal law (Eq. 9.28, black diamond curve) for all $F/\sharp$. Both FDTD and the ray model (blue and red curve) are in good agreement for all $F/\sharp$, wavelengths, and material combinations (see Figure 10.8 in Section 10.6). However, they disagree with the TEA model.

10.4.3 Accuracy of the ray model

This section computes the ray model accuracy (i.e. its divergence from FDTD) using the root mean squared error (RMSE), presented in Table 10.1.



RMSE	Strehl ratio	LCA			
		F/10	F/15	F/20	F/30
ZnS – Air – IRG25	16.1 %	5.7 mm	6.5 mm	3.8 mm	5.6 mm
ZnS – Air – Ge	15.3 %	5.8 mm	7.5 mm	6.6 mm	4.2 mm

Table 10.1: Table displaying the RMSE values between the ray model and FDTD (reference) for the two studied MLDOE designs. The numerical values are based on the 10 subfigures displayed in Figures 10.5, 10.6 and 10.8.

The RMSE metric describes the standard deviation of the residuals (the averaged distance between FDTD predictions and the ray model predictions). We interpret the LCA RMSE value (given in millimetres) by comparing it to the LCA maximum extent (for the FDTD model). The infrared Strehl ratio RMSE is $\simeq 15\%$ for both designs, meaning that the ray model cannot give accurate image quality values. Nonetheless, it provides a useful infrared Strehl ratio tendency, especially compared to the TEA model.

The ray model accurately predicts the LCA with an RMSE between 4 and 8 mm. The highest LCA extent (40 mm) occurs in Figure 10.8a. and is associated with a 6 mm RMSE in Table 10.1. It leads to a worst-case relative RMSE of 15%. This value decreases to less than 10% for higher F-numbers, regardless of the design.

10.5 Conclusion

We have shown that our ray model was more accurate than the TEA model for F-numbers between 10 and 30, at least for the two presented MLDOE designs. The ray model is not suited for accurately determining optical metrics and must be complemented by rigorous calculations, such as FDTD. In the meantime, the ray model accurately models the evolution of the Strehl ratio and LCA, which is not the case for the TEA model. The latter matches the asymptotic behaviour of the ray model for very high F-numbers (Figure 10.5). The latter allows us to extend the scalar theory to low F-numbers and high aspect ratios. Because of its scalar nature, the ray model is fast to compute and can even be used for off-axis MLDOE simulations. It is valuable since FDTD simulations take considerable processing time and require complex MLDOE layouts. When multiple wavelengths, materials, and F-numbers are studied (e.g., chromatic aberration estimation), the FDTD time requirements become challenging to handle. Therefore, if FDTD is necessary for verification purposes, the ray model should be used as an optimisation and estimation tool.

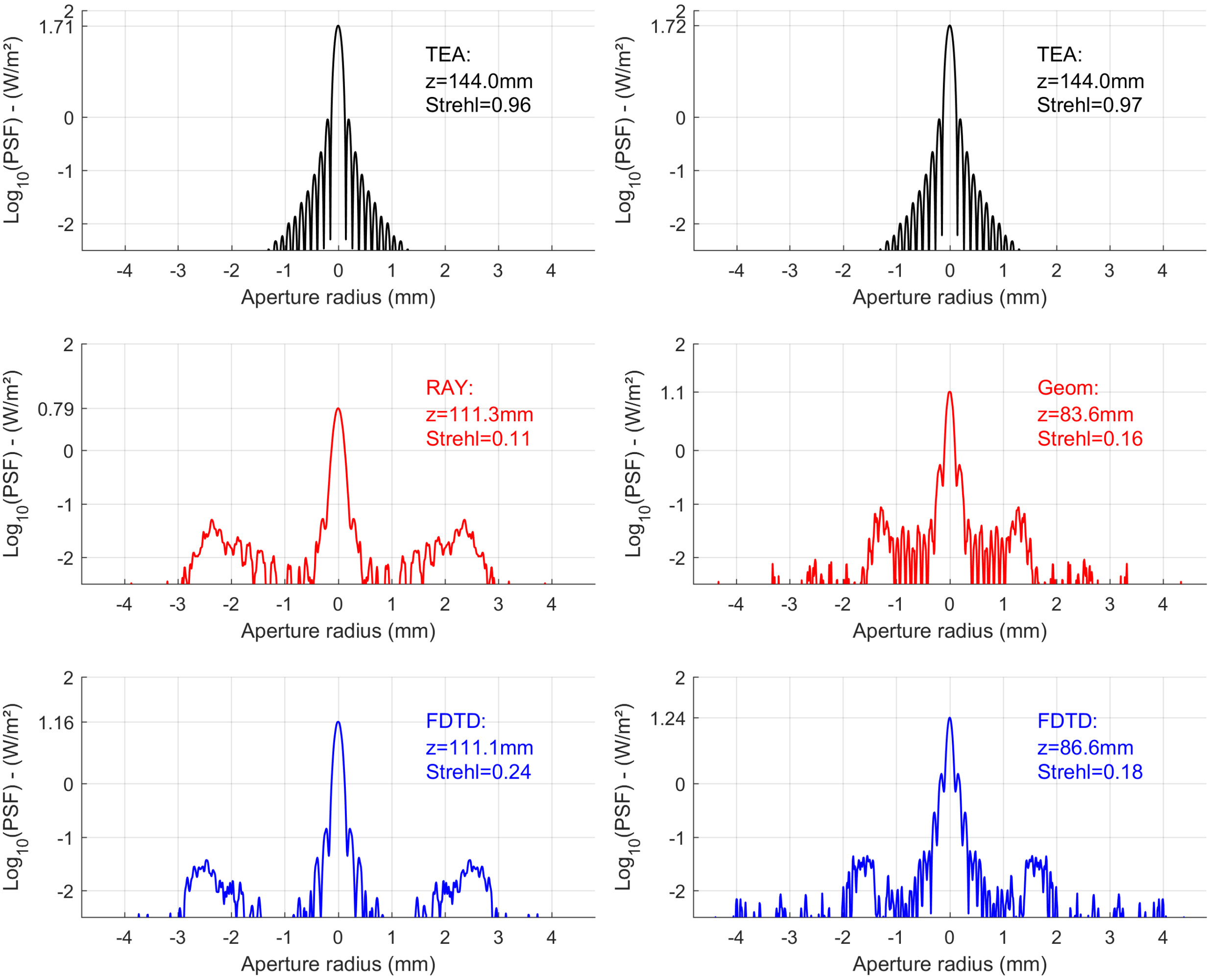
In addition, in the two studied cases, the ray model provides pessimistic results compared with FDTD, which is a significant advantage in predicting MLDOE performance with an approximate model. Thereby, the ray model offers a form of image quality insurance, being a worst-case scenario, at least in the two presented cases. We interpret it as a higher number of "operating" zones for FDTD than for the ray model, leading to increased peak intensity at the focal point. Physically, interferences inside the MLDOE lead to a converging wavefront even inside zones of high aspect ratio, for which no geometrical rays can pass (total internal reflection). Therefore, the ray model provides pessimistic estimations because it neglects the wave nature of light propagating inside the MLDOE.

In contrast to the TEA model, the ray model can compute an average transmission and considers total internal reflection. These quantities can be used as additional metrics to evaluate the performance of MLDOE design without requiring Fourier optics wave propagation (see Chapter 11). The ray model examined here is a powerful designing tool when studying numerous materials, F-numbers, and wavelengths. Further optimisation studies, carried out in Chapters 11 and 12, will try to increase the overall MLDOE performance.

10.6 Annex

A one-dimensional cut of the irradiance at the best focal plane for both MLDOE designs (corresponding to the yellow lines in Figures 10.3 and 10.4) is displayed in Figure 10.7. The irradiance pattern predicted by the ray model and FDTD differ from the TEA approach, for the two MLDOE designs.

The LCA for the ZnS-air-IRG25 MLDOE is presented in Figure 10.8, for various F-numbers: Similarly to Figure 10.6, Figure 10.8 demonstrates that the LCA does not



(a) ZnS-air-Ge MLDOE.

(b) ZnS-air-IRG25 MLDOE.

Figure 10.7: One-dimensional cut of the logarithm of the PSF, at the best focus, for the two F/15 MLDOE designs: a) ZnS-air-Ge, b) ZnS-air-IRG25. The TEA model, ray model and FDTD are respectively displayed in black, red and blue. Figures a) and b) show that the TEA model provides a diffraction-limited PSF. In contrast, the ray model and FDTD predict a very different PSF shape, with a secondary ring of light and a weaker central lobe. Contrary to the TEA model prediction, the PSF shape strongly depends on the MLDOE design.

follow the analytical curve as expected by the TEA model.

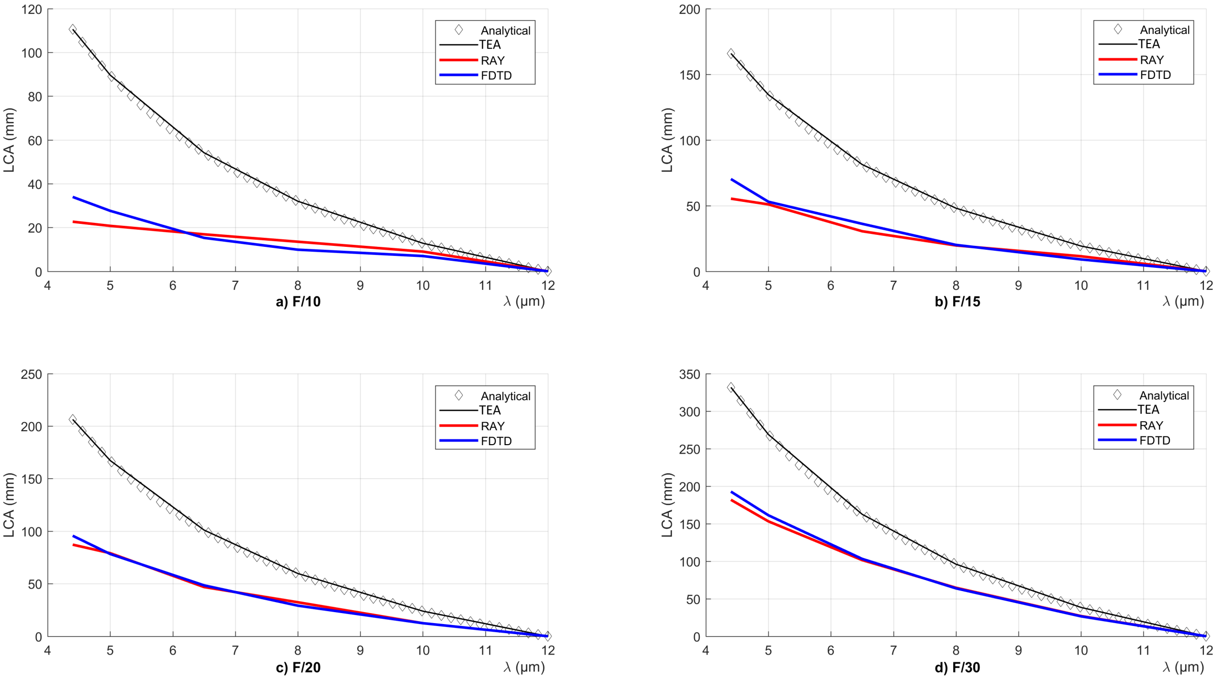
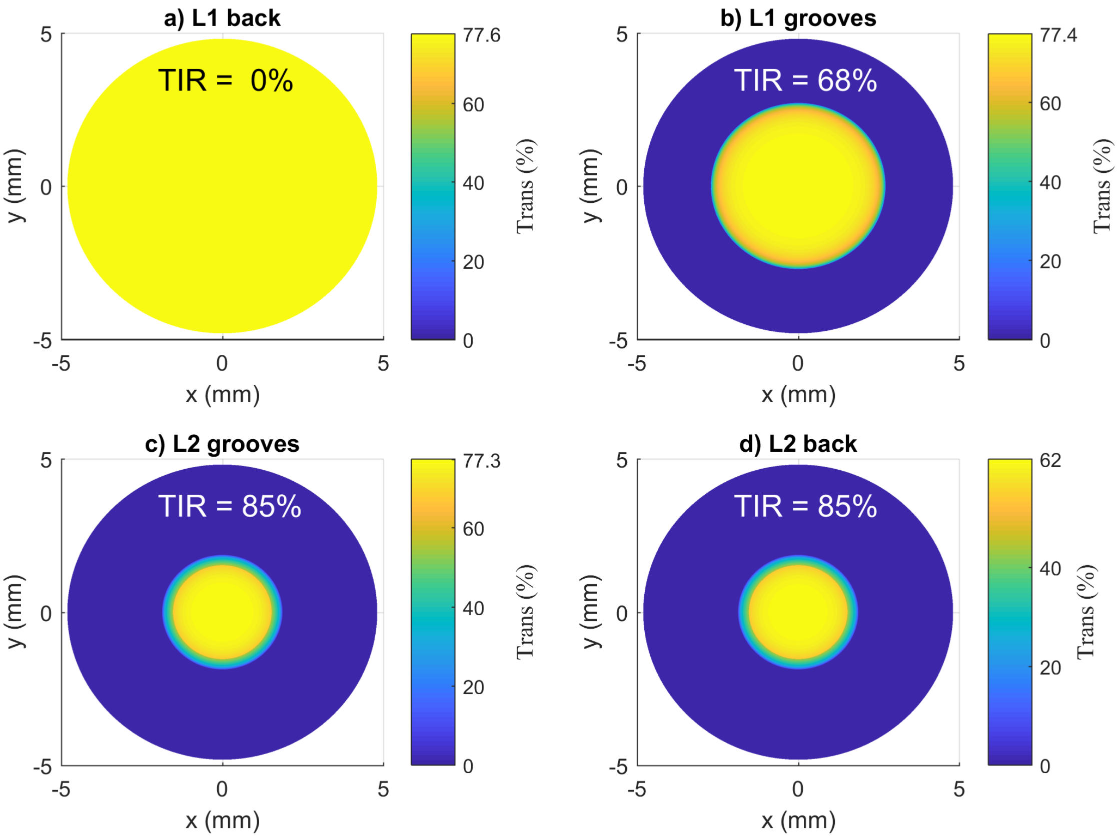


Figure 10.8: ZnS-air-IRG25 MLDOE LCA for various $F/\sharp$: a) F/10, b) F/15, c) F/20 and d) F/30. The "best" focal plane is defined as the plane $z > 0$ with the highest Strehl ratio. The TEA model, ray model and FDTD are respectively displayed in black, red and blue. The analytical LCA is shown with black diamonds (Eq. 9.28). FDTD is the reference curve.


a) L1 back
TIR = 0%
b) L1 grooves
TIR = 68%
c) L2 grooves
TIR = 85%
d) L2 back
TIR = 85%
y (mm)
x (mm)
5
0
-5
-5
0
5
Trans (%)
77.6
77.4
77.3
62
60
40
20
0

Chapter 11

MLDOE material selection methods

This chapter proposes a method to select appropriate MLDOE material combinations. The polychromatic integral diffraction efficiency (PIDE) metric is generally used to select the most suitable materials for multilayer diffractive optical elements (MLDOEs). However, this method is based on the thin element approximation (TEA), which yields inaccurate results for thick diffractive elements such as MLDOEs. We propose a new material selection approach based on three metrics: the transmission, the total internal reflection and the optical component's total thickness. This approach, called "geometric optics material selection methods" (GO-MSM), is tested in mid-wave and long-wave infrared bands. Finite-difference time-domain (FDTD) is used to study the optical performance (Strehl ratio) of the "optimal" MLDOE combinations obtained with the PIDE metric and the GO-MSM. Only the proposed method can provide MLDOE designs that are performing, but not to a sufficient extent. This finding leads to Chapter 12 MLDOE shape optimisation study. We also demonstrated that an MLDOE gap filled with a low-index material (air) strongly degrades the image quality.

Contents

11.1 Introduction

This chapter is dedicated to the study of MLDOE material combinations. A widely used material selection method [85] is based on the following observation: an "optimal" MLDOE design should undergo the smallest diffraction efficiency drop for the highest incidence angle. As described in Section 11.2, this method uses the PIDE, derived for off-axis incidence under the TEA. Although this method is fast and gives interesting results, its applicability is limited by the validity domain of the TEA. The latter becomes progressively inaccurate when the heights of the diffractive microstructures increase, as shown in Chapter 9.

This chapter proposes a new material selection method based on three evaluation metrics: the transmission, the total internal reflection (TIR) and the total MLDOE thickness. According to this alternative approach, an "optimal" MLDOE design has high transmission, low TIR and is as thin as possible. In the following, we refer to this material selection procedure as the "geometric optics material selection method" (GO-MSM) since it is based on ray-tracing and Fresnel equations to assess the TIR and the MLDOE transmission. The GO-MSM provides MLDOE designs but does not directly evaluate their image quality. To validate this approach and compare its result with the existing PIDE material selection method (PIDE-MSM), we use rigorous Finite-Difference Time-Domain (FDTD) electromagnetic calculations. Similarly to previous chapters, the optical performance of an MLDOE design is determined by the Strehl ratio, computed at the "best" focal plane. The Strehl ratio is retrieved based on FDTD near-field calculation and Fourier optics, as detailed in [67] (see Chapters 9 and 10). The results presented in this chapter have been published in [86]

11.2 PIDE material selection method (PIDE-MSM)

11.2.1 MLDOE off-axis modelling and PIDE definition

Each MLDOE zone can be approximated by the N-step structure of Figure 11.1:

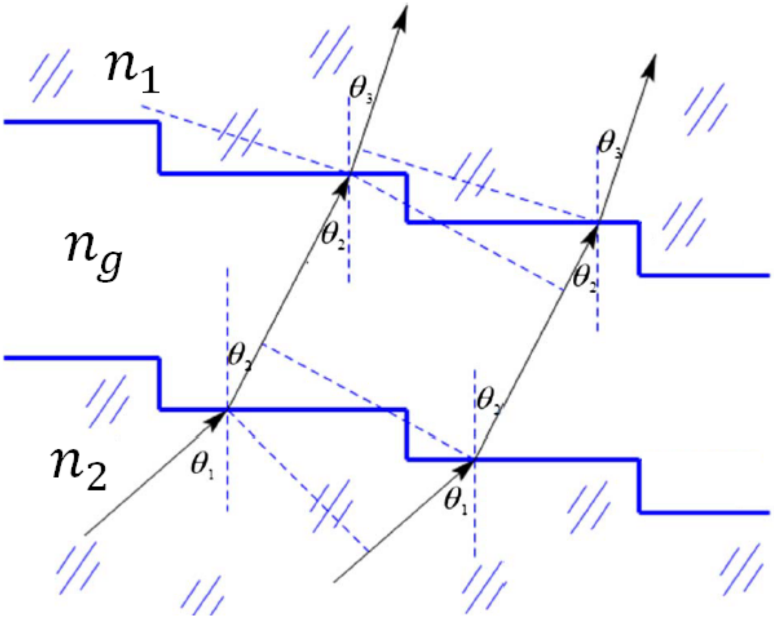


Figure 11.1: Modeling of ray propagation inside an MLDOE for off-axis incidence [85]. The continuous surface of each zone is modelled with an N-step profile instead.

Figure 11.1 is used to compute the optical path differences (OPDs) for layers 1 and 2, with an off-axis incidence. The calculation, detailed in [87], leads to the analytical expression of the total off-axis MLDOE phase delay:

$$\Phi(\lambda,\theta) = H_1[n_1(\lambda)\cos\theta - \sqrt{n_g^2(\lambda) - n_1^2(\lambda)\sin^2\theta}] + H_2[\sqrt{n_2^2(\lambda) - n_1^2(\lambda)\sin^2\theta}] \quad (11.1)$$

Where n_1, n_2 and n_g are respectively the refractive index of layer 1, 2 and the gap. H_1 and H_2 denotes the respective microstructure heights of layers 1 and 2, expressed in Equation 9.11.

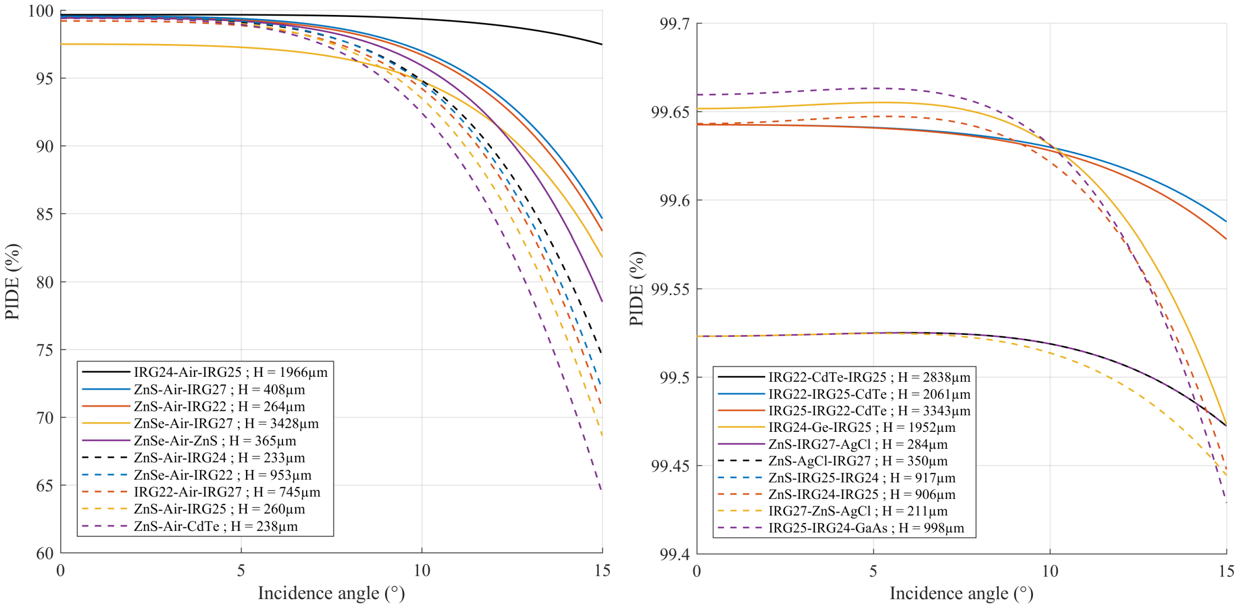
The PIDE (polychromatic integral diffraction efficiency) is a well-known evaluation metric to study diffractive optics [72]. In this chapter, the PIDE is based on the angle-dependant phase expression of Equation 11.1. For clarity, we recall the PIDE expression, previously defined in Equation 9.23:

$$\begin{cases} \overline{\eta}(\theta,j) = \dfrac{1}{\lambda_{max} - \lambda_{min}} \displaystyle\int_{\lambda_{min}}^{\lambda_{max}} sinc^2\left(j - \frac{\Phi(\lambda,\theta)}{2\pi}\right) \mathrm{d}\lambda \\ \eta(\theta,j) = \dfrac{1}{2}\overline{\eta}_{MWIR}(\theta,j) + \dfrac{1}{2}\overline{\eta}_{LWIR}(\theta,j) \end{cases} \quad (11.2)$$

where λ_{max} and λ_{min} represent the boundary wavelengths of the considered waveband. The total infrared PIDE $\eta(\theta,j)$ is the weighted sum of two PIDEs, one calculated in MWIR and the other in LWIR. We consider only the first diffractive order $j = 1$, as it is the main operating order for MLDOEs. The PIDE depends on the incident angle, the chosen wavebands and most importantly, the material refractive indices. We define an *MLDOE material combination* as a triplet of materials (m_1, m_g, m_2), and the denominations "two-layer DOE" and "three-layer DOE" are employed when the gap material m_g is filled by air or another IR material.

11.2.2 Results of the PIDE-MSM

This section focuses on applying the PIDE material selection method (PIDE-MSM) [70] and analysing its potential limits. The PIDE variation with the incident angle is displayed in Figure 11.2 for multiple MLDOE combinations:



(a) Two-layer DOEs (air gap). (b) Three-layer DOEs.

Figure 11.2: Angle-dependant PIDE for various MLDOE configurations: a) two-layer DOEs and b) three-layer DOEs. An air gap is imposed in a) while the gap material is variable in b). Only the ten configurations with maximal PIDE at 15° incidence are plotted. The PIDE variations are similar to [70]. H represents the total MLDOE thickness, computed using Equation 9.11.

Figure 11.2a displays the "best" two-layer DOEs combinations, according to [70]: Only the ten MLDOEs with the highest PIDE at maximal incidence (15°) are shown. The air gap is imposed for manufacturing reasons since linking two HDOEs with an IR material can be difficult. Figure 11.2a shows an important diffraction efficiency drops when the incidence increases. All ten combinations start with $> 95\%$ PIDE for on-axis incidence and end with PIDE values between $65\% - 85\%$ at 15° incidence, except for the IRG24-air-IRG25 combination (97% PIDE).

Similarly to Figure 11.2a, Figure 11.2b displays the ten "best" three-layer DOEs combinations. Any IR gap material is considered, regardless of the manufacturability of the resulting MLDOEs. All the solutions depicted in Figure 11.2b have negligible efficiency decrease (the worst PIDE being 99.4% at 15°). Since no thickness metric is accounted for in the PIDE method [70], we define the IRG22-IRG25-CdTe as an optimal three-layer DOE for the PIDE-MSM.

This analysis shows that the on-axis PIDE is nearly independent of the material combination since it is above 95% for all combinations depicted in Figures 11.2a and 11.2b. This behaviour results from Equation 11.1 and the choice of design wavelengths: For any material combination, an MLDOE will always provide 100% diffraction efficiency at the design wavelengths $\lambda_{1,2}$, assuming the TEA validity.

11.2.3 Limits of the PIDE method

We have shown in Chapter 9 [67] that the TEA was only valid for very few MLDOE combinations, with F-numbers higher than 40. This section provides a numerical application to test the PIDE method validity in the case of thick IR MLDOEs.

First, the off-axis MLDOE geometrical model depicted in Figure 11.1 is incompatible with the "paraxial" TEA requirement [78]. Even when on-axis PIDE is considered, according to [77], the aspect ratio of each layer and zone should remain smaller than 1/6. Recall that the mth zone aspect ratio $\mathcal{R}$, assuming layer i, is defined as:

$$\mathcal{R}(i,m) = \frac{H_i}{T_m} \tag{11.3}$$

We take Figure 11.2a most optimal configuration as an example: The IRG24-air-IR25 MLDOE. For standard values $N = 10$ diffractive zones and $F/\sharp = 15$, the 1^{st} zone period is T_1 ~3 mm (Equation 9.32). Following Equation 9.11, each layer is ~1 mm thick. The resulting aspect ratio is $\mathcal{R} = 0.33 > 1/6$, meaning that the TEA is not applicable for this F-number.

Imposing a 1/6 aspect ratio for the 1^{st} zone only leads to an F/30 design of $D = 19$ mm aperture diameter, keeping ten diffractive zones. This example shows that only a very constraining MLDOE design can ensure the validity of the TEA, and only for the largest zone. If extreme zones (i.e., lowest period) are considered, the TEA cannot fulfill the "thin" restriction mentioned in [78].

Furthermore, applying geometric optics laws to the IRG24-air-IRG25 MLDOE shows that it has a huge transmission drop due to total internal reflection (TIR). TIR occurs at the IRG24-air interface (refractive index ratio of 0.38) when an on-axis incident ray makes a 22.5° angle with the grooves interface normal. It happens inside the 1^{st} zone, starting from the radius $r = 0.48$ mm. As a result, only 10% of the MLDOE aperture can transmit light, leading to a nearly opaque MLDOE design. Therefore, the 100% on-axis diffraction efficiency predicted in Figure 11.2a is very likely overestimated, considering an F/15 MLDOE design. This numerical analysis is rigorously verified in Section 11.5 when studying the optical performance of the IRG24-air-IRG25 and IRG22-IRG25-CdTe solutions using FDTD.

11.3 Geometric optics material selection method (GO-MSM)

This section presents an alternative MLDOE material selection method based on geometric optics called the GO-MSM. The GO-MSM relies on three evaluation metrics: the total internal reflection (TIR), the transmission at each interface and the total MLDOE thickness. These metrics can be calculated rapidly for any MLDOE design and incident angle using a ray-tracing engine (ASAP [84] or the analytical ray-tracing calculator developed in Chapter 5). According to the GO-MSM, the most "optimal" MLDOE combination has the highest transmission and the lowest TIR and thickness. We present only the on-axis selection process, but we provide the off-axis results in Section 11.7 since they do not bring any change in the results.

Figures 11.3 and 11.4 graphically describe how we evaluate these three metrics. Figure 11.3 presents the shape and defines the optical interfaces of an arbitrary selected MLDOE (IRG23-IRG22-IRG25):

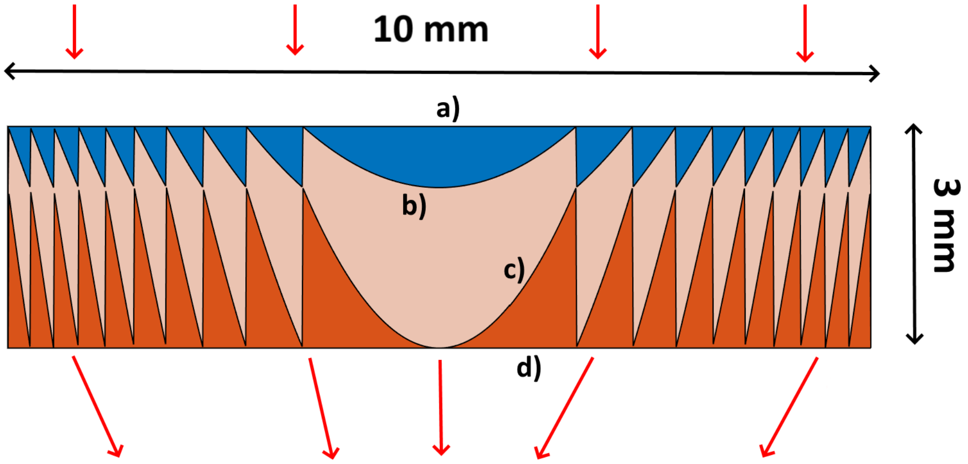


Figure 11.3: Dimensions and shape of an arbitrary selected MLDOE: IRG23-IRG22-IRG25. The top blue layer is IRG23, the gap (salmon colour) is IRG22, and the orange bottom layer is IRG25. The red arrows display the incident light's direction. The optical interfaces are denoted: a) Layer 1 back, b) Layer 1 grooves, c) Layer 2 grooves, and d) Layer 2 back. The MLDOE was designed according to Chapter 9.

Figure 11.4 displays the transmission map, obtained with geometric optics and Fresnel reflection equations, for each interface defined in Figure 11.3:

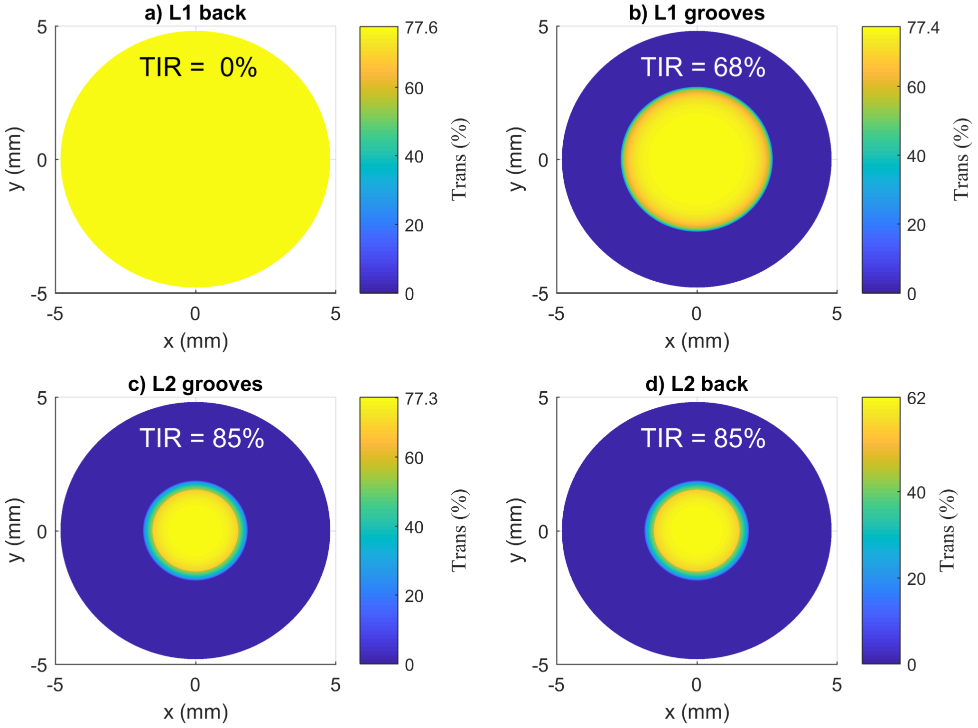


Figure 11.4: Transmission map for each interface of the IRG23-IRG22-IRG25 MLDOE depicted in Figure 11.3. The blue portion represents TIR (null transmission), whereas the white areas are outside the circular aperture.

Figure 11.4 shows the evolution of transmission and TIR inside the arbitrary selected MLDOE. The TIR percentage metric consists of the area covered by TIR over the circular aperture area. In this particular example, the working portion of the MLDOE is only composed of the central zones, thus potentially reducing its performance. The average measure of TIR is defined as the maximal amount of TIR after all interfaces, in this case, 85%, while the average transmission equals 62%.

The presented evaluation metrics (TIR, transmission and thickness) are used in multiple selections processes, considering two-layer DOEs in Section 11.4.1 and three-layer DOEs in Section 11.4.2. The optimisation variables are the layer/gap materials.

11.4 Results of the GO-MSM

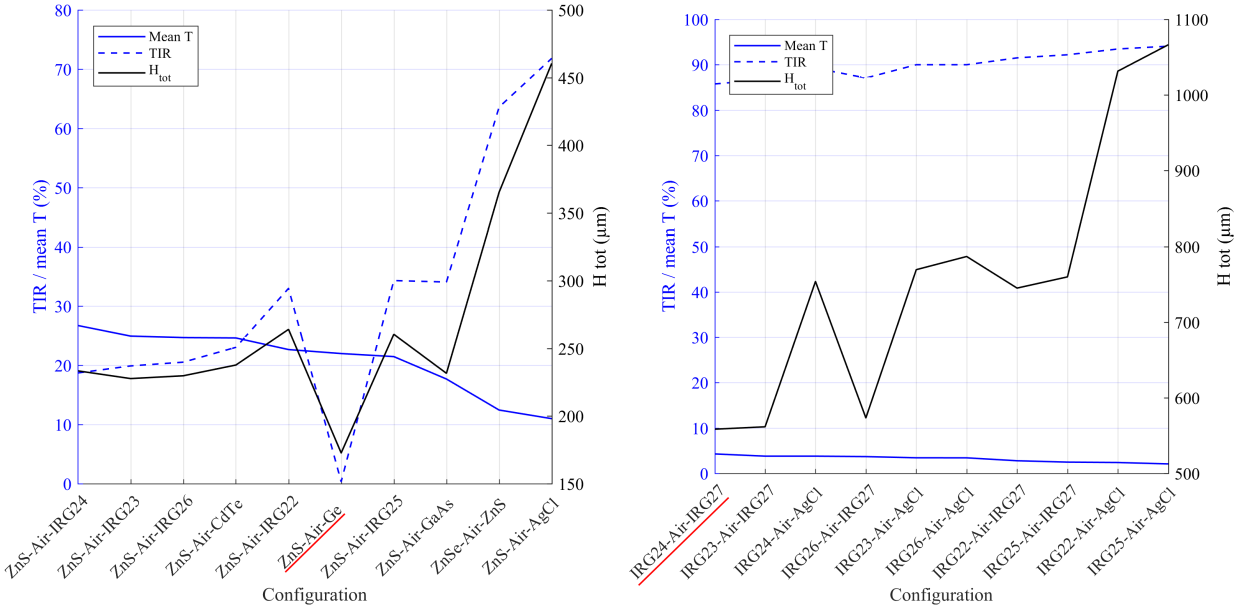
In this section, we use the proposed GO-MSM to provide a selection of MLDOE combinations for two and three-layer DOEs. The validation of this method is performed in Section 11.5, using FDTD as a reference.

We fix the same design parameters as in Section 11.2.2: The $F/\sharp$ is 15 at $\lambda = 8$ µm and the number of zones N is set to 10, giving an aperture radius $D = 10$ mm. These standard parameters provide a relatively small number of zones and aperture diameter, considering a potential manufacturing process.

In this chapter, we classify IR materials as "hard" (ZnS, Ge, ZnSe, CdTe and GaAs) or "soft" (IRG22-27 and AgCl). These denotations come from the relative hardness property of these materials (See Annex 16). Since a softer material is more easily manufactured, we perform the GO-MSM considering all IR materials or only soft materials. These two analyses are made for two-layer DOEs and three-layer DOEs, resulting in four different optimisation results.

11.4.1 two-layer DOE optimal configuration

The GO-MSM presented in Section 11.3 is applied in the case of two-layer DOEs. The gap is made of air, while the materials of both layers are variable. The results of the GO-MSM are presented in Figures 11.5a and 11.5b, respectively for all-IR materials and soft-only materials.



(a) All-IR materials.

(b) Soft materials.

Figure 11.5: On-axis geometric selection method results for the ten "best" material configurations with an air gap. The design parameters are $D = 10$ mm, F/15 and $N = 10$. Three evaluation metrics are shown: transmission (blue line, left axis), TIR (blue dotted line, left axis) and thickness (black line, right axis). The "best" MLDOE combinations are underlined in red and correspond to solutions having a low TIR, thickness, and high transmission.

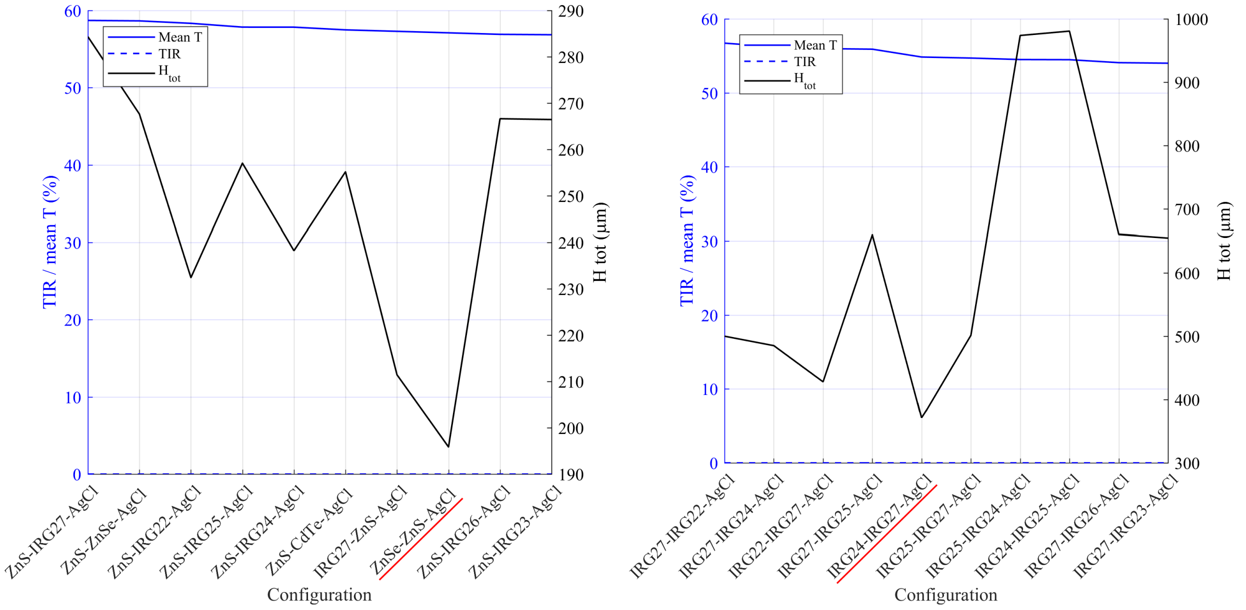
Interestingly, Figures 11.5a and 11.5b provide very different results. No combination from the soft-only selection process can be used for an efficient optical design: the average transmission is only 5% while the TIR averages 90%. The "best" solution is IRG24-air-IRG27, displaying the lowest thickness in this case.

In Figure 11.5a, while the material of the first layer is a variable, the final solutions are all made of ZnS-air-X. Nevertheless, while ZnS seems to couple well with an air gap, especially in the ZnS-air-Ge configuration, the remaining possibilities are rather bad, displaying low transmission and high TIR. ZnS-air-Ge is the only combination that generates no TIR, which coincides with its shallow thickness compared to the other solutions.

Contrary to the PIDE results of Figure 11.2a, where at least four air-gap configurations had high imaging performance, even at 15° incidence, the result of the GO-MSM is much less optimistic. Note that the "optimal" configurations are different between our GO-MSM and the reviewed PIDE-MSM.

11.4.2 three-layer DOE optimal configuration

To improve the performance, we consider three-layer DOEs, replacing the air gap with an IR material. We keep the same distinction between all-IR materials and soft-only materials. The results are depicted in Figure 11.6.



(a) All-IR materials.

(b) Soft materials.

Figure 11.6: On-axis geometric selection method results for the ten "best" material configurations with all layer/gap materials variables. The design parameters are $D = 10$ mm, F/15 and $N = 10$. Three evaluation metrics are shown: transmission (blue line, left axis), TIR (blue dotted line, left axis) and thickness (black line, right axis). The "best" MLDOE combinations are underlined in red and correspond to solutions having a low TIR and thickness, and a high transmission.

Figure 11.6 leads to much better solutions than Figure 11.5. The optimal solution considering all-IR materials (Figure 11.6a is ZnSe-ZnS-AgCl. It has the lowest thickness (195 µm), a relatively high transmission (57%) and null TIR. The best solution, obtained with soft-only materials (Figure 11.6b, is IRG24-IR27-AgCl. It has the same transmission value as the ZnSe-ZnS-AgCl solution but a higher thickness (370 µm).

In the following Section (11.5), we derive the optical performance of all the presented solutions for the PIDE-MSM and the GO-MSM. We analyze if their expected optical quality matches their FDTD performance. If they do, it means that the selection model is reliable, regardless of the actual performance level of the configurations.

11.5 Validation of the GO-MSM: polychromatic Strehl ratio

This comparative section draws the image quality of the optimal solutions provided by the PIDE-MSM and our GO-MSM. We use the Strehl ratio as an evaluation metric computed at the best focal plane. FDTD rigorous wave simulation is performed using OptiFDTD software [43]. The FDTD sampling is the same as in Chapter 4. The simulation process (Similar to Figures 4.1 and 10.1) is described in Figure 11.7 [86]:

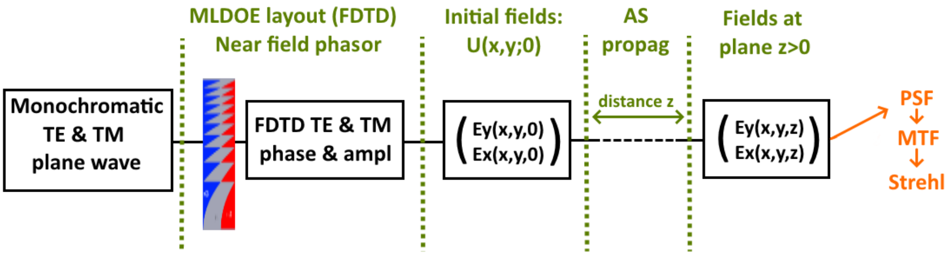


Figure 11.7: Numerical simulation process to evaluate the MLDOE optical performance (PSF, MTF, Strehl ratio). An input plane wave is delayed by an MLDOE, and the resulting vectorial near field is retrieved using the rigorous FDTD method. This input field ($z = 0$) is propagated to any plane $z > 0$ along the optical axis using Fourier optics angular spectrum, generating a vectorial output field. The "best" focal plane is defined as the plane providing the highest Strehl ratio value, where optical metrics are generated [86].

Figure 11.7 hybrid process is used to study the output performance of MLDOE designs obtained with the PIDE-MSM and our GO-MSM. The aim is not to compare MLDOE combinations but to analyze if the GO-MSM or the PIDE-MSM had predicted their optical performance. Practically, we compute the Strehl ratio at the best focal plane of each "optimal" solution derived in the previous sections:

- PIDE-MSM: IRG24-air-IRG25 and IRG22-IRG25-CdTe, Figure 11.2

- Two-layer GO-MSM: ZnS-air-Ge and IRG24-air-IRG27, Figure 11.5

- Three-layer GO-MSM: ZnSe-ZnS-AgCl and IRG24-IRG27-AgCl, Figure 11.6

For each of the six combinations, multiple FDTD calculi have been made to obtain the polychromatic variation of the Strehl ratio. As stated in Section 11.4.1 and 11.4.2, the input design parameters are the same for all MLDOE combinations: $D = 10$ mm, F/15 and $N = 10$.

Figure 11.8 shows the polychromatic evolution of the Strehl ratio for the MLDOE design solutions mentioned above:

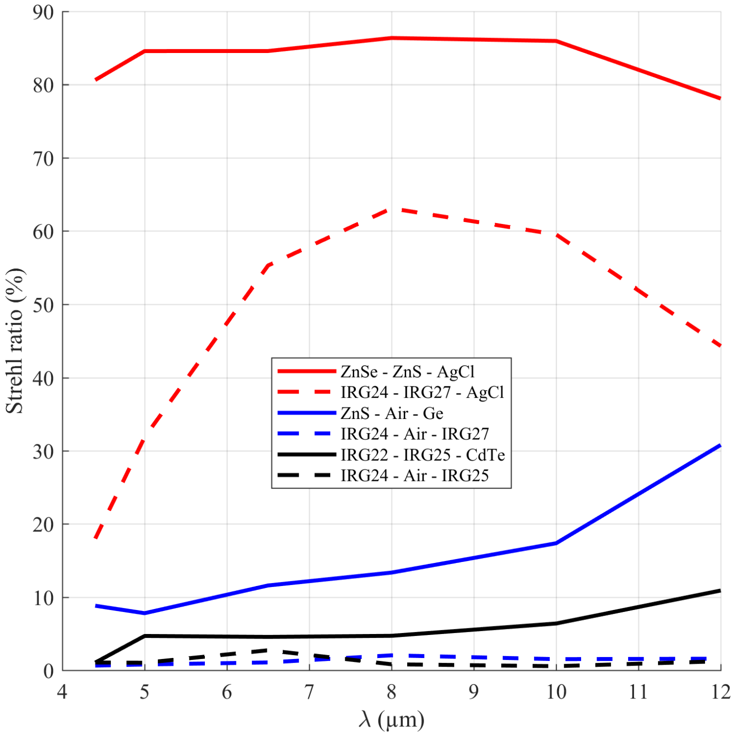


Figure 11.8: Polychromatic evolution of the Strehl ratio, computed at the "best" focal plane, following Figure 11.7 procedure. PIDE-MSM (black curves), two-layer GO-MSM (blue curves) and three-layer GO-MSM (red curves). The polychromatic Strehl ratios of the PIDE-MSM solutions are worse than predicted in Figure 11.2.

Figure 11.8 displays the six "best" MLDOE material combinations obtained using the GO-MSM and the PIDE-MSM. The IRG24-air-IRG27 configuration, resulting from Figure 11.5b, has a polychromatic Strehl ratio of less than 5%, which is in agreement with its very low transmission, very high TIR and thickness. Similarly, the ZnS-air-Ge configuration also performs poorly, as expected by its transmission value of only 22% in Figure 11.5a.

The two "optimal" three-layer DOEs configurations: ZnSe-ZnS-AgCl and IRG24-IRG27-AgCl, resulting from Figure 11.6, display a relatively high performance for all wavelengths, especially ZnSe-ZnS-AgCl. Both have a ∼60% transmission and no TIR, but the ZnSe-ZnS-AgCl MLDOE has a ∼ 200 µm thickness while the IRG24-IRG27-AgCl MLDOE has a ∼ 400 µm thickness, explaining its lower polychromatic Strehl ratio.

Finally, looking back at Figures 11.2a and 11.2b, both IRG24-air-IRG25 and IRG22-IRG25-CdTe combinations provided extremely high PIDE values (∼ 99% on-axis). This result does not concur with the black curves in Figure 11.8. The TEA, not accounting for the transmission losses and the TIR (as pointed out in Section 11.2.3), entails considerably optimistic results.

A clear link arises between the GO-MSM metrics (transmission, TIR, thickness) and the expected performance at the focal plane. An MLDOE combination with high transmission, null TIR, and low thickness presents a high polychromatic Strehl ratio (ZnSe-ZnS-AgCl). Inversely, very thick MLDOEs present high TIR and may perform very poorly (IRG24-air-IRG25). Consequently, this study shows the high potential of the GO-MSM when designing MLDOEs.

11.6 Results discussion

This chapter has proved the PIDE's inadequacy in studying and selecting MLDOE configurations. The analytical PIDE does not consider any refraction effect, thus neglecting transmission losses at optical interfaces (Fresnel reflection) and TIR. According to the phase expression in Equation 11.1, the least angle-dependant combinations are obtained when the refractive indexes of each layer are very close (with no dependency on the gap material), such as IRG24-air-IRG25. On the other hand, based on Equation 9.11, the zone height design leads to very thick MLDOE combinations, which are more likely to encounter TIR. Therefore, the "thickness" metric is of prime importance as it defines the susceptibility of an MLDOE to generate TIR, which strongly impacts its transmission.

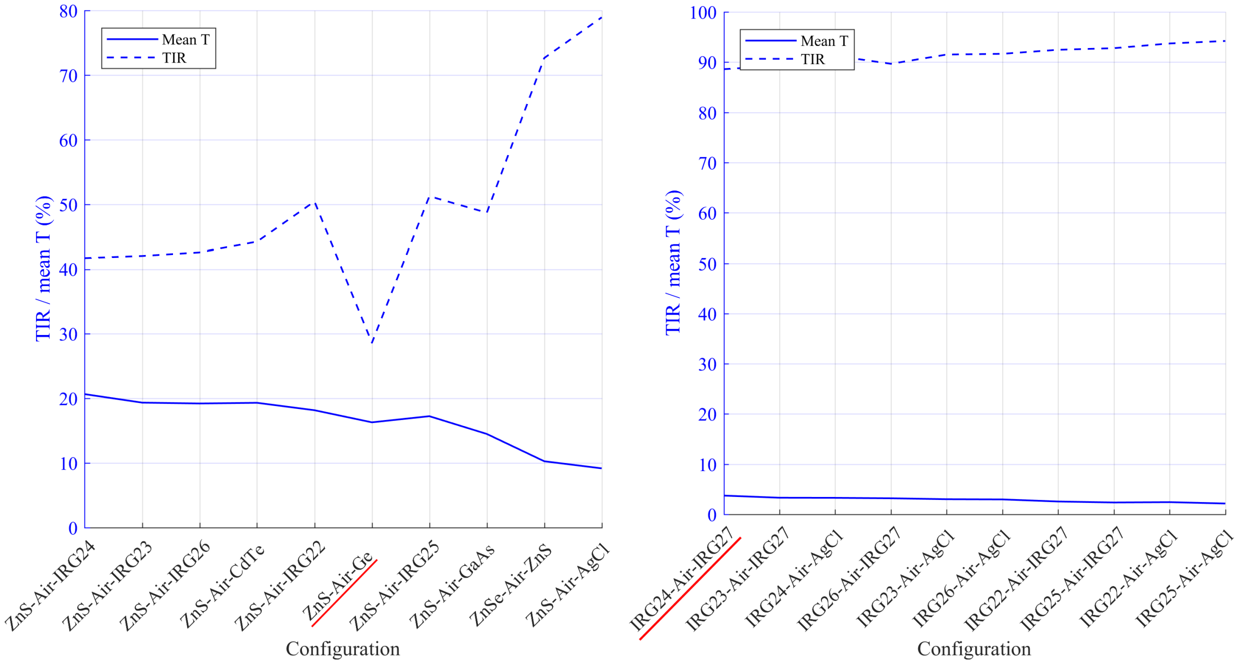
Figure 10.5 clearly shows the performance difference between MLDOE combination with and without an air gap. Due to sharper refractive index transitions between interfaces, the air gap (or any low-index gap material) potentially decreases MLDOE imaging performance. To solve this issue, an MLDOE might require a filling material with a relatively high index to avoid TIR and enhance its IR transmission. High-index materials, such as ZnS or ZnSe, tend to decrease the component thickness and provide valuable combinations. On the other hand, these materials are more difficult and expensive to manufacture due to their hardness.

Although three-layer DOEs might be challenging to manufacture and two-layer DOEs seem to perform poorly (Figure 11.5), MLDOEs are still very useful dual-band IR components. Thereby, one could use a different groove design model such as the extended scalar theory [88, 89] or the effective area method [81, 82]. The performance of two-layer DOEs might increase tremendously since these methods lower the aspect ratio of extreme zones, limiting TIR.

Our geometric optics material selection method (GO-MSM) is fast, reliable and does not require time-consuming wave simulations, except for validating the results. Therefore, it is a beneficial optimisation tool. Rigorous wave propagation has been used to confirm the GO-MSM results, which has proven to perform much better than the PIDE material selection method for studying thick MLDOEs. This chapter has also highlighted the benefit of adding a high index gap material to reduce the TIR and enhance optical performance. Since the performance obtained in this chapter are overall not excellent, we propose in Chapter 12 to optimise the shape of the studied MLDOEs, increasing their optical quality in both MWIR and LWIR.

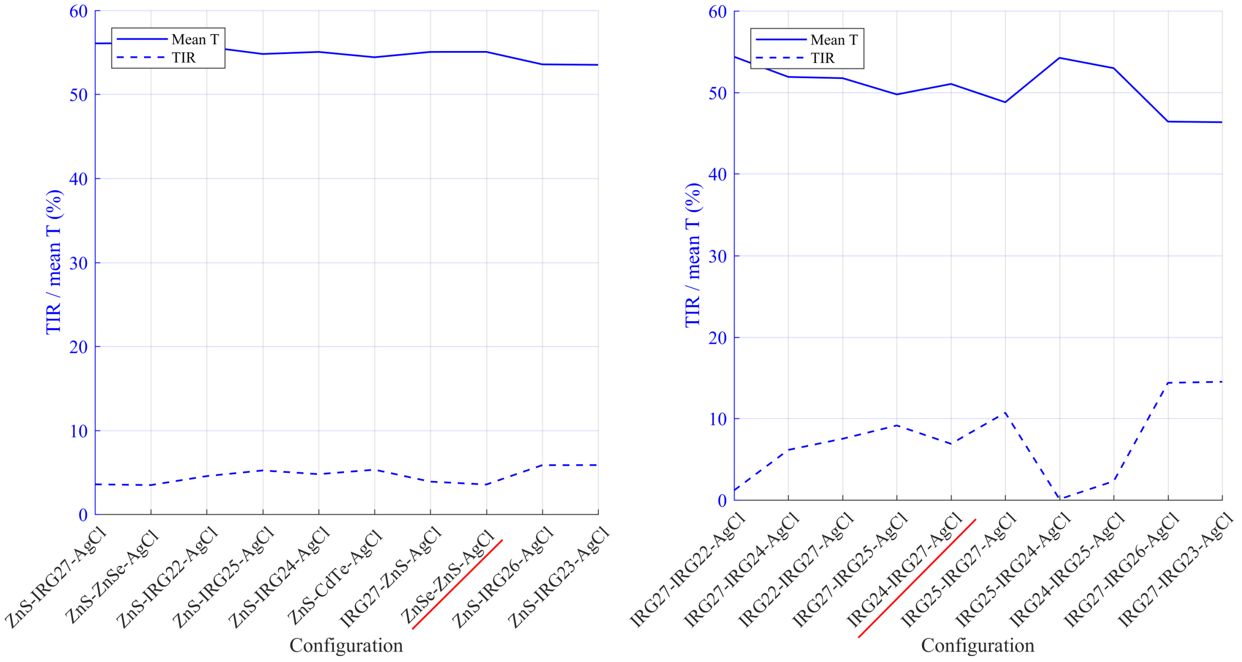
11.7 Annex

In this section, the same optimisations as in Sections 11.4.1 and 11.4.2 are performed with an incident angle of 15°. The results are displayed in Figure 11.9:



(a) Two-layer DOEs, all-IR materials.

(b) Two-layer DOEs, soft materials.



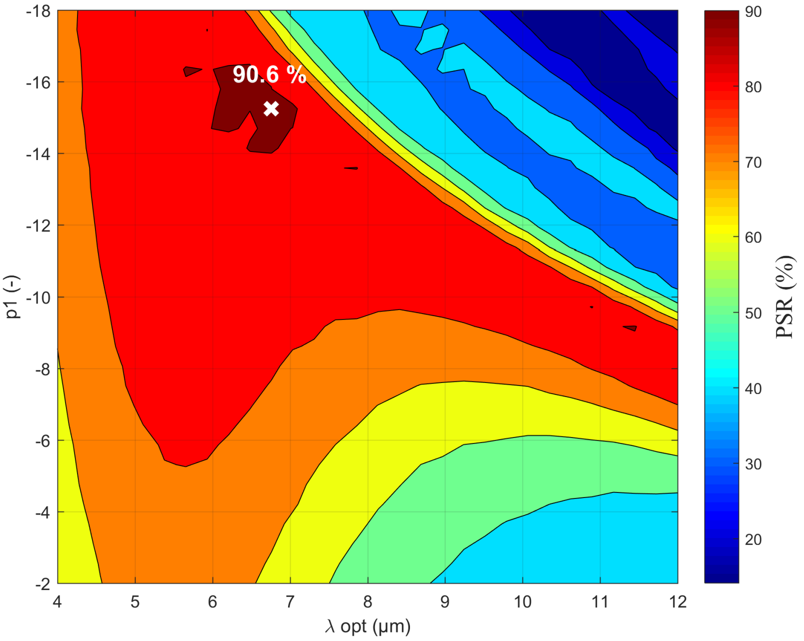
(c) Three-layer DOEs, all-IR materials.

(d) Three-layer DOEs, soft materials.

Figure 11.9: Off-axis geometric selection method results for the ten "best" material configurations considering: a) Two-layer DOEs, all-IR materials, b) two-layer DOEs, soft materials, c) Three-layer DOEs, all-IR materials, d) three-layer DOEs, soft materials. The same materials and x-label configuration have been kept from Figures 11.5 and 11.6 to ease the comparison. The design parameters are $D = 10$ mm, F/15 and $N = 10$. Three evaluation metrics are shown: transmission (blue line, left axis), TIR (blue dotted line, left axis) and thickness (black line, right axis). The "best" MLDOE combinations selected in Figures 11.5 and 11.6 are underlined in red.

As a result of Figure 11.9, an increase in TIR is noticed, especially in a) and d), but they have no impact on the selected configurations of Sections 11.4.1 and 11.4.2. The thickness is not depicted here since it does not change when off-axis incidence is considered. In

Figure 11.9d, the IRG24-IRG27-AgCl is still considered the "best" combination since it has the lowest thickness, but the IRG24-IRG27-AgCl also appears as a very good solution.


90.6 %
-18
-16
-14
-12
-10
-8
-6
-4
-2
p1 (-)
4
5
6
7
8
9
10
11
12
λ opt (μm)
90
80
70
60
50
40
30
20
PSR (%)

Chapter 12

Alternative MLDOE design: Extended scalar theory

In this chapter, we present an algorithm based on the extended scalar theory to improve an MLDOE performance by modifying its zone heights. The extended scalar theory (EST) is an alternative thickness optimisation method that depends on the diffractive order and the optimisation wavelength. We research suitable EST input parameters by combining ray-tracing and Fourier optics to provide a performance estimate for each EST parameter pair. Considering three different material combinations, we test the performance of our height optimisation algorithm using rigorous Finite-Difference Time-Domain (FDTD). Although relying on an approximate theory, our algorithm can provide much more performing MLDOE designs than the TEA in all three cases. Nonetheless, the algorithm does not ensure an optimal design. Depending on the MLDOE configuration, the performance predictions are partially accurate. At the end of this chapter we provide a very efficient MLDOE design, concluding the research of a dual-band diffractive element carried out since Chapters 7 and 8.

Contents

12.1 Introduction

Alternative scalar theories have been researched to fill the gap between the TEA and rigorous numerical methods. The effective-area method applies geometric optics to sawtooth MLDOEs to model the "shadowing effect" and perform more accurate computations of diffraction efficiency [81, 82]. The extended scalar theory (EST), introduced by Swanson [89] and recently applied to MLDOEs [88] combines the grating equation and Snell's law to optimise the profile of a DOE (reducing the shadowing effect). We have also shown how Fourier optics could be combined with ray-tracing to estimate MLDOE's optical performance at the focal plane [67] and provide a material selection method [86].

In this chapter, we propose an algorithm to find suitable groove thickness for any MLDOE. This process is based on the EST (Section 12.2) and combines the ray model (Chapter 10) with Fourier optics (Chapter 4). According to the EST, the thickness of each MLDOE layer only depends on the diffractive order p and the optimisation wavelength λ_{opt}. The algorithm performs a parametric study to estimate the pair (p, λ_{opt}) providing the "best" MLDOE design. An MLDOE's optical performance is computed at its focal plane and evaluated over the MWIR-LWIR wavebands using the Strehl ratio metric. Similarly to Chapter 10, the Strehl ratio is retrieved using the ray model and a free-space Fourier optics propagator (Section 12.3). This algorithm is applied in Section 12.5 for three MLDOE combinations, ensuing from the MLDOE material selection process introduced in Chapter 11 [86]. As a result, a suitable EST input pair (p, λ_{opt}) allows for generating an "optimal" MLDOE profile for each combination. In Section 12.6, we test the validity of the proposed algorithm by comparing the predicted optical performance with accurate FDTD simulations. The results are finally discussed in Section 12.7.

12.2 MLDOE EST design

This section describes the extended scalar theory (EST), introduced by Swanson [89], and its application to MLDOEs [88]. The EST uses Snell's law of refraction to compute the so-called "shadowing factor" for diffraction blazed grating. In opposition to the TEA, the EST considers finite DOE thickness, accounting for the negative impact of the shadowing factor on the diffraction efficiency. Consequently, Swanson has described a method to optimise any DOE's thickness based on its period and working order. By combining Snell's law of refraction and the grating equation, this method provides a grating thickness which reduces the shadowing effect.

Following this reasoning, alternative MLDOE layer thickness equations have been proposed [88]. They are based on the layout depicted in Figure 12.1.

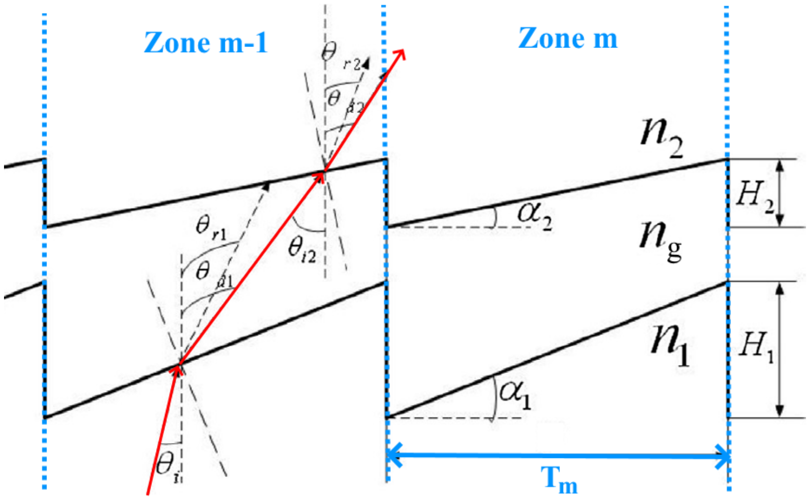


Figure 12.1: Extended scalar theory applied to an MLDOE. The deviation angle θ_d is obtained using the local grating equation while θ_r results from Snell's law. Initially, the layer heights H_1 and H_2 are constant, defined by the TEA. EST alternative profile heights are obtained when $\theta_d = \theta_r$ for both layers. T_m is the *mth* zone period. The light trajectory is displayed in red. Adapted from [88].

Following Figure 12.1 layout, we define n_1, n_g and n_2 the refractive indices of layer 1, the gap and layer 2, respectively. Similarly to a diffractive lens, the *mth* zone's aperture radius r_m and period size T_m are defined by:

$$\begin{cases} r_m^2 = m r_1^2 = 2 m f \lambda_d \\ T_m = r_{m+1} - r_m \end{cases} \tag{12.1}$$

where λ_d represents one of the two design wavelengths, and f is the associated MLDOE focal length. $\alpha_1 = H_1/T_m$ and $\alpha_2 = H_2/T_m$, with H_1 and H_2 the microstructure heights of layer 1 and 2 given in Equation 9.11. The EST height calculation for the first layer is detailed in the following.

Snell's law of refraction is applied to the first layer:

$$n_1 \sin(\theta_i + \alpha_1) = n_g \sin(\theta_{r1} + \alpha_1) \tag{12.2}$$

θ_i and θ_{r1} are respectively the incident and refracted angles. The grating equation, applied to the first layer at local period T_m, wavelength λ and diffractive order p_1, gives:

$$n_g(\lambda) \sin\theta_{d1} - n_1(\lambda) \sin\theta_i = \frac{p_1 \lambda}{T_m} \tag{12.3}$$

For a particular wavelength $\lambda = \lambda_{opt}$ and order p_1, we can equate the refractive and diffractive deviation angles ($\theta_{d1} = \theta_{r1}$) in Equation 12.3. It results in a single layer height value, $H_1^{EST}(T_m)$, for each zone m of layer 1:

$$H_1^{EST}(T_m) = \frac{p_1 \lambda_{opt}}{n_1 \cos\theta_i - \sqrt{n_g^2 - \left(\frac{p_1 \lambda_{opt}}{T_m} + n_1 \sin\theta_i\right)^2}} \tag{12.4}$$

The MLDOE layer height depends on the chosen "optimisation" wavelength λ_{opt} and order p_1. Refractive indices in Equation 12.4 are defined at λ_{opt}. The second layer EST height $H_2^{EST}(T_m)$ follows a similar calculation, detailed in [88]:

$$H_2^{EST}(T_m) = \frac{p_2\lambda_{opt}}{\sqrt{n_2^2 - \left(\frac{(p_1+p_2)\lambda_{opt}}{T_m} + n_1\sin\theta_i\right)^2} - \sqrt{n_g^2 - \left(\frac{p_1\lambda_{opt}}{T_m} + n_1\sin\theta_i\right)^2}} \tag{12.5}$$

where p_2 corresponds to the operating diffractive order of layer 2. Since this chapter focuses on the on-axis MLDOE design and performance, we assume $\theta_i = 0°$ in the following. A major difference with the TEA arises since the optimal heights are now period-dependent. Consequently, each MLDOE diffractive zone m will have a unique thickness $H_{EST}(T_m)$. Equation 12.5 supposes that $p_1+p_2 = 1$, which is imposed in this whole thesis to ensure that the studied MLDOE designs are converging (see Chapter 9).

The diffractive orders p_1 and p_2 have been expressed in Chapter 9 and in [88], based on the harmonic diffractive optical elements (HDOEs) theory [64, 65]:

$$\begin{cases} p_1 = \dfrac{H_1}{\lambda_{opt}}[n_1(\lambda_{opt}) - n_g(\lambda_{opt})] \\ p_2 = \dfrac{H_2}{\lambda_{opt}}[n_2(\lambda_{opt}) - n_g(\lambda_{opt})] \\ \Rightarrow p_1 + p_2 = 1 \end{cases} \tag{12.6}$$

The relation $p_1 + p_2 = 1$ is ensured using the TEA heights definition of Equation 9.11. Since Equation 12.6 is based on HDOE's design equations, and therefore relies on the TEA, it might provide unreliable order values (see Section 12.6).

12.3 Monochromatic Strehl ratio simulation

In this section, we detail a procedure to retrieve the Strehl ratio and the focal length of any MLDOE design. This simulation process has been introduced in Chapter 10 [67] (see Figure 10.1) and is depicted in Figure 12.2:

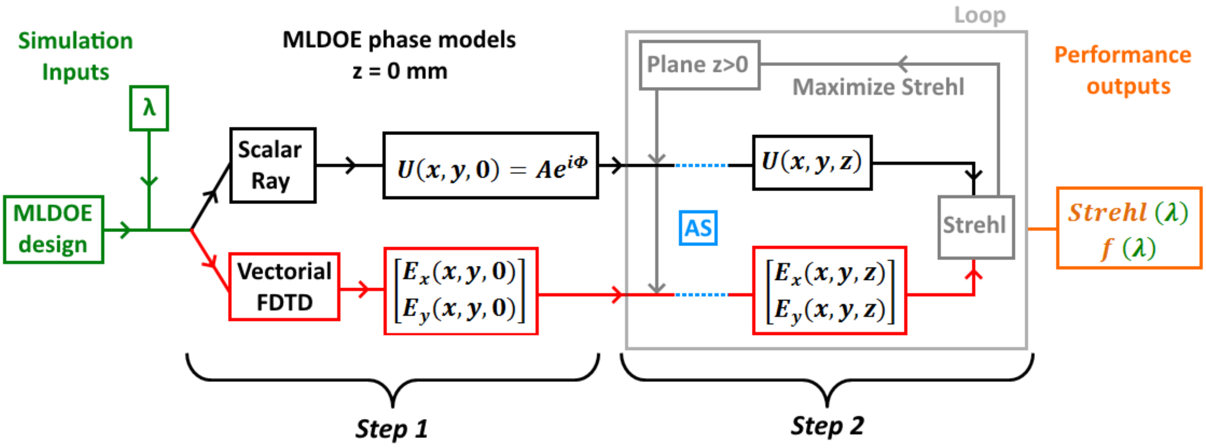


Figure 12.2: Diagram showing the hybrid procedure to retrieve the Strehl ratio and the focal length of any MLDOE design. This procedure is divided into two steps: 1) modelling the MLDOE near-field and 2) using Fourier optics (angular spectrum method) in free space to retrieve the two outputs mentioned above. Step 1 involves two approaches: one is scalar (top black), based on ray-tracing and called "ray model" [67], while the other is vectorial (bottom red), based on the Finite Differences Time Domain method (FDTD).

The Strehl ratio retrieval procedure depicted in Figure 12.2 combines near-field modelling (step 1) and free-space optical propagation (step 2). The former involves the scalar ray model (Chapter, 10 [67]) as well as the rigorous FDTD approach (Chapter 6). The FDTD sampling corresponds to Figure 6.10.

The optical free-space propagation is simulated using the angular spectrum of plane waves described in Equation 4.22. The procedure depicted in Figure 12.2 takes an MLDOE design as input (i.e. the periods T_m and thicknesses $H_{T_m}^{EST}$ for any zone m) and is purely monochromatic. The loop, displayed in grey, allows determining precisely the "best" focal plane $f(\lambda)$. The input wavelength λ is called "simulation wavelength" and is unrelated to the previously defined optimisation wavelength λ_{opt}. It clearly appears from Figure 12.2 loop that the "best" focal plane is found iteratively and might not correspond to the analytical expression (Equation 9.28), as already shown in Chapter 10.

Finally, we refer to Figure 12.2 as "monochromatic ray simulation" when the ray model is used or as "monochromatic FDTD simulation" when FDTD is used instead.

12.4 MLDOE profile optimisation algorithm

In this section, we develop an algorithm to determine suitable layer thicknesses for any MLDOE. This algorithm researches the EST input pair (order p_1 and optimisation wavelength λ_{opt}) leading to the most performing MLDOE design. As shown in Equations 12.4 and 12.5, selecting the pair (p_1, λ_{opt}) greatly impacts the resulting MLDOE layer heights.

Recall that the diffractive orders of both layers are linked ($p_1 + p_2 = 1$) and that Equation 12.6 relates p_1 to λ_{opt}. Since this relation is based on the HDOE theory [64, 65], which involves the TEA, we perform two parametric studies. On one hand we consider λ_{opt} as the only design variable (p_1 is defined according to Equation 12.6) while, on the other hand, we vary p_1 and λ_{opt} independently.

Some global design parameters are arbitrarily fixed for the remaining of the chapter. The studied MLDOE designs have a $D = 10$ mm aperture diameter for a total of $N = 10$

diffractive zones. The F-number (F/#) is set to 15 (at $\lambda = 8$ µm) and the design wavelengths are selected in the middle of both MWIR and LWIR wavebands: $\lambda_1 = 4.7$ µm and $\lambda_2 = 10$µm. The period T_m of each zone m is defined in Equation 9.1. Figure 12.3 describes the proposed thickness optimisation algorithm:

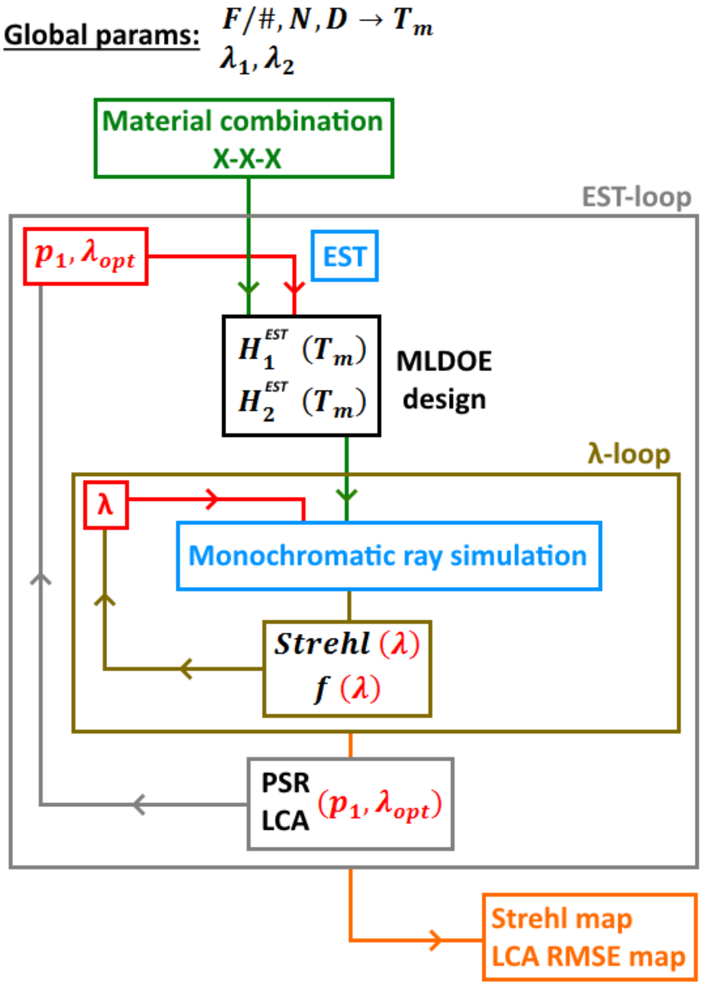


Figure 12.3: Diagram describing the proposed MLDOE thickness optimisation algorithm. It is divided into two loops and searches the "best" EST input doublet (p_1, λ_{opt}). A set of layer materials and a pair (p_1, λ_{opt}) defines an "MLDOE design". The green loop involves the ray model and Figure 12.2 diagram. It is performed over the MWIR-LWIR wavebands. At the end of the loop, the polychromatic Strehl ratio (PSR) is obtained for a pair (p_1, λ_{opt}). This process is repeated in the main loop (gray) for each value of (p_1, λ_{opt}), resulting in either a PSR curve (if Equation 12.6 is applied) or a PSR two-dimensional map. Given the above-mentioned global parameters, this output map (or curve) is specific for each MLDOE material combination.

The procedure depicted in Figure 12.3 provides a polychromatic Strehl ratio (PSR) mapping. According to Equations 12.4 and 12.5, each pair (p_1, λ_{opt}) provides a specific MLDOE design (thicknesses H_2^{EST} and H_1^{EST}), that can be evaluated using the PSR. We define the PSR over a waveband $\Delta_\lambda = \lambda_{max} - \lambda_{min}$ as:

$$PSR_{\Delta_\lambda}(p_1, \lambda_{opt}) = \frac{1}{\lambda_{max} - \lambda_{min}} \int_{\lambda_{min}}^{\lambda_{max}} Strehl(\lambda; p_1, \lambda_{opt})\, \mathrm{d}\lambda \tag{12.7}$$

The total infrared PSR (denoted by "PSR" in Figure 12.3 for conciseness) is the weighted sum of two PSRs, one calculated in MWIR and the other in LWIR:

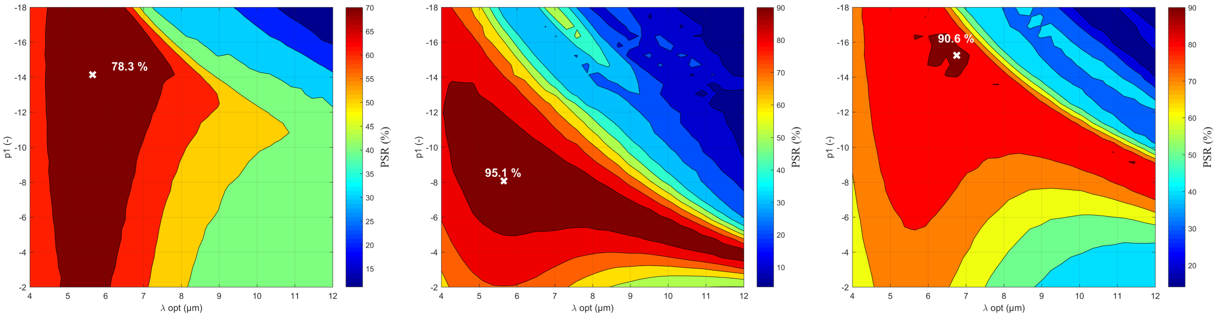
$$PSR(p_1, \lambda_{opt}) = \frac{1}{2} PSR_{MWIR}(p_1, \lambda_{opt}) + \frac{1}{2} PSR_{LWIR}(p_1, \lambda_{opt}) \tag{12.8}$$

For each pair (p_1, λ_{opt}), the PSR is retrieved using the monochromatic ray simulation (Figure 12.2, ray model) for multiple MWIR-LWIR wavelengths (green loop).

12.5 Results of the optimisation algorithm

In this section, we detail the results of the proposed algorithm (Figure 12.3). We apply this EST input selection process to three MLDOE combinations, ensuing from the MLDOE material selection process introduced in Chapter 11 [86]. We have selected the IRG24-air-IRG27, ZnS-air-Ge and IRG24-IRG27-AgCl combinations, regardless of their potential manufacturing issues. The IRG24 and IRG27 materials belong to SCHOTT's chalcogenide family [53]. They are "softer" than ZnS and Ge, with lower refractive indices and can potentially be moulded (see Chapter 13). Since two-layer DOEs (i.e. air-gap MLDOEs) are more easily manufactured, we have selected an "all chalcogenide" solution (IRG27-air-IRG27), a more "classical" solution (ZnS-air-Ge) and finally, a three-layer solution (IRG24-IRG27-AgCl) made of "soft" materials. It has been shown in Chapter 11 [86] that these MLDOE combinations performed relatively poorly (see Section 12.6), especially for IRG24-air-IRG27 and ZnS-air-Ge.

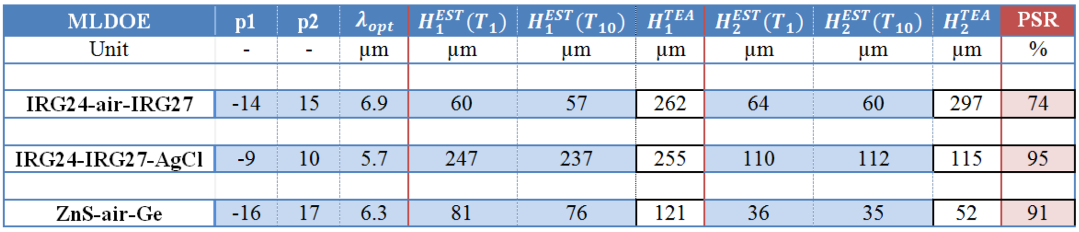
The result of the proposed optimisation algorithm is displayed in Figure 12.4, assuming that p_1 and λ_{opt} vary independently:



(a) IRG24-air-IRG27 (b) IRG24-IRG27-AgCl (c) ZnS-air-Ge

Figure 12.4: PSR maps for the three selected MLDOE combinations: a) IRG24-air-IRG27, b) IRG24-IRG27-AgCl, and c) ZnS-air-Ge. These maps result from the application of Figure 12.3 algorithm, with order p_1 and wavelength λ_{opt} both variable and independent. A white cross displays the maximal PSR value. λ_{opt} varies within the MWIR-LWIR wavebands while $|p_1|$ is kept relatively small to obtain "thinner" layers.

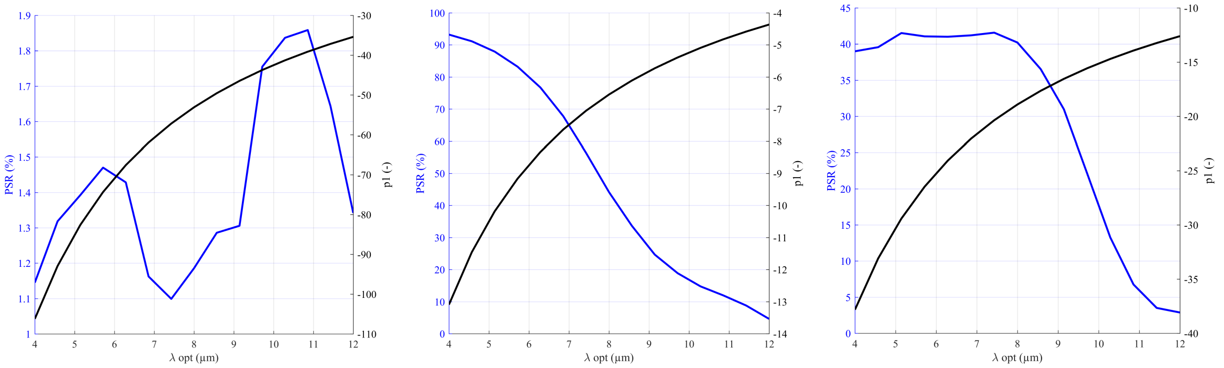
The PSR maps depicted in Figure 12.4 provide "optimal" pairs (p_1, λ_{opt}) for all three selected MLDOEs. The results are shown in Table 12.1:



MLDOE	p1	p2	λ_{opt}	$H_1^{EST}(T_1)$	$H_1^{EST}(T_{10})$	H_1^{TEA}	$H_2^{EST}(T_1)$	$H_2^{EST}(T_{10})$	H_2^{TEA}	PSR
Unit	-	-	µm	µm	µm	µm	µm	µm	µm	%
IRG24-air-IRG27	-14	15	6.9	60	57	262	64	60	297	74
IRG24-IRG27-AgCl	-9	10	5.7	247	237	255	110	112	115	95
ZnS-air-Ge	-16	17	6.3	81	76	121	36	35	52	91

Table 12.1: Algorithm design results assuming both (p_1, λ_{opt}) are variable and independent. Their selected values correspond to the maximum PSR for all three MLDOE combinations in Figure 12.4. $H_{1,2}^{EST}(T_1)$ and $H_{1,2}^{EST}(T_{10})$ are the resulting layer heights, respectively corresponding to the central zone (period T_1) and extreme zone (period T_{10}). For comparison, TEA-based heights are computed for both layers using Equation 9.11.

The next study relies on Equation 12.6 to define the diffractive order p_1 [88]. Only λ_{opt} varies freely, leading to the one-dimensional curve outputs of Figure 12.5:



(a) IRG24-air-IRG27 (b) IRG24-IRG27-AgCl (c) ZnS-air-Ge

Figure 12.5: PSR and diffractive order p_1 curves for various optimisation wavelengths λ_{opt}, for the three selected MLDOE combinations: a) IRG24-air-IRG27, b) IRG24-IRG27-AgCl and c) ZnS-air-Ge. The PSR curves (blue, left axis) result from the application of Figure 12.3 algorithm, assuming that the order p_1 (black curves, right axis) is defined by Equation 12.6. λ_{opt} varies within the MWIR-LWIR wavebands.

The PSR curves shown in Figure 12.5 provide different (p_1, λ_{opt}) pairs than in Figure 12.5. The diffractive order values obtained by Equation 12.6 are rather different, especially for the IRG24-air-IRG27 and the ZnS-air-Ge combinations. For these two designs, the maximal PSR output values, according to the ray model, are also much lower than in Figure 12.4, particularly for IRG24-air-IRG27. The "optimal" results of the proposed algorithm, in this case, are displayed in Table 12.2:

MLDOE	p1	p2	λ_{opt}	$H_1^{EST}(T_1)$	$H_1^{EST}(T_{10})$	H_1^{TEA}	$H_2^{EST}(T_1)$	$H_2^{EST}(T_{10})$	H_2^{TEA}	PSR
Unit	-	-	µm	µm	µm	µm	µm	µm	µm	%
IRG24-air-IRG27	-39	40	11	260	144	262	264	151	297	1.8
IRG24-IRG27-AgCl	-13	14	4	253	242	255	115	118	115	93
ZnS-air-Ge	-19	20	8	124	106	121	53	50	52	40

Table 12.2: Algorithm design results assuming a variable λ_{opt} and an order p_1 following Equation 12.6. The selected (p_1, λ_{opt}) values correspond to the maximum PSR for each of the three MLDOE combinations. $H_{1,2}^{EST}(T_1)$ and $H_{1,2}^{EST}(T_{10})$ are the resulting layer heights, respectively corresponding to the central zone (period T_1) and extreme zone (period T_{10}). TEA-based heights are computed for both layers using Equation 9.11, for comparison.

When p_1 and λ_{opt} are variable, the proposed algorithm can find profile heights leading to high MLDOE PSRs for each of the three studied MLDOEs. It is not the case anymore for the IRG24-air-IRG27 and ZnS-air-Ge combinations when Equation 12.6 is applied. Their PSR values drop drastically, meaning that the design Equation 12.6 is not reliable, at least according to the ray model. Since the latter is a scalar approximation, we use the rigorous electromagnetic FDTD method to validate the presented results.

12.6 FDTD results validation

In this Section, we use the "optimal" EST input pairs (p_1, λ_{opt}) obtained in Section 12.5. The denomination "map optimisation" refers to Figure 12.4, whereas the "curve optimisation" refers to Figure 12.6, where Equation 12.6 is applied.

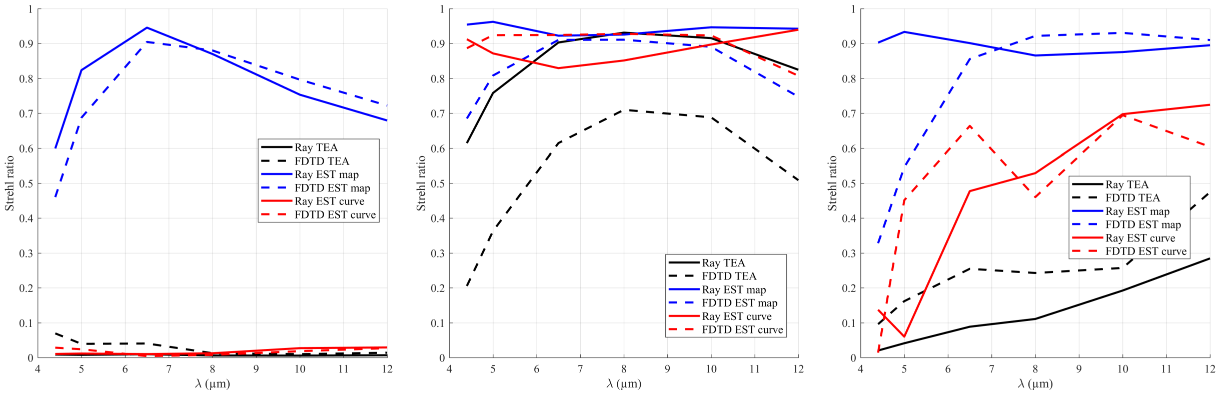
Each MLDOE design's performance is retrieved over various wavelengths in the MWIR-LWIR wavebands, using the monochromatic FDTD simulation procedure described in Figure 12.2. This process is accurate since both FDTD and the angular spectrum method provide exact fields (as long as the sampling is fine enough, see Figure 6.10).

In Figure 12.6, "Ray TEA" and "FDTD TEA" refer to Strehl ratio curves respectively obtained using the ray model and FDTD, assuming the standard TEA groove profile (Equation 9.11). Similarly, we refer to "Ray EST map", "FDTD EST map", "Ray EST curve", and "FDTD EST curve" as the Strehl ratio curves retrieved with the ray model or FDTD, considering the map or curve optimisation.

The results shown in Figure 12.6 are very variable, depending on the MLDOE combination. For example, in Figure 12.6a, the proposed algorithm is able to increase significantly the Strehl ratio. Inversely, the design Equation 12.6 is not reliable since it provides a very low Strehl ratio.

The IRG24-IRG27-AgCl combination studied in Figure 12.6b already has a relatively high TEA Strehl ratio. The proposed algorithm still finds a better design, though the ray model is overly optimistic, especially in MWIR. If Equation 12.6 is used to define the order (Figure 12.5b), the Strehl ratio increases to 90% (red dashed curve).

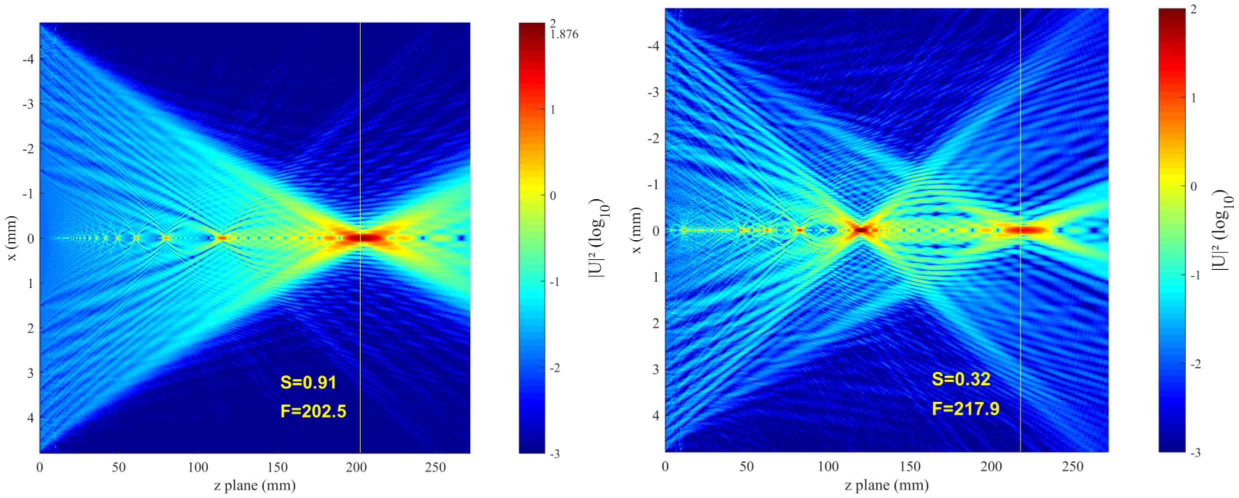
Finally, Figure 12.6c shows that the ZnS-air-Ge MLDOE can reach 90% Strehl ratio in LWIR (blue dashed curve), but the ray model does not predict the high MWIR Strehl ratio decrease. Applying Equation 12.6 provides an "average" design, with a higher Strehl ratio than in the TEA case (although still very low in MWIR) but lower than using the map (Figure 12.4c).



(a) IRG24-air-IRG27 (b) IRG24-IRG27-AgCl (c) ZnS-air-Ge

Figure 12.6: Strehl ratio for various wavelengths in the MWIR-LWIR wavebands, considering the three studied MLDOE configurations: a) IRG24-air-IRG27, b) IRG24-IRG27-AgCl, and c) ZnS-air-Ge. For each combination, six curves are depicted: The continuous curves are made with the ray model, whereas the dashed curves are obtained with FDTD and serve as a reference. The map optimisation, curve optimisation and TEA profile (i.e. no optimisation) results are displayed in blue, red and black. The black curves are exactly the same as in Figure 11.8 [86].

We now study the MWIR divergence between the ray model and FDTD for the ZnS-air-Ge combination. We consider the blue curve in Figure 12.6c, for a wavelength of 4.4 µm. Using the monochromatic MLDOE simulation process, we record the irradiance at many propagation planes $z > 0$. as depicted in Figure 12.7:

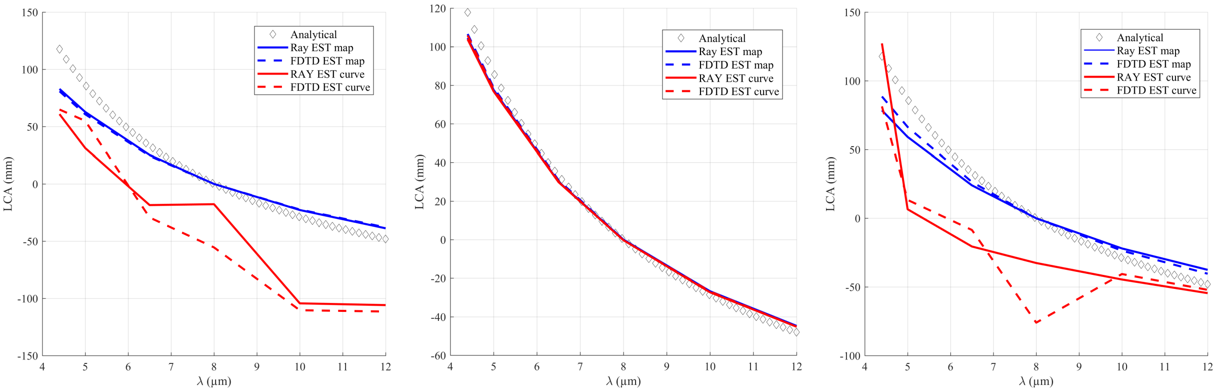


(a) ZnS-air-Ge: Ray model (b) ZnS-air-Ge: FDTD

Figure 12.7: Irradiance along the optical axis for the ZnS-air-Ge combination, using the setup given in Table 12.1 at a simulation wavelength of 4.4 µm. This simulation is made with the monochromatic MLDOE simulation process considering: a) the ray model and b) FDTD (reference). The focal length (F) and Strehl ratio (S) are displayed in yellow and correspond to the values obtained in Figure 12.6 at this wavelength (blue curves). Only FDTD can model the second diffractive order, generating a much lower Strehl ratio than the approximate ray model.

In Figure 12.7, the ray model cannot predict the second diffractive order, degrading the "best" focal plane performance. As a result, the ray model overestimates the Strehl ratio for this wavelength, which impacts the EST parameter selection and leads to a sub-optimal design choice.

Finally, the longitudinal chromatic aberration (LCA) over the MWIR-LWIR wavebands provides a design-oriented evaluation. In addition to the Strehl ratio, the monochromatic MLDOE simulation retrieves the "best" focal length $f(\lambda)$. The theoretical LCA is deduced from Equation 9.28. The LCA is displayed in Figure 12.8 for all three MLDOE combinations and both map and curve optimisations.



(a) IRG24-air-IRG27 (b) IRG24-IRG27-AgCl (c) ZnS-air-Ge

Figure 12.8: LCA over the MWIR-LWIR wavebands, considering the three studied MLDOE configurations: a) IRG24-air-IRG27, b) IRG24-IRG27-AgCl, and c) ZnS-air-Ge. For each combination, five curves are depicted: The continuous curves are made with the ray model, whereas the dashed curves are obtained with FDTD and serve as a reference. The map optimisation, curve optimisation and theoretical LCA are displayed in blue, red and black diamonds.

Considering the map optimisation in Figure 12.8, all three MLDOEs have slowly varying LCA with different slopes than the theoretical one. The ray model and FDTD provide near-identical LCA curves. In the case of the curve optimisation, except for the IRG24-IRG27-AgCl design, the resulting LCA curves are much steeper, with high variations. FDTD and ray model curves still have similar variations but display higher divergences.

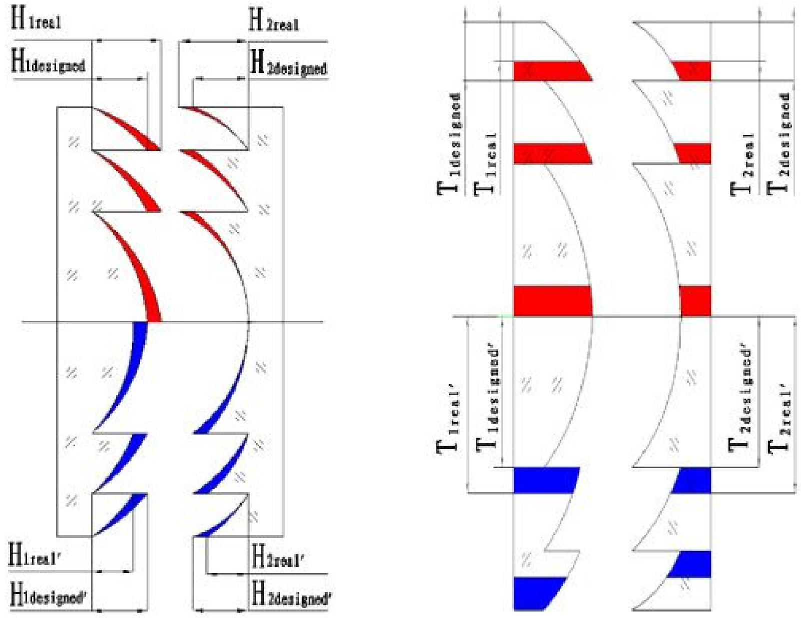
12.7 Conclusion

It has been shown in Figures 12.6 and 12.8 that the curve optimisation was not reliable. Since Equation 12.6 is based on the HDOE's theory (i.e. the TEA), it cannot provide suitable MLDOE profile heights for every MLDOE combination. Depending on the considered materials, it can generate highly performing designs (Figures 12.6b and 12.8b) as well as poorly performing profiles (Figures 12.6a and 12.8a). Therefore, the proposed algorithm must primarily use independent diffractive orders and optimisation wavelengths.

Based on the approximate ray model, our algorithm has enhanced the optical performance of all three MLDOE designs compared to their initial Strehl ratios (black curves in Figure 12.6). The resulting LCA curves are accurate compared to FDTD and as close as possible to the analytical variation. Consequently, the studied MLDOEs can be combined with refractive systems to correct the chromatic focal shift, assuming that the focal lengths are well selected.

However, as shown in Figures 12.6a, b, and c, the accuracy of the ray model in determining the Strehl ratio is variable. By extension, our proposed algorithm also provides mitigated results. For instance, the ray model is relatively accurate in Figure 12.6a, but highly overestimates the MWIR Strehl ratio of the ZnS-air-Ge configuration in Figure 12.6c. As shown in Figure 12.7, a second unpredicted diffractive order largely impacts the beam shape and degrades the Strehl ratio of the first diffractive order.

In addition, the map optimisation does not ensure that the EST parameter choice is optimal since it provides a lower performance in Figure 12.6b than the curve optimisation. In conclusion, at least for the presented MLDOE configurations, our algorithm can provide an MLDOE thickness design that enhances the TEA performance. The latter does not necessarily provide the "best" solution and is less accurate than FDTD, which must always be used for validation. The "best" focal plane position determination is precise in all studied cases. Finally, we have derived a very performing IRG24-IRG2-AgCl dual-band MLDOE design (Table 12.2).


H1real
H1designed
H2real
H2designed
H1real′
H1designed′
H2real′
H2designed′
T1designed
T1real
T2real
T2designed
T1real′
T1designed′
T2designed′
T2real′

(a) (b)

Chapter 13

MLDOE tolerances and manufacturing

Previous chapters have detailed various MLDOE shape and material optimisations, based on the ray model and verified by FDTD. Chapter 12 has demonstrated a performing MLDOE configuration (IRG24-IRG27-AgCl) excellently fitting in a dual-band camera (as in Chapter 7). To conclude the thesis and this particular MLDOE study, this chapter presents a tolerance and manufacturing analysis, based on a literature review and FDTD simulations. The IRG24-IRG27-AgCl MLDOE optical quality is tested regarding assembly and fabrication errors, through the study of various tilt, decenters, passive facets width and groove thickness manufacturing errors. This chapter finally reviews major fabrication method for micro-optics and chalcogenide moulding.

Contents

13.1 MLDOE tolerancing

This section uses the best MLDOE design obtained in Chapter 12: The IRG24-IRG27-AgCl MLDOE, designed using the EST. Recall that Table 12.2 describes its working orders and layer heights. The considered wavelength is 8 µm throughout this chapter. The following FDTD analysis use the same sampling as in Figure 6.10. After adding various errors (decenters, gap width, facets) to the IRG24-IRG27-AgCl FDTD layout, the near-fields E_x and E_y are simulated, leading to the Strehl ratio VS manufacturing error evolution, according to the process depicted in Figure 12.2.

13.1.1 Layer decenter and gap shift

We study two types of decenter tolerances. This first one is a layer decenter depicted in Figure 13.1:

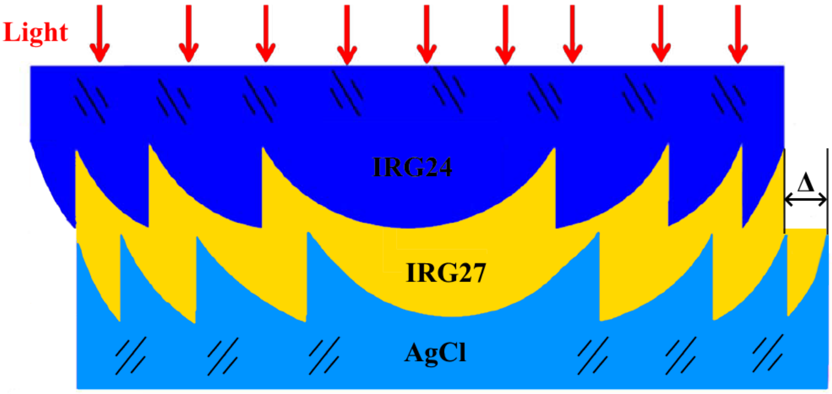


Figure 13.1: Schematic representation of a layer decenter: layer 2 is shifted by an amount Δ compared to layer 2.

A geometrical decenter tolerance model is used in [90, 91, 92], considering oblique incidence. According to this model, a decenter adds an extra optical path length (OPL_Δ) to the ideal OPL. Based on Snell's laws, its expression is derived in [92]:

$$OPL_\Delta = \Delta \frac{\sin\beta_2}{A}\left[A\sqrt{n_2^2 - B^2} + B^2 - 1\right] \tag{13.1}$$

where $B = n_1 \sin\theta$, $A = \sqrt{1 - B^2}$ and $\sin\beta_2 = H_2/\sqrt{T^2 + H_2^2}$. Recall that n_1 and n_2 represents the refractive indexes of the first and second layer, associated with the layer heights H_1 and H_2. T denotes the zone period width, and θ is the incident angle.

Figure 13.2 shows the PIDE evolution with the decenter Δ for multiple incident angles [92] and period widths:

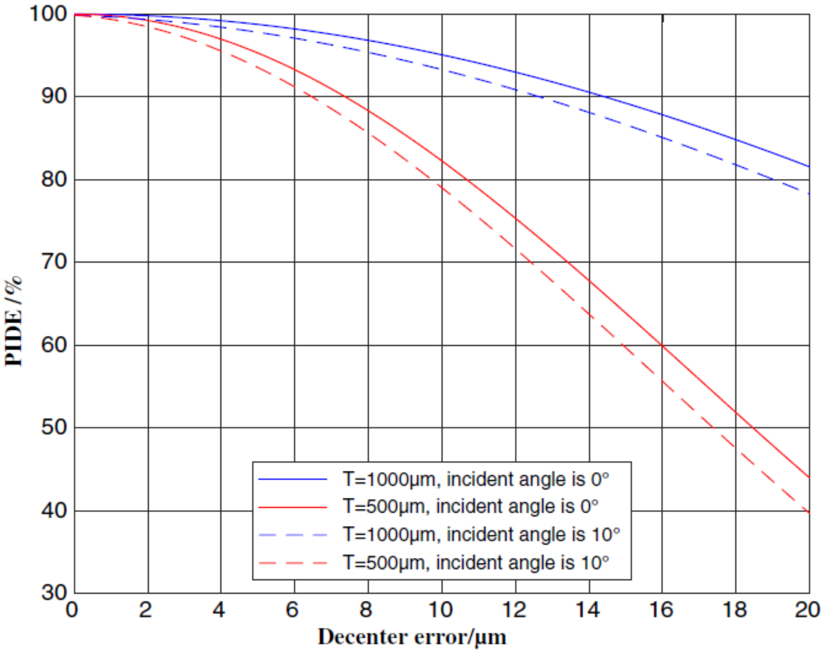


Figure 13.2: PIDE VS decenter error (µm) for various incident angles and period widths. Original source [91].

According to Figure 13.2, the diffraction efficiency drops below 80% for a 10 µm decenter and $T = 500$ µm period width. Doubling the period doubles the decenter (keeping an 80% PIDE). The effect of off-axis incidence is minimal.

This analysis shows the impact of the period size but does not provide conclusions for a whole MLDOE with varying zone periods. Furthermore, the extra decenter OPL is added to the ideal OPL to form a total phase delay. However, the latter is computed using the thin element approximation [91, 92] which is often not valid for thick MLDOEs (Figure 9.11).

In this section, we propose an alternative analysis using the MLDOE design described previously. Previous Figure 13.1 depicts the decentered layout designed and simulated in OptiFDTD for on-axis incidence. The resulting Strehl ratio for various decenters is shown in Figure 13.3

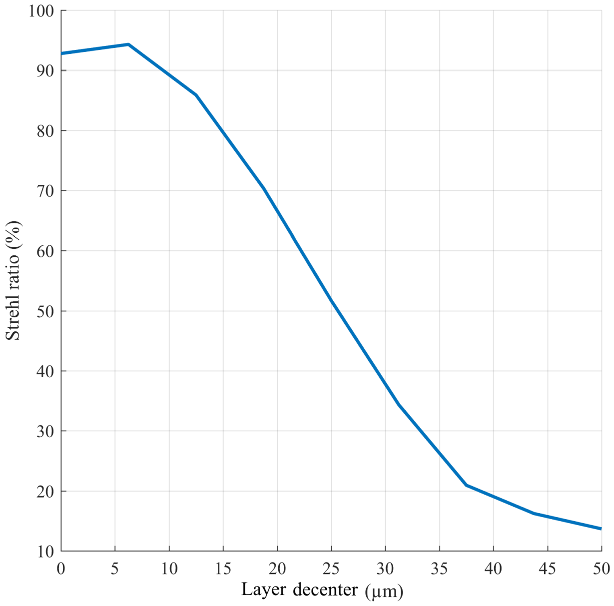


Figure 13.3: Strehl ratio for various decenter errors (μm). On-axis incidence is considered for the FDTD simulations. The "best" focal plane is constant, defined for $\Delta = 0$ μm.

In Figure 13.3, the on-axis Strehl ratio for the total MLDOE design (with varying zone periods) remains above 80% for a 15 μm decenter. This prediction is slightly more optimistic than in Figure 13.2. Following the same procedure, we simulate the impact of a gap decenter as depicted in Figures 13.4 and 13.5:

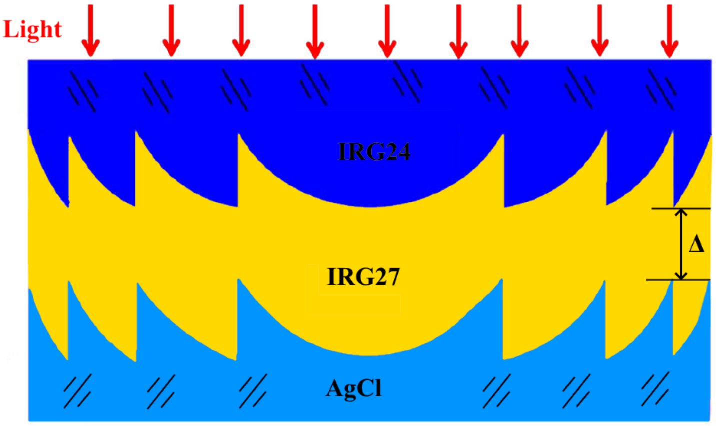


Figure 13.4: Layout showing a gap decenter Δ, increasing the distance between the two layers.

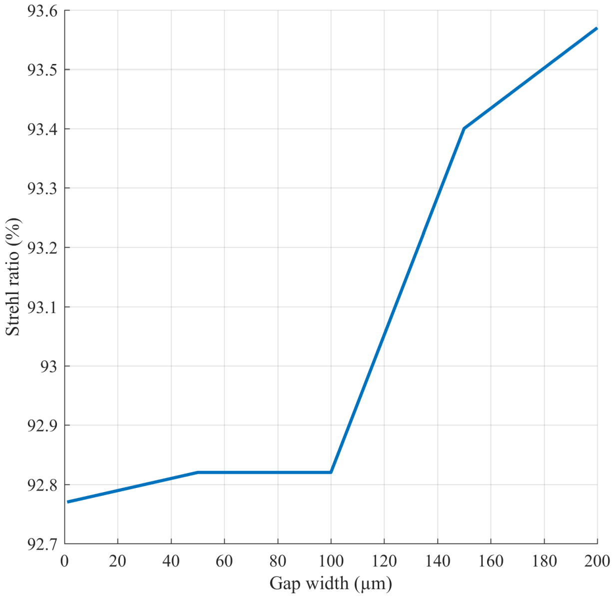


Figure 13.5: Strehl ratio for various gap decenters (µm). On-axis incidence is considered for the FDTD simulations. The "best" focal plane is constant, defined for $\Delta = 0$ µm.

Increasing the gap width has a negligible impact on the Strehl ratio and even improves it. Therefore, we conclude that an MLDOE is insensitive to gap width errors.

13.1.2 Passive facet

Cutting sharp vertical edges when manufacturing an MLDOE can be challenging. This section studies the impact of non-vertical groove edges under normal incidence. Each MLDOE diffractive zone now has a passive facet as depicted in Figure 13.6

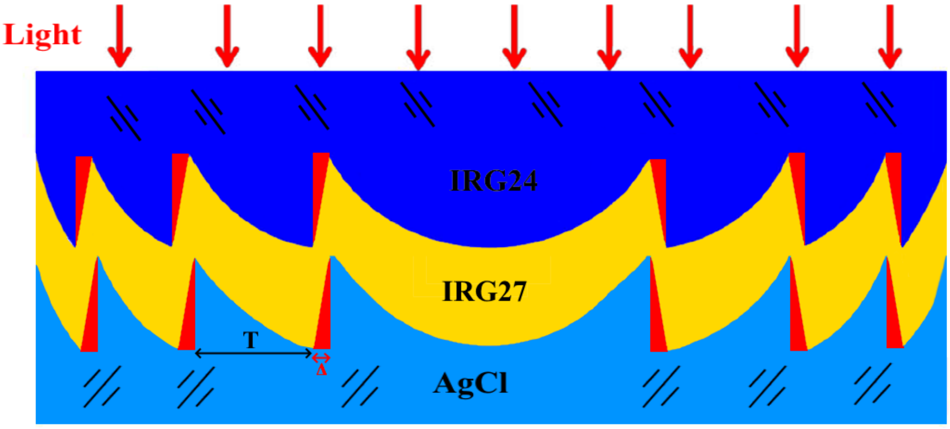


Figure 13.6: Layout displaying passive facets caused by processing errors. In this model, passive facets of constant width Δ are added to each diffractive zone in place of the ideal vertical edge.

A geometrical model has been proposed in [93] to model the influence of passive facets using the PIDE metric. The passive facet does not contribute to the diffraction efficiency and degrades it. This model is similar to the effective area method (EAM) described in [81, 82]. Practically, due to passive facets of width t_1 (layer 1) and t_2 (layer 2), the "effective MLDOE height" is reduced, following [93]:

$$\begin{cases} h_1 = H_1 \dfrac{T - t_1 - t_2.s_1}{T - t_1} \\ h_2 = H_2 \dfrac{T - t_2 - t_1.s_2}{T - t_2} \end{cases} \tag{13.2}$$

where H_1 and H_2 respectively are layer 1 and layer 2 heights, T the considered zone period. The parameters s_1 and s_2 are detailed in [93] and account for reflection, re-reflection, and refraction on the passive surface. The diffraction efficiency is expressed in [93] as:

$$\eta = \mathrm{sinc}^2 \left(1 - \frac{\Phi}{2\pi}\right) \left(1 - \frac{t_2}{T}\right)^2 \left(1 - \frac{t_1}{T}\right)^2 \tag{13.3}$$

In the last two-term model, the shadowing factor introduced by the passive facet decreases the diffraction efficiency. The result of the study proposed in [93] is shown in Figure 13.7 for the visible waveband:

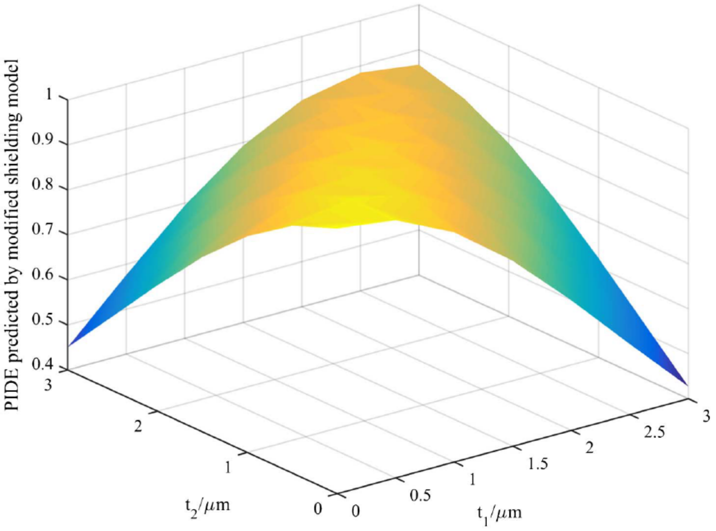


Figure 13.7: Visible PIDE predicted by the shielding model when the parameters $s_1 = 1$ and $s_2 = H_2/H_1$. Original source [93]

.

As shown in Figure 13.7, when $t_1 = t_2$, the PIDE stays at very high levels regardless of the value of t_1, t_2. The PIDE strongly decreases when the values diverge. Similar results are shown in [93] for multiple materials in the visible and for various period widths T.

A similar shadowing model, based on the layout displayed in Figure 13.8, is used to compute the effect of period width errors (with no passive facet) [94], displayed in Figure 13.9:

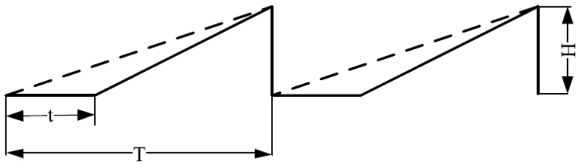


Figure 13.8: Period width error layout. Original source [94]

.

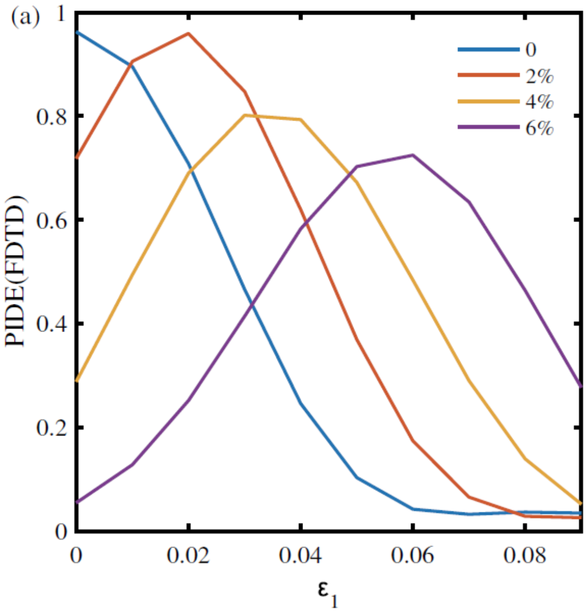


Figure 13.9: PIDE analysis based on FDTD for various period width errors. ϵ_1 and ϵ_2 represents the relative period error ($\Delta T/T$) of layer 1 and 2. Original source [94]. The visible waveband is considered.

.

Figure 13.9 provides a result similar to Figure 13.7. The PIDE keeps a high values even at high period width errors if $\epsilon_1 = \epsilon_2$.

For an IR MLDOE with varying period widths, we have studied the evolution of the Strehl ratio with the passive facet width. This study, displayed in Figure 13.10, uses the FDTD layout depicted in Figure 13.6:

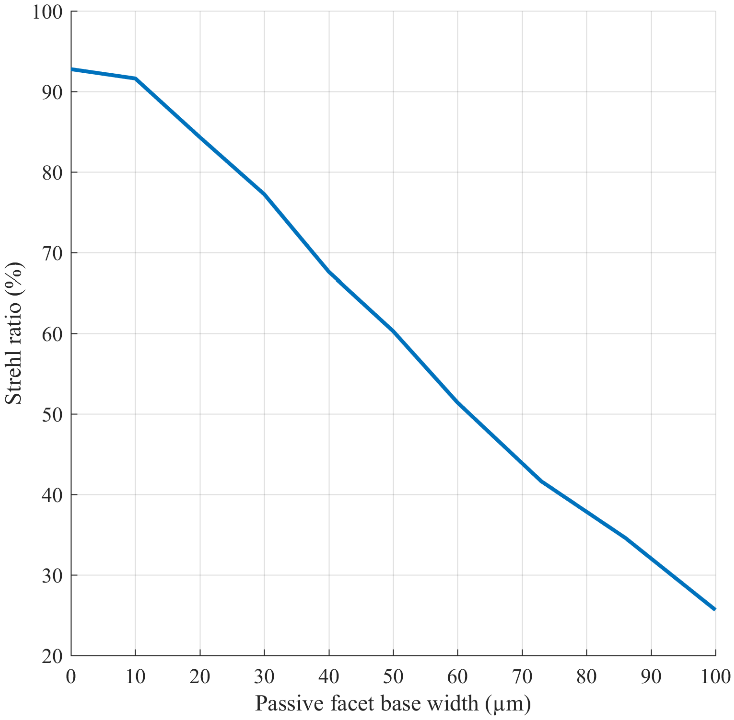


Figure 13.10: Cutting error leading to passive facets. The Strehl ratio is computed for multiple passive facet widths using FDTD. The "best" focal plane is constant, defined for $\Delta = 0$ µm.

In Figure 13.10 we consider the same passive facet widths Δ for all zones and both layers. It appears that passive facets do not degrade the Strehl ratio until $\Delta = 20$ µm, corresponding to a rather important manufacturing error. Therefore, passive facets manufacturing errors can be set on a low tolerance level.

13.1.3 Groove thickness variation

In this section, we study the impact of groove thickness errors on the IRG24-IRG27-AgCl MLDOE performance. These errors are called depth-scaling errors in [95]. Their impact is geometrically computed using the PIDE, assuming the analytical phase model derived in Equation 9.14:

$$\eta = \operatorname{sinc}^2 \left\{ m - \sum_{i=1}^{N} \frac{(d_{i,0} + \Delta d_i)}{\lambda} \left[\sqrt{n_{i,j}^2 - (n_{1,j} \sin\theta_{1,j})^2} - \sqrt{n_{i,t}^2 - (n_{1,j} \sin\theta_{1,j})^2} \right] \right\} \tag{13.4}$$

N MLDOE layers are assumed in [95]. $n_{i,j}$ and $n_{i,t}$ denote the refractive indices of the incident and emerging medium of the ith layer. Δd_i is layer i average depth error, based on the depth error of each diffractive zone of this layer. $d_{i,0}$ represents layer i ideal thickness. The oblique incidence PIDE with depth scaling error is deduced from Equation 13.4. The provided example considers an air-spaced double layer DOE made of PMMA (polymethyl methacrylate) and PC (polycarbonate). Figure 13.11 displays the result of this analytical study, assuming $\Delta d_1/d_{1,0} = \Delta d_2/d_{2,0}$:

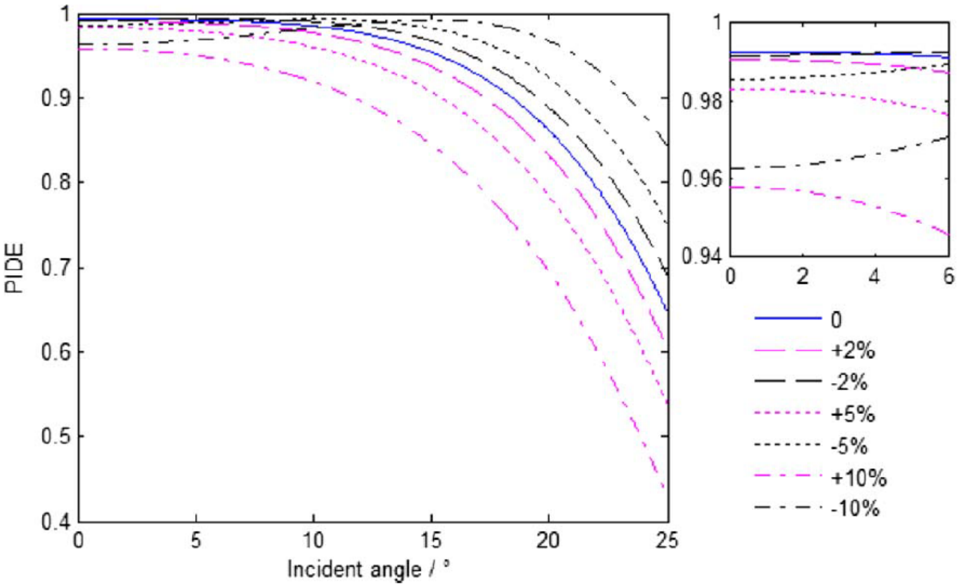


Figure 13.11: PIDE versus incident angle with various depth-scaling errors. Original source [95]. An air-spaced PMMA-PC MLDOE is considered in the visible waveband.

According to Figure 13.11 the selected PMMA-PC MLDOE performs near-perfectly until 15° incidence, even for a depth error to ideal thickness ratio of 10%. The PMMA microstructure ideal thickness is 16.46 µm, meaning that this MLDOE can withstand an average depth error $\Delta d = 1.6$ µm.

Since analytical equations rely on the TEA model, which is rarely valid for infrared MLDOEs, we provide a depth error study based on the ray model. We propose a Monte Carlo analysis computing the Strehl ratio for various MLDOE depth (or thickness) errors. A tolerance Δ defines the standard deviation of a normal distribution $3\sigma = \Delta$. This distribution is used to generate a random thickness variation for each zone, resulting in an altered MLDOE design as illustrated in Figure 13.12:

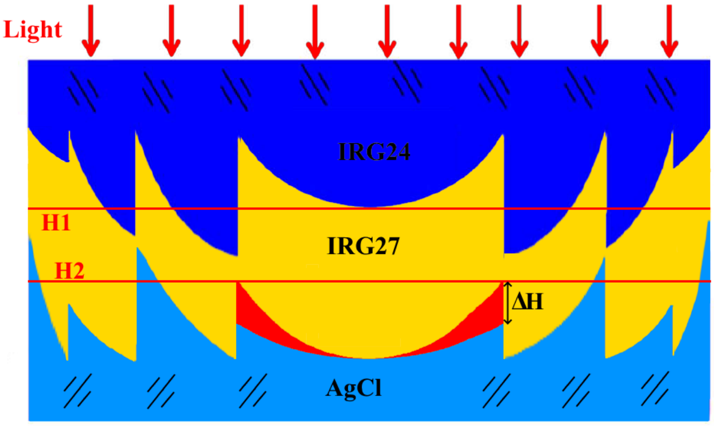


Figure 13.12: MLDOE layout with varying zone thicknesses. Δ_H is different for each zone, based on a normal distribution and a defined tolerance.

.

This process is repeated 1000 times for each tolerance value Δ. Figure 13.13 displays the complete Monte Carlo analysis:

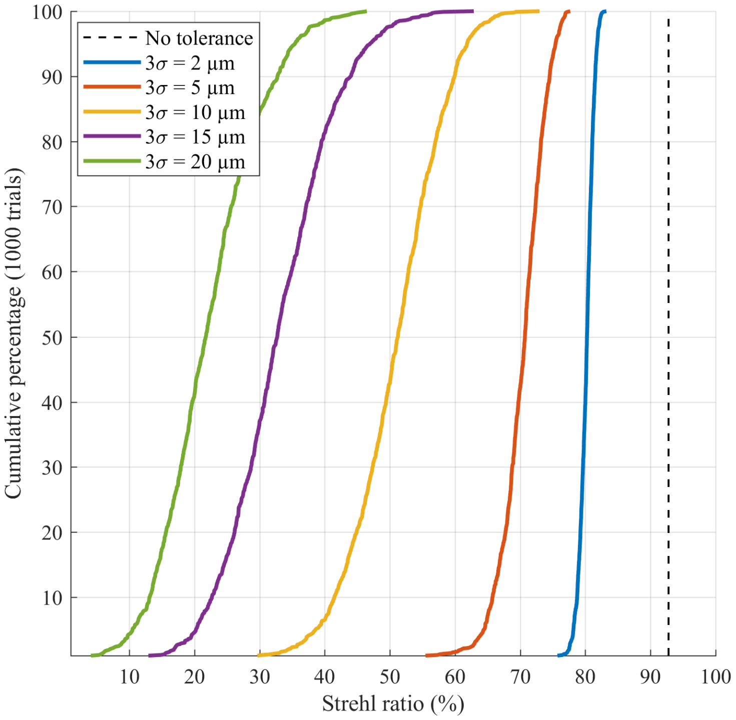


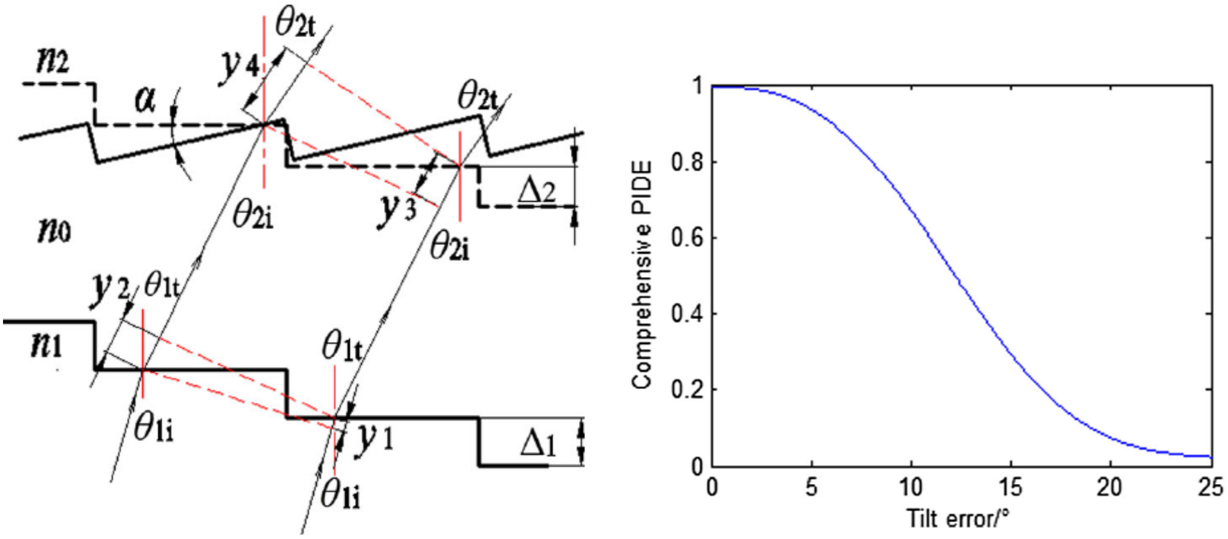
Figure 13.13: Monte Carlo analysis using the the Strehl ratio as an optical evaluation metric. Multiple tolerance levels are studied. The dotted black curve shows the ideal design's Strehl ratio.

Figure 13.13 describes the design performance evolution with the tolerance level. The higher the level and the more "spread out" the performance. For instance, with the lowest tolerance level, $3\sigma = 2$ µm, nearly all 1000 random designs have a Strehl ratio of 80%. Increasing the tolerance level decreases the overall performance and the number of high or low-performance designs. For $3\sigma = 20$ µm, the average performance is 20% Strehl ratio, with 10% of the designs having $> 30\%$ Strehl ratio and 10% having $< 15\%$ Strehl ratio.

This analysis shows the tight thickness tolerance required for this MLDOE design since a 5 µm thickness tolerance already decreases the Strehl ratio by 20% on average. Note that the IRG24 layer has a 255 µm maximal thickness while the AgCl layer has a 115 µm maximal thickness. Consequently, the 5 µm tolerance represents 2% and 4% of the first and second layer thicknesses. Compared to the visible case, where a 10% tolerance ratio was acceptable until 15° incidence, IR MLDOEs seem much more challenging to manufacture.

13.1.4 Tilt error

This section analyses the impact of a tilt error when assembling an MLDOE. It is based on the analytical geometrical model described in [90, 92]. Since the visible waveband and the TEA have been considered, the result provided in this section are indicative, but a rigorous study must be performed in the infrared bandwidth. The modelling layout and the PIDE variation for multiple tilt errors are displayed in Figure 13.14:



(a) Tilt error (α) modeling layout. (b) Comprehensive PIDE for various tilt errors.

Figure 13.14: a) Tilt error (α) modelling layout. y_1, y_2 and y_3, y_4 are the OPL variations when the incident rays pass through the first and second MLDOE surface. b) Comprehensive PIDE analytical analysis. The comprehensive PIDE metric integrates incidence angles between 0° and 20°. The visible waveband is considered. Original source [92].

The analytical OPL variation due to the tilt error α is detailed in [92], using Snell's law of refraction. Figure 13.14 shows that the PIDE keeps a high value ($> 90\%$) for a 6° tilt, which is a very loose tolerance. This result includes off-axis incidences integrated by the comprehensive PIDE metric between 0° and 20°. However, remind that this study is only valid for thin MLDOEs in the visible waveband and should be adapted to the IR waveband.

13.2 Manufacturing processes

To the Author's knowledge, the manufacturing of MLDOEs in the infrared waveband has never been discussed or studied. However, single-point diamond turning (SPDT), micro-lithography and material moulding are the most promising MLDOE manufacturing techniques.

13.2.1 Single point diamond turning (SPDT)

The optimal diamond turning tool shape in the visible waveband is studied in [96]. Based on a shadowing model (effective area method, see Section 13.1.2), they concluded that a half-round SPDT tool was necessary to generate accurate vertical edges. The optimal tool curvature radius and feed rate are described in Equation 13.5:

$$\begin{cases} R_{optim} \approx 0.4T.f \\ f_{min} = T/200 \end{cases} \tag{13.5}$$

where T is the considered zone period. The optimal tool radius R_{optim} is expressed in µm while the minimal feed rate f_{min} is expressed in µm/r (µm/rotation). the feed rate characterizes the displacement speed of the cutting tool when the machined object spins. A representation of the SPDT process is illustrated in Figure 13.15:

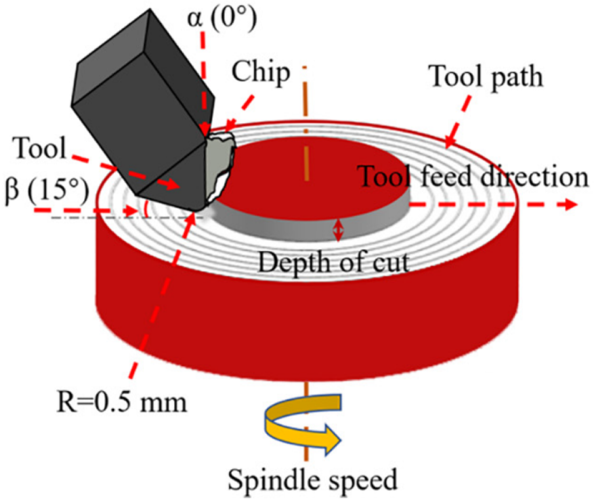


Figure 13.15: A schematic of the ultra-precision SPDT cutting process. The tool angles and radius values are arbitrary, taken from [97].

The variation of the PIDE with the cutting tool radius and feed rate is shown in Figure 13.16:

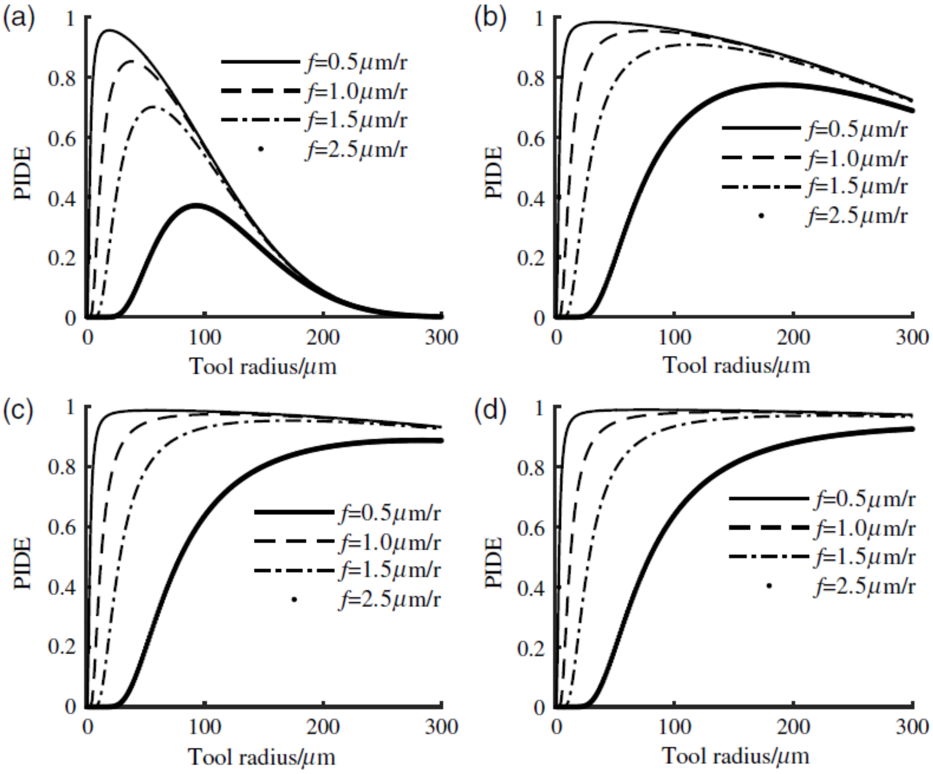


Figure 13.16: The relation between cutting tool radius and PIDE with different feed rates. (a) $T = 100$ µm, (b) $T = 200$ µm, (c) $T = 300$ µm, and (d) $T = 400$ µm.

13.2.2 Lithography and microfluid

A very recent example of MLDOE fabrication in the visible, called microfluidic diffractive optical element (MFDOE), is described in [98]. Though not currently extended in the infrared due to a lack of materials, this process has provided high diffraction efficiency in the visible. In [98], both MFDOE layers have been manufactured using a digital micromirror device (DMD) through maskless lithography (DMDML). The writing process involves a femtosecond laser direct writing (FsLDW). Finally, the gap material is casted inside the MFDOE using a microfluid (ethylene glycol or methanol). The lithographic setup is displayed in Figure 13.17:

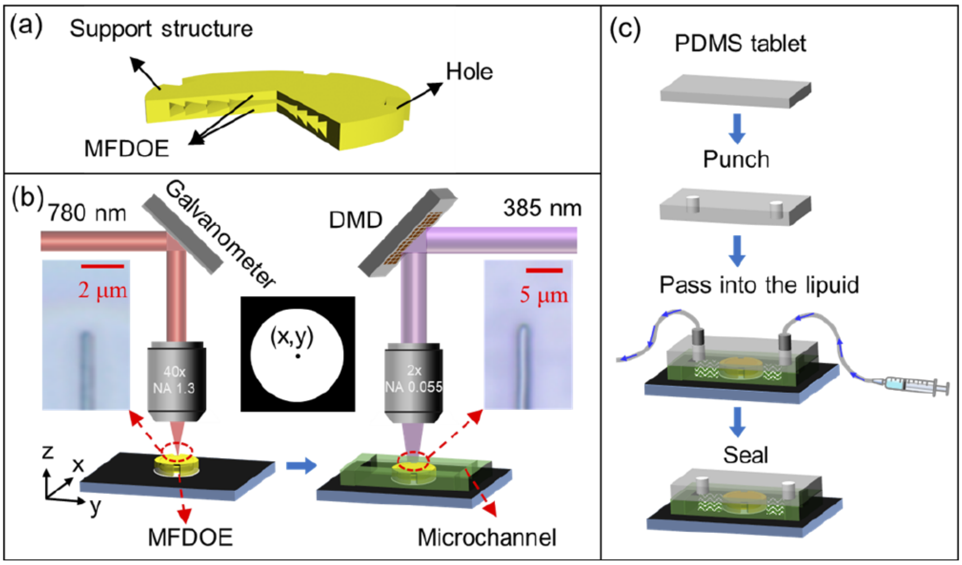


Figure 13.17: (a) Schematic diagram of the three-dimensional structure of MFDOE. (b) Fabrication steps for the MFDOE. (c) The sealing step of MFDOE. Original source [98].

Three-dimensional printing is widely used in micro-optics. The FsLDW allows fabricating both MFDOE layers as shown in Figure 13.17a. An understructure is added to both layers to maintain their integrity. To achieve the alignment of the two processes depicted in Figure 13.17b, a circular mark on the glass substrate is prepared in advance. The MFDOE is processed by FsLDW based on the centre of the circular mark. The microchannel is only used to avoid the volatilisation or flow of liquid after sealing (Figure 13.17c), so high processing precision is not required. The detailed processing is explained in depth in [98].

13.3 Chalcogenide moulding

To the Author's knowledge, no study has been made yet concerning the feasibility of infrared moulded MLDOEs. Nonetheless, Chaclcogenides have been widely used for high accuracy microlens fabrication due to their moulding characteristics [99]. The moulding process is depicted in Figure 13.18:

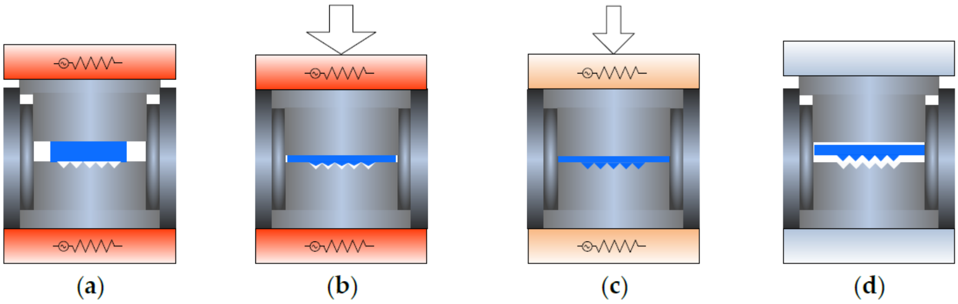
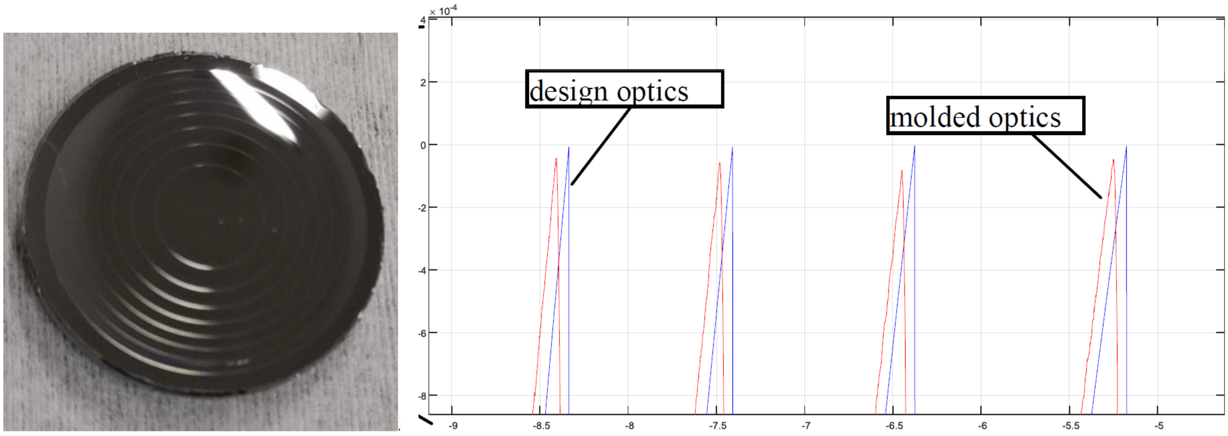


Figure 13.18: Four stages of a PGM (precision glass moulding) cycle: (a) heating, (b) pressing, (c) annealing, and (d) cooling. Original source [99]

SPDT is often used for engraving the microstructure pattern on the mould. A recent example of diffractive optics moulding using chalcogenide glass is shown in the following, based on [100]. They follow a moulding process similar to Figure 13.18 detailed in [100]. The result is depicted in Figure 13.19:



(a) Microphoto of the moulded DOE.

(b) Comparison between moulded parts and the designed DOE.

Figure 13.19: a) Micro-photo of the moulded DOE and b) Comparison between moulded parts and the designed DOE. Original source [100].

As shown in Figure 13.19 and according to [100], the local filling of the microstructure is accurate. The annealing rate significantly influences the surface profile, followed by the annealing and temperature and then by the holding pressure. Flat diffraction moulds processed with diamond tools can achieve even better filling of microstructures.

Chapter 14

Conclusion and perspectives

14.1 Overall conclusion on the MLDOEs conception

This thesis has treated MLDOE conception in the context of Earth infrared remote sensing. We have described multiple infrared applications in chapter 1, explaining their spatial, temporal, thermal and spectral requirements. Three prominent applications have been detailed, with a common dual-band infrared bandwidth requirement. The remote sensing of a field's canopy temperature, combined with the soil temperature knowledge, allows for estimating a crop's water stress. Controlled water stress irrigation is an innovative and sustainable water management strategy that requires precise plant water stress control and irrigation schedules. Infrared remote sensing greatly serves fire detection in distinguishing actual fires from false alarms. The study of fire radiative energy and burned areas provide insight into fire spreading mechanisms and improve fire prevention. A similar idea applies to volcano monitoring, where infrared images significantly add value in determining lava flow and eruption characteristics. A dual-band 10 m resolution camera enables water management and has sufficient resolution for "hot events" monitoring in case of fire or eruption occurring in its field of view. A spacecraft constellation ideally achieves daily coverage.

The targeted applications provide high-level requirements such as spectral and spatial resolution. Nonetheless, they are not sufficient to entirely constrain an optical design. In Chapter 2, we have reviewed some state-of-the-art infrared detection technologies. We have discussed the prominent detector noises and figures of merit, providing additional constraints to the optical design. The high-level requirements and detection constraints have been integrated into the signal-to-noise ratio study in Chapter 3. Starting from the source physical radiance model, we have detailed a comprehensive parametric study, expressing the NEDT and SNR as a function of fundamental optical design parameters (pixel size, F-number, and aperture diameter). Assuming some detector characteristics and the crop water management application in LEO orbit, we have selected a suitable optical F-number and pixel size. The resulting MWIR and LWIR SNRs for this low-temperature application are 1 and 25, associated with a 12 m GSD in MWIR and a 24 m GSD in LWIR. When high temperatures applications are considered, the MWIR SNR increases to 30.

Chapters 4, 5, and 6 are aside chapters since they each introduced a different modelling tool to study complex diffractive optics. In Chapter 4, we developed a free-space optical wave propagator in Matlab, following the work of Goodman and Matsushima on the angular spectrum method. This method has proven exceptionally useful in later chapters, generating optical metrics at the focal plane. Since Fourier optics already provided free-space field propagation, Chapters 5 and 6 focused on modelling the field behaviour inside an MLDOE. Chapter 5 described an analytical optical path length calculator (ray-tracing) specifically designed for studying MLDOEs and implemented in Matlab. Much faster than professional ray tracing software, the Matlab analytical optical path length calculator is perfectly accurate and fully parameterisable. Finally, Chapter 6 details an exact electromagnetic field calculation method known as FDTD. This chapter has explained the MLDOE layout creation in FDTD and how to adequately define spatial and temporal sampling, ensuring trustful simulations. This study allowed for considering FDTD simulations as references in the following chapters.

Chapter 7 has proposed to optimise a MWIR-only design using a standard diffractive lens and a powerful athermalisation algorithm. The latter has been detailed, based on various literature references, and applied step-by-step to the design of a MWIR optical system. It has been implemented on the professional ray-tracing software CodeV. We have been able to extend the MWIR design thermal range from ±5°K to ±50°K. This chapter constitutes a starting point that needs to be extended in LWIR, justifying the research of a dual-band diffractive component in the following chapters.

Chapter 8 has introduced the concept of diffractive optics and explained their advantages in compact, achromat optical systems. The constructive interference requirement between each diffractive zone, associated with Fermat's principle, is the cornerstone of designing diffractive elements. The thin element approximation has been presented as an analytical solution to compute the phase delay induced by a DOE. This chapter has shown multiple DOE designs, varying the operating order and the harmonic parameter. It resulted in standard DOEs having a high diffraction efficiency only at their design wavelength, making them incompatible with dual-band designs. Multi-order DOEs could go around this issue but suffer from a discontinuous chromatic focal shift, hardly suitable for hybrid refractive-diffractive designs.

Multilayer diffractive optical elements (MLDOEs) have been introduced in Chapter 9 as dual-band extensions of standard DOEs. The designing principles remain the same as for DOEs, except that a second design wavelength is considered. This is permitted by increasing the number of diffractive layers (and materials). By design, an MLDOE ideally provides maximal diffraction efficiency at each design wavelength (one in MWIR, one in LWIR), making it perfectly suitable for dual-band applications. Its chromatic focal shift has the standard negative variation, making the coupling with refractive lenses very beneficial for achromatisation. This ideal MLDOE behaviour relies on the validity of the thin element approximation. While it is always verified for standard DOEs, we have shown, using some literature references, that it was no more the case for MLDOEs. Another modelling technique must be applied to accurately assess MLDOE performance.

The ray model has been proposed as an alternative MLDOE modelling strategy in Chapter 10. It uses ray-tracing to compute optical path lengths inside an MLDOE instead of relying on the TEA. Our Matlab OPL calculator has been specifically used to derive an MLDOE phase approximation. Fourier optics remains a powerful free-space propagation method and has constituted the ideal tool to retrieve the "best" focal plane, the PSF, and the Strehl ratio. This chapter has compared the performance results of the ray model and the TEA. The former provides pessimistic results, while the second, by design, always predicts ideal performance. FDTD simulations have been conducted to adjudicate the issue, demonstrating a clear-cut validity of the ray model over the TEA model. Nonetheless, the ray model Strehl ratio predictions have not proven as accurate as FDTD ones.

The following chapters have presented two optimisation methods. The *geometric optics material selection method* presented in Chapter 11 has allowed for selecting the most suitable MLDOE material combinations. This method uses the total internal reflection, the transmission and the total thickness as evaluation metrics to determine if a material combination has imaging potential. As a result, three-layer DOEs have displayed the highest performance upon the studied configurations, but in many cases, they could not outclass standard DOEs. Chapter 12 has proposed to optimise the diffractive zone height using the ray model. To this end, the extended scalar theory has provided extremely valuable zone design equations. A parametric study involving the ray model has been conducted to determine the best EST parametrisation for various MLDOE configurations. As a result, our optimisation algorithm has improved the imaging capabilities of each MLDOE design, but not to the expected extent (when compared to FDTD). Still, the optimised IRG24-IRG27-AgCl MLDOE design has proven to be especially performing in the MWIR and LWIR wavebands.

The IRG24-IRG27-AgCl MLDOE has the highest polychromatic Strehl ratio of all the tested configurations. Nonetheless, this predicted performance, supported by exact FDTD simulations, assumes an ideal MLDOE. Even the finest optical fabrication process brings manufacturing errors, and the optical assembly is never perfect. Chapter 13 has proposed understanding the effect of such defaults on the IRG24-IR27-AgCl MLDOE performance. The impact of a layer decenter, an addition of passive facets, and a variation of thickness has been evaluated using FDTD, simulating a non-ideal MLDOE. The IRG24-IRG27-AgCl MLDOE is very sensitive to thickness errors. Nevertheless, loose tolerances can be applied to decenters or period width errors. A literature review has highlighted possible modelling strategies for decenters, tilt, and period width errors in the visible waveband. Finally, prominent micro-optics manufacturing processes (diamond turning, lithography and moulding) have been evocated for completeness.

14.2 Suggested improvements and perspectives

So far, this thesis has focused on accurately retrieving an MLDOE's optical performance and determining promising designs. We have described at least one very performing dual-band MLDOE and started studying the impact of various manufacturing/assembly errors. To the Author's knowledge, the ray model and Fourier optics have been pushed to their limits for the modelling of MLDOEs, with mitigated success. The hybrid combinations of these two methods have improved the polychromatic Strehl ratio of most of the studied MLDOEs, but often to an insufficient extent. Furthermore, the approximate ray model does not ensure finding the most optimal design and only provides a design starting point. Only rigorous approaches such as the FDTD method are accurate enough to deliver trustful results in the frame of thick diffractive optical elements. We suggest the creation of a specific MLDOE design software implementing the following procedure:

- Eliminate un-manufacturable or too expensive material combinations.
- For each remaining material combination, use the ray model to optimise the layer shape. It will serve as a starting point for the subsequent FDTD analysis.
- Use the geometric material selection method to eliminate "improper" designs (i.e. designs having low transmission or high thickness/TIR). We assume that these improper solutions barely have any chances to be improved after an FDTD optimisation.
- Conceive a core FDTD calculator to optimise further the MLDOE shape. A genetic algorithm can be employed to find an optimal solution as efficiently as possible.
- For the most performing designs, further FDTD tolerancing studies must be carried out to determine fine manufacturing requirements.
- A monte-Carlo analysis using the FDTD core is mandatory for precise cross-tolerance evaluation. It provides a manufacturability metric and a price estimation.

This procedure allows for designing MLDOE comprehensively and reliably. It necessitates a fine MLDOE manufacturing knowledge that may not currently exist. An FDTD calculation core is required for automation and fast parametrisation, as it constitutes the cornerstone of the optimisation process.

A further step would be to include MLDOEs and refractive lenses in a unique design software to estimate their combined performance and correct their aberrations. This step would consists in extending Chapter 7 camera to LWIR using MLDOEs. Since ray tracing cannot be sufficient to model MLDOE, at least to the Author's knowledge, this designing software should apply some Field-Tracing paradigms. It must propagate an input field through various optical components, selecting adequate modellings: Fourier optics in free space, FDTD inside an MLDOE, and ray-tracing for lenses.

In this thesis, we only applied the athermalization algorithm to a hybrid MWIR design without MLDOEs. Similarly to tolerances, an MLDOE's thermal behaviour is discussed in the literature but should be deepened using an environmental tool similar to CodeV's *ENV* macro. Dual-band athermalization is discussed in some papers and is a keystone of hybrid MLDOE design for dual-band imaging.

Future perspectives

Infrared remote sensing offers a vast range of applications, mostly addressing make-or-break issues of our century. Most applications, from wildfire detection to crop irrigation management, tremendously profit from multispectral (or even dual-band) imaging. In the meantime, infrared applications require increasing temporal and spatial resolution, coinciding with the current image processing capabilities, physical phenomena knowledge, and monitoring strategies. Recent detection technologies have enabled small pixels, increasing the achievable infrared spatial resolution. Applications that necessitate daily revisit traditionally rely on geostationary satellites but can now be fulfiled by high-resolution nanoSat constellations, enabling high temporal resolution.

In this context, having compact, passive dual-band optical designs is a crucial technologic milestone. NanoSat or even SmallSat constellations cannot board bulky multispectral imagers. The latter are often constituted by light-splitting devices and multiple optical arms (each for a specific waveband). Furthermore, they require space and cannot be manufactured at high cadency. In contrast, MLDOEs achieve both compactness and dual-band capabilities, displaying considerable potential for infrared imaging. In addition, the rising interest in MLDOEs might trigger further development in dual-band detectors or innovative manufacturing processes.

Moulding processes may enable the production of performing dual-band cameras in large quantities, considerably reducing their cost. In parallel, it is worth mentioning the development of dual-band detection technologies that will gain a huge added value if dual-band components such as MLDOEs and their manufacturing processes become standardised. This statement leads to the conclusion that a virtuous cycle may arise from the miniaturisation of multispectral infrared systems.

Chapter 15

Bibliography

[1] J. A. Sobrino, F. Del Frate, M. Drusch, J. C. Jimenez Munoz, and P. Manunta, "Review of High Resolution Thermal Infrared Applications and Requirements: The Fuegosat Synthesis Study," in *Thermal Infrared Remote Sensing. Sensors, Methods, Applications.*, vol. 17, ch. 10, Springer, 2013.

[2] M. Gerhards, *Advanced Thermal Remote Sensing for Water Stress Detection of Agricultural Crops*. PhD thesis, Universität Trier, 2 2018.

[3] P. J. Pinter Jr, J. L. Hatfield, J. S. Schepers, E. M. Barnes, M. S. Moran, C. S. Daughtry, and D. R. Upchurch, "Remote sensing for crop management," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 6, pp. 647–664, 2003.

[4] P. Colaizzi, S. Evett, S. O'Shaughnessy, and T. Howell, "Using plant canopy temperature to improve irrigated crop management," *Proceedings of the 24th Annual Central Plains Irrigation Conference Colby*, 02 2012.

[5] N. Liu, Z. Deng, H. Wang, Z. Luo, H. A. Gutiérrez-Jurado, X. He, and H. Guan, "Thermal remote sensing of plant water stress in natural ecosystems," *Forest Ecology and Management*, vol. 476, p. 118433, 2020.

[6] S. Taghvaeian, J. L. Chávez, J. Altenhofen, T. Trout, and K. DeJonge, "Remote sensing for evaluating crop water stress at field scale using infrared thermography: potential and limitations," *Hydrology Days 2013*, 2013.

[7] R. G. Allen, L. S. Pereira, D. Raes, M. Smith, *et al.*, "Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56," *Fao, Rome*, vol. 300, no. 9, p. D05109, 1998.

[8] D. M. Szpakowski and J. L. R. Jensen, "A review of the applications of remote sensing in fire ecology," *Remote Sensing*, vol. 11, no. 22, 2019.

[9] M. J. Wooster, G. Roberts, *et al.*, "Thermal Remote Sensing of Active Vegetation Fires and Biomass Burning Events," in *Thermal Infrared Remote Sensing. Sensors, Methods, Applications.*, vol. 17, ch. 18, Springer, 2013.

[10] A. Calle and J. L. Casanova, "Forest fires and remote sensing," in *Integration of Information for Environmental Security* (H. G. Coskun, H. K. Cigizoglu, and M. D. Maktav, eds.), (Dordrecht), pp. 247–290, Springer Netherlands, 2008.

[11] M. Fabrizia Buongiorno, D. Pieri, and M. Silvestri, “Thermal Analysis of Volcanoes Based on 10 Years of ASTER Data on Mt. Etna,” in *Thermal Infrared Remote Sensing. Sensors, Methods, Applications.*, vol. 17, ch. 20, Springer, 2013.

[12] L. Spampinato, S. Calvari, C. Oppenheimer, and E. Boschi, “Volcano surveillance using infrared cameras,” *Earth-Science Reviews*, vol. 106, no. 1, pp. 63–91, 2011.

[13] M. Patrick, A. Harris, M. Ripepe, J. Dehn, D. Rothery, and S. Calvari, “Strombolian explosive styles and source conditions: Insights from thermal (flir) video,” *Bulletin of Volcanology*, vol. 69, p. 769–784, 05 2007.

[14] A. Rogalski and K. Chrzanowski, “Infrared devices and techniques (revision),” *Metrology and Measurement Systems*, vol. 21, 12 2014.

[15] O. Riou and L. Audaire, “Technologie des dÉtecteurs des camÉras thermiques,” tech. rep., Centre d‘études et de recherche de thermique et d‘environnement des systèmes, IUT de Sénart, 2001.

[16] N. K. Dhar, R. Dat, and A. K. Sood, “Advances in infrared detector array technology,” *Optoelectronics-Advanced Materials and Devices*, vol. 1, 2013.

[17] Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*, vol. 1. Springer, 2014.

[18] A. Daniels, *Field Guide to Infrared Systems, Detectors, and FPAs, Third Edition*. SPIE, 2018.

[19] SCD, “Pelican-d 640.” `https://www.scd.co.il/products/pelican-d-640/`. Accessed: 2022-04-15.

[20] A. Jain and A. Banerjee, “Design, development, characterization and qualification of infrared focal plane area array detectors for space-borne imaging applications,” in *Earth Observing Missions and Sensors: Development, Implementation, and Characterization IV* (X. J. Xiong, S. A. Kuriakose, and T. Kimura, eds.), vol. 9881, pp. 205 – 213, International Society for Optics and Photonics, SPIE, 2016.

[21] H. Yu, *Time-delay-integration CMOS image sensor design for space applications*. PhD thesis, Nanyang Technological University, 2016.

[22] J. Vilaboa Perez, “Multi-band ir sensor for earth observation,” Master’s thesis, Liege University, 2019. https://matheo.uliege.be/handle/2268.2/6781.

[23] A. Rogalski, *Infrared and terahertz detectors*. CRC Press, 2019.

[24] J. M. Lloyd, *Performance Summary Measures*, pp. 166–211. Boston, MA: Springer US, 1975.

[25] Leonardo, “Condor ii - dwir.” `https://electronics.leonardo.com/en/products/condor-ii`. Accessed: 2022-04-15.

[26] A. Rogalski, J. Antoszewski, and L. Faraone, “Third-generation infrared photodetector arrays,” *Journal of Applied Physics*, vol. 105, 2009.

[27] R. L. Goff, F. Tanguy, P. Fuss, and P. Etcheto, “Technological development of multispectral filter assemblies for micro bolometer,” in *International Conference on Space Optics — ICSO 2012* (B. Cugny, E. Armandillo, and N. Karafolas, eds.),

vol. 10564, pp. 212 – 216, International Society for Optics and Photonics, SPIE, 2017.

[28] J. M. Lloyd, *Thermal Radiation Theory and Atmospheric Transmission*, pp. 18–67. Boston, MA: Springer US, 1975.

[29] A. Dueñas Herrero, “Radiometric budget,” in *Study of a Cubesat dual band infrared imager for hydric stress observation from space*, ch. 4, The University of Liege, Belgium, 2018.

[30] G. C. Holst and R. G. Driggers, “Small detectors in infrared system design,” *Optical Engineering*, vol. 51, no. 9, pp. 1 – 10, 2012.

[31] G. C. Holst, “Imaging system performance based upon FL/d,” *Optical Engineering*, vol. 46, no. 10, pp. 1 – 10, 2007.

[32] D. Lohrmann, R. T. Littleton, C. E. Reese, D. Murphy, and J. N. Vizgaitis, “Uncooled long-wave infrared small pixel focal plane array and system challenges,” *Optical Engineering*, vol. 52, no. 6, pp. 1 – 7, 2013.

[33] G. C. Holst, “Are reconstruction filters necessary?,” in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII* (G. C. Holst, ed.), vol. 6207, pp. 219 – 229, International Society for Optics and Photonics, SPIE, 2006.

[34] J. W. Goodman, “The angular spectrum of plane waves,” in *Introduction to Fourier Optics*, ch. 3, McGraw-Hill, 1996.

[35] D. Voelz, “Scalar Diffraction and Propagation Solutions,” in *Computational Fourier Optics, a Matlab tutorial*, ch. 4, SPIE press, 2011.

[36] K. Matsushima and T. Shimobaba, “Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields,” *Optics Express*, vol. 12, no. 22, 2009.

[37] Wikipedia, “Rectangular function.” `https://en.wikipedia.org/wiki/Rectangular_function`. Accessed: 2022-23-01.

[38] K. Matsushima, “Shifted angular spectrum method for off-axis numerical propagation,” *Optics Express*, vol. 18, no. 17, 2010.

[39] H. Sauer, P. Chavel, and G. Erdei, “Diffractive optical elements in hybrid lenses: modeling and design by zone decomposition,” *Appl. Opt.*, vol. 38, pp. 6482–6486, Nov 1999.

[40] A. Nemes-Czopf, D. Bercsényi, and G. Erdei, “Simulation of relief-type diffractive lenses in zemax using parametric modelling and scalar diffraction,” *Appl. Opt.*, vol. 58, pp. 8931–8942, Nov 2019.

[41] Wikipedia, “Aspheric lens.” `https://en.wikipedia.org/wiki/Aspheric_lens`. Accessed: 2022-09-02.

[42] J. J. Rousseau, “Expression vectorielle des lois de descartes.” `http://ressources.univ-lemans.fr/AccesLibre/UM/Pedago/physique/02/optigeo/descartesvecto.html`. Accessed: 2020-11-11.

[43] O. photonic software, “Optifdtd.” `https://optiwave.com/optifdtd-overview/`. Accessed: 2022-10-02.

[44] Optiwave, “Optifdtd documentation – technical background,” tech. rep., Optiwave Systems Inc, Ontario, Canada, 2020.

[45] M. J.Riedl, “Thermal Effects,” in *Optical Design Fundamentals for Infrared Systems, Second edition* (I. R. Inc. and U. of Central Florida, eds.), ch. 7, SPIE PRESS, 2001.

[46] T. H. Jamieson, “Thermal effects in optical systems,” *Optical Engineering 20*, 1981.

[47] Y. Tamagawa and T. Tajime, “Expansion of an athermal chart into a multilens system with thick lenses spaced apart,” *Optical Engineering, Vol. 35 No. 10*, 1996.

[48] Y. Tamagawa and T. Tajime, “Multilens system design with an athermal chart,” *Applied Optics Vol.33 N. 34*, 1994.

[49] B.-I. Ahn, Y.-S. Kim, and S.-C. Park, “Athermal and Achromatic Design for a Night Vision Camera Using Tolerable Housing Boundary on an Expanded Glass Map,” *Current Optics and Photonics Vol.1 N. 2*, April 2017.

[50] V. Moreno, J. R. Salgueiro, and J. F. Román, “High efficiency diffractive lenses: Deduction of kinoform profile,” *American Journal of Physics*, 1997.

[51] I. Bertrand, “State of the art : optical materials,” in *Space validation of infrared materials for multispectral imager on board a Cubesat mission*, ch. 4, The University of Liege, Belgium, 2018.

[52] CRYSTRAN, “Crystran.” `https://www.crystran.co.uk/optical-materials`. Accessed: 2019-28-11.

[53] SCHOTT, “Reliable solutions for the infrared industry.” `https://www.schott.com/advanced_optics/english/products/optical-materials/ir-materials/infrared-chalcogenide-glasses/index.html`. Accessed: 2020-24-02.

[54] J. L. Tosi and K. M. Khajurivala, “Common infrared optical materials and coatings: A guide to properties, performance and applications.” `https://www.photonics.com/a25495/`. Accessed: 2019-28-11.

[55] J. T. Daiker, “Athermalization techniques in infrared systems,” *Opti521*, 2010.

[56] P. . J. Rogers and M. Roberts, “Thermal Compensation Techniques,” in *Handbook of Optics, Second edition, Vol I* (U. o. C. F. CREOL, ed.), ch. 39, OSA, 1995.

[57] J.-P. Goure, *L’optique dans les instruments: Généralités*. Lavoisier, 2011.

[58] T.-Y. Lim, Y.-S. Kim, and S.-C. Park, “Graphical selection of optical materials using an expanded athermal glass map and considering the housing material for an athermal an achromatic design,” *Journal of the Optical society of Korea Vol.19 N. 5*, October 2015.

[59] T.-Y. Lim and S.-C. Park, “Achromatic and athermal lens design by redistributing the element powers on an athermal glass map,” *Optics Express*, 2016.

[60] N. Xie, Q. Cui, and J. Wang, “Optical athermalization in the visible waveband using the 1+σ method,” *Applied Optics Vol.58 N. 3*, January 2019.

[61] D. A. Pommet, M. Moharam, and E. B. Grann, “Limits of scalar diffraction theory for diffractive phase elements,” *JOSA*, vol. 11, no. 6, 1994.

[62] F. Languy, "Design of broadband high-efficiency diffractive lenses," in *Achromatization of nonimaging Fresnel lenses for photovoltaic solar concentration using refractive and diffractive patterns*, ch. 4, University of Liege, 2012.

[63] D. A. Buralli and M. G. Morris, "Effect of diffraction efficiency on the modulation transfer function of diffractive lenses," *Applied Optics*, vol. 31, no. 22, 1992.

[64] C. W. Sweeney and G. E. Sommargren, "Harmonic diffractive lenses," *Applied Optics*, vol. 34, pp. 2462–2475, 1995.

[65] D. Faklis and G. M. Morris, "Spectral properties of multiorder diffractive lenses," *Applied Optics*, vol. 34, pp. 2462–2468, 1995.

[66] V. Laborde, J. Loicq, and S. Habraken, "Modeling infrared behavior of multilayer diffractive optical elements using fourier optics," *Appl. Opt.*, vol. 60, pp. 2037–2045, Mar 2021.

[67] V. Laborde, J. Loicq, J. Hastanin, and S. Habraken, "Hybrid ray-tracing/fourier optics method to analyze multilayer diffractive optical elements," *Appl. Opt.*, vol. 61, pp. 4956–4966, Jun 2022.

[68] Y. Arieli, "Kinoforms for wideband use," *SPIE*, 1995.

[69] Y. Arieli, S. Noach, S. Ozeri, and N. Eisenberg, "Design of diffractive optical elements for multiple wavelengths," *Applied Optics*, vol. 37, no. 26, 1998.

[70] B. Zhang, Q. Cui, and M. Piao, "Effect of substrate material selection on polychromatic integral diffraction efficiency for multilayer diffractive optics in oblique incident situation," *Optics Communications*, vol. 415, pp. 156–163, 2018.

[71] C. Xue, Q. Cui, T. Liu, Y. Liangliang, and B. Fei, "Optimal design of multilayer diffractive optical element for dual wavebands," *Optics Letters*, vol. 35, no. 24, 2010.

[72] D. A. Buralli and G. M. Morris, "Effects of diffraction efficiency on the modulation transfer function of diffractive lenses.," *Applied optics*, vol. 31 22, pp. 4389–96, 1992.

[73] Y. Soskind, "Diffractive optics technologies in infrared systems," *Proc. of SPIE*, vol. 9451, no. 94511T-1, 2015.

[74] N. Takehiko and O. Hideki, "Research on multi-layer diffractive optical elements and their application to camera lenses," *DOMO,DMA2*, 2002.

[75] C. Fan, Z.Wang, L.Lin, M. Zhang, and H. Fan, "Design of infrared telephoto-optical system with double layer harmonic diffractive element," *Chinese Physics Letters*, vol. 24, 2007.

[76] B. Zhang, Q. Cui, M. Piao, and Y. Hu, "Design of dual band infrared zoom lens with multilayer diffractive optical element," *Applied Optics*, vol. 58, no. 8, 2019.

[77] G. Greisukh, G. Danilov, E. Ezhov, S. Stepanov, and B. Usievich, "Comparison of electromagnetic and scalar methods for evaluation of efficiency of diffractive lenses for wide spectral bandwidth," *Optics Communication*, vol. 338, pp. 54–57, 2015.

[78] H. Zhong, S. Zhang, F. Wyrowski, and H. Schweitzer, "Parabasal thin element approximation for the analysis of the diffractive optical elements," *Proc. of SPIE*, vol. 9131, no. 913114-2, 2014.

[79] G. Greisukh, G. Danilov, S. Stepanov, A. Antonov, and B. Usievich, “Spectral and angular dependences of the efficiency of three-layer relief-phase diffraction elements of the ir range,” *Optics and Spectroscopie*, vol. 125, no. 1, 2018.

[80] G. Greisukh, G. Danilov, S. Stepanov, A. Antonov, and B. Usievich, “Minimization of the total depth of internal saw-tooth reliefs of a two-layer relief-phase diffraction microstructure,” *Optics and Spectroscopie*, vol. 124, no. 1, 2018.

[81] H. Yang, C. Xue, C. Li, and J. Wang, “Optimal design of multilayer diffractive optical elements with effective area method,” *Applied Optics*, vol. 55, no. 7, 2016.

[82] C. Yang, H. Yang, C. Li, and C. Xue, “Optimization and analysis of infrared multilayer diffractive optical elements with finite feature sizes,” *Applied Optics*, vol. 58, no. 10, 2019.

[83] F. Wyrowski and M. Kuhn, “Introduction to field tracing,” *Journal of Modern Optics*, vol. 58, no. 5-6, pp. 449–466, 2010.

[84] B. R. Org, “Asap nextgen.” `https://breault.com/asap/`. Accessed: 2022-04-22.

[85] H. Yang, C. Xue, C. Li, J. Wang, and R. Zhang, “Diffraction efficiency sensitivity to oblique incident angle for multilayer diffractive optical elements,” *Appl. Opt.*, vol. 55, pp. 7126–7133, Sep 2016.

[86] V. Laborde, J. Loicq, J. Hastanin, and S. Habraken, “Multilayer diffractive optical element material selection method basedon transmission, total internal reflection, and thickness,” *Appl. Opt.*, vol. 61, pp. 7415–7423, Sep 2022.

[87] T. Wang, H. Liu, H. Zhang, H. Zhang, Q. Sun, and Z. Lu, “Effect of incidence angles and manufacturing errors on the imaging performance of hybrid systems,” *Journal of Optics*, vol. 13, p. 035711, feb 2011.

[88] F. Huo, W. Wang, and C. Xue, “Limits of scalar diffraction theory for multilayer diffractive optical elements,” *Optik*, vol. 127, 2016.

[89] G. Swanson, “Binary optics technology: Theoretical limits on the diffraction efficiency of multilevel diffractive optical elements,” tech. rep., Lincoln Laboratory, Massachusetts Institute of Technology, 1991.

[90] L. Gao, S. To, H. Yang, X. Nie, T. Liu, and C. Xue, “Effect of assembling errors on the diffraction efficiency for multilayer diffractive optical elements,” *Appl. Opt.*, vol. 53, pp. 7341–7347, Nov 2014.

[91] S. Mao and J. Zhao, “Tolerance analysis on decenter error of multilayer diffractive optical elements based on polychromatic integral diffraction efficiency,” *Appl. Opt.*, vol. 58, pp. 2422–2428, Mar 2019.

[92] L. Yang, C. Liu, R. Guo, and Y. Zhao, “Comprehensive polychromatic integral diffraction efficiency sensitivity to tilt error for multilayer diffractive optical elements with oblique incidence,” *Appl. Opt.*, vol. 59, pp. 508–514, Jan 2020.

[93] H. Yang and C. Xue, “Influence of passive facet of multilayer diffractive optical elements,” *Appl. Opt.*, vol. 57, pp. 2604–2609, Apr 2018.

[94] H. Yang and C. Xue, “Sensitivity of diffraction efficiency to period width errors for multilayer diffractive optical elements,” *Appl. Opt.*, vol. 57, pp. 855–860, Feb 2018.

[95] L. Yang, C. Liu, and S. Li, "Optimal design of depth-scaling error for multilayer diffractive optical elements with oblique incidence," *Appl. Opt.*, vol. 56, pp. 4532–4536, May 2017.

[96] P. Zhou, C. Xue, C. Yang, C. Liu, and X. Liu, "Diffraction efficiency evaluation for diamond turning of harmonic diffractive optical elements," *Appl. Opt.*, vol. 59, pp. 1537–1544, Feb 2020.

[97] H. Zhiyuan, Y. Dai, Z. Li, G. Zhang, C. Chang, and J. Ma, "Investigation on surface morphology and crystalline phase deformation of al80li5mg5zn5cu5 high-entropy alloy by ultra-precision cutting," *Materials and Design*, vol. 186, p. 108367, 11 2019.

[98] M. Tan, L. Huang, J. Cao, H. Zhang, S. Zhao, M. Liu, Z. Jia, R. Zhai, and H. Liu, "Microflow multi-layer diffraction optical element processed by hybrid manufacturing technology," *Opt. Express*, vol. 30, pp. 24689–24702, Jul 2022.

[99] T. Zhou, Z. Zhu, X. Liu, Z. Liang, and X. Wang, "A review of the precision glass molding of chalcogenide glass (chg) for infrared optics," *Micromachines*, vol. 9, no. 7, 2018.

[100] Y. Zhang, Z. Wang, F. Zhang, J. Li, and Z. Zhang, "Research on molding of chalcogenide glass diffractive optical element," in *Seventh Asia Pacific Conference on Optics Manufacture and 2021 International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021)* (J. Tan, X. Luo, M. Huang, L. Kong, and D. Zhang, eds.), vol. 12166, p. 1216651, International Society for Optics and Photonics, SPIE, 2022.

[101] U. of reading, "Germanium (ge)." `https://www.reading.ac.uk/ir-infraredmaterials-ge.aspx`. Accessed: 2021-19-05.

[102] H. H. Li, "Refractive index of zns, znse, and znte and its wavelength and temperature derivatives," *J. Phys. Chem.*, vol. 13, no. 1, pp. 121,135, 1984.

[103] T. Skauli, "Improved dispersion relations for gaas and applications to nonlinear optics," *Journal of Applied Physics*, vol. 94, no. 10, 2003.

[104] G. J. Hawkins, "Infrared materials absorption theory," in *Spectral Characterisation of Infrared Optical Materials and Filters*, ch. 1, pp. 20,27, The University of Reading, UK, 1998.

[105] R. Index.info, "Refractive index.info." `https://refractiveindex.info/?shelf=main&book=AgCl&page=Tilton`. Accessed: 2021-05-17.

[106] CRYSTRAN, "Crystran." `https://www.crystran.co.uk/optical-materials/silver-chloride-agcl`. Accessed: 2021-05-17.

Chapter 16

Appendix A: IR materials

16.1 Germanium Ge

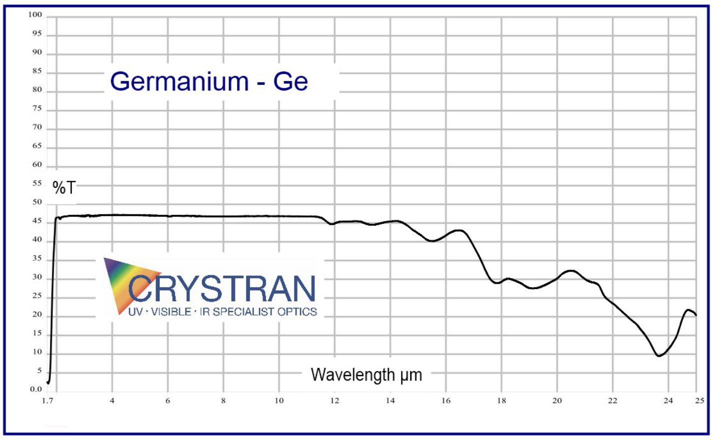


Figure 16.1: IR transmission of uncoated Ge.

Ge refractive index model comes from [101]. It is valid from 2.5 µm to 14 µm and from 50°K to 300°K.

$$\begin{cases} n^2(\lambda, T) = A + \dfrac{B\lambda^2}{\lambda^2 - C} + \dfrac{D\lambda^2}{\lambda^2 - E} \\ A = -6.040 * 10^{-3}T + 11.05128 \\ B = 9.285 * 10^{-3}T + 4.00536 \\ C = -5.392 * 10^{-4}T + 0.599034 \\ D = 4.151 * 10^{-4}T + 0.09145 \\ E = 1.51408T + 3426.5 \end{cases} \tag{16.1}$$

Ge thermal and chromatic powers are displayed in Figure 16.2 for both MWIR and LWIR bands:

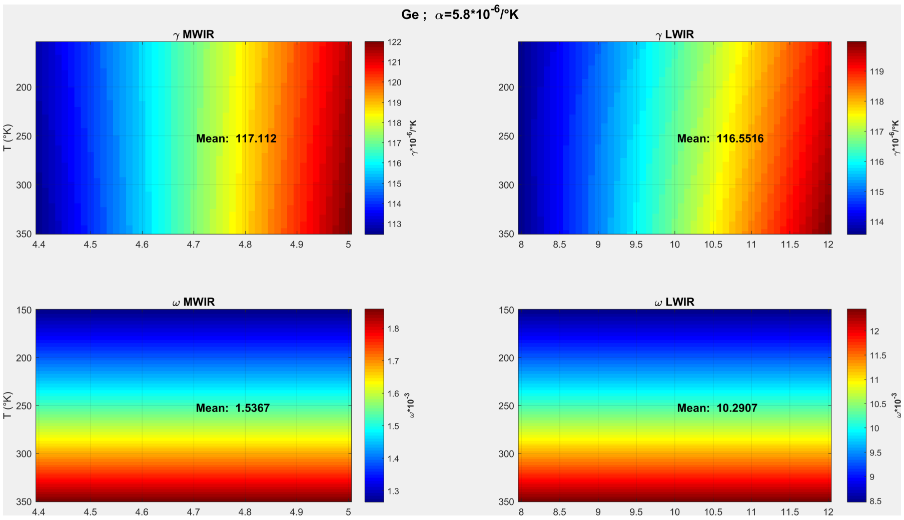


Figure 16.2: Ge thermal and chromatic powers in MWIR and LWIR wavebands.

16.2 Zinc Selenide ZnSe

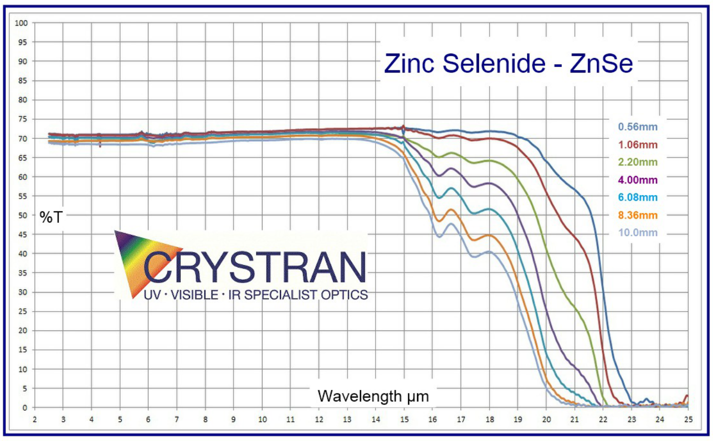


Figure 16.3: IR transmission of uncoated ZnSe.

ZnSe refractive index model comes from [102]. We define $t = T - 293(°\text{K})$:

$$\begin{cases} n^2(\lambda, T) = E(t) + \dfrac{A(t)}{\lambda^2 - \lambda_u^2} + \dfrac{B(t)}{\lambda^2/\lambda_I^2 - 1} \\ \lambda_u = 0.29934 + 1.004 * 10^{-4} t \\ \lambda_I = 48.38 + 6.29 * 10^{-3} t \\ E(t) = 9.01536 + 1.44190 * 10^{-3} t + 3.32973 * 10^{-7} t^2 - 1.08159 * 10^{-9} * t^3 - 3.88394 * 10^{-12} t^4 \\ A(t) = 0.24482 + 2.77806 * 10^{-5} t + 1.01703 * 10^{-8} t^2 - 4.51746 * 10^{-11} t^3 + 4.18509 * 10^{-13} t^4 \\ B(t) = 3.08889 + 1.13495 * 10^{-3} t + 2.89063 * 10^{-7} t^2 - 9.55657 * 10^{-10} t^3 - 4.76123 * 10^{-12} t^4 \end{cases} \tag{16.2}$$

ZnSe thermal and chromatic powers are displayed in Figure 16.4 for both MWIR and LWIR bands:

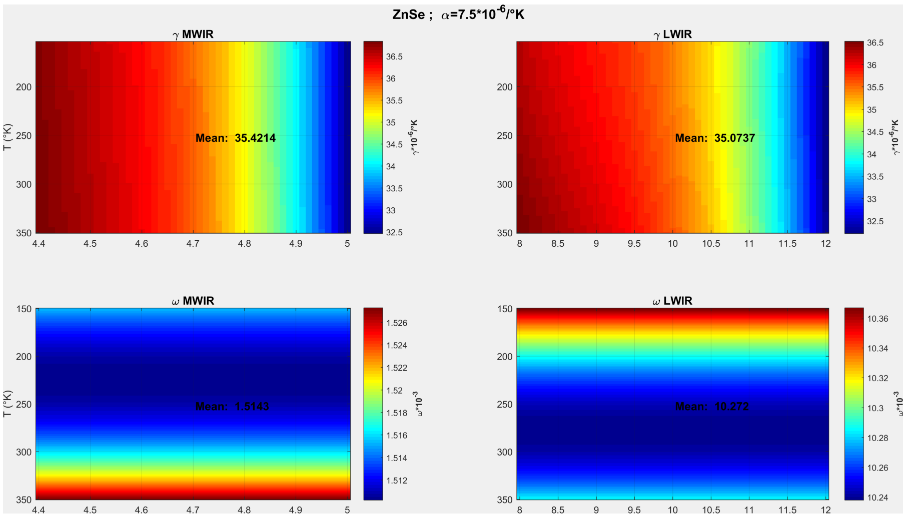


Figure 16.4: ZnSe thermal and chromatic powers in MWIR and LWIR wavebands.

16.3 Zinc Sulfide ZnS

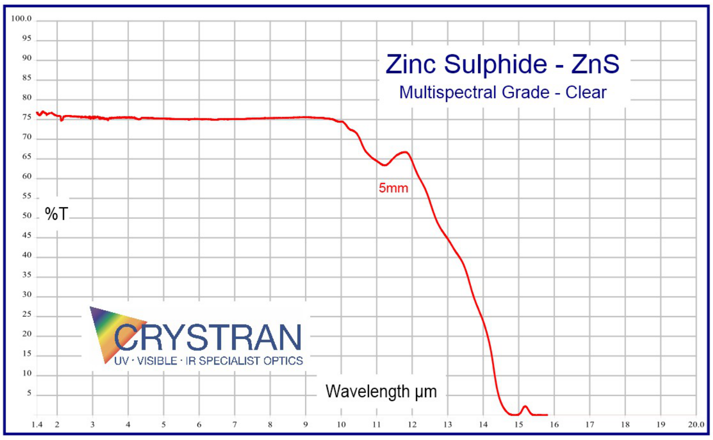


Figure 16.5: IR transmission of uncoated ZnS.

ZnS refractive index model comes from [102]. We define $t = T - 293(°\mathrm{K})$:

$$\begin{cases} n^2(\lambda, T) = E(t) + \dfrac{A(t)}{\lambda^2 - \lambda_u^2} + \dfrac{B(t)}{\lambda^2/\lambda_I^2 - 1} \\ \lambda_u = 0.23979 + 4.841 * 10^{-5} t \\ \lambda_I = 36.525 + 4.75 * 10^{-3} t \\ E(t) = 8.34096 + 1.29107 * 10^{-3} t + 4.68388 * 10^{-7} t^2 - 1.31683 * 10^{-9} t^3 - 6.64356 * 10^{-12} t^4 \\ A(t) = 0.1454 + 1.13319 * 10^{-5} t + 1.05932 * 10^{-8} t^2 + 1.06004 * 10^{-10} t^3 - 2.27671 * 10^{-13} t^4 \\ B(t) = 3.23924 + 1.096 * 10^{-3} t + 4.20092 * 10^{-7} t^2 - 1.1135 * 10^{-9} t^3 - 7.2992 * 10^{-12} t^4 \end{cases} \tag{16.3}$$

ZnS thermal and chromatic powers are displayed in Figure 16.6 for both MWIR and LWIR bands:

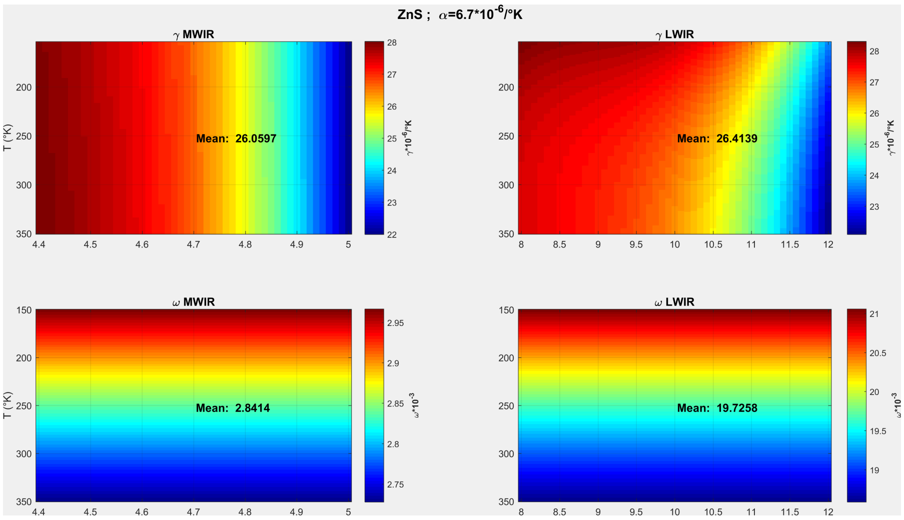


Figure 16.6: ZnS thermal and chromatic powers in MWIR and LWIR wavebands.

16.4 Gallium Arsenide GaAs

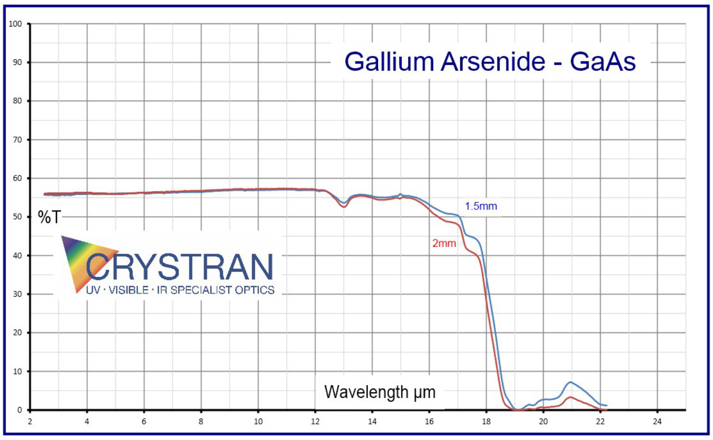


Figure 16.7: IR transmission of uncoated GaAs.

GaAs refractive index model comes from [103]. We define $t = T - 293(°\mathrm{K})$:

$$\begin{cases} n^2(\lambda, T) = G_0 + \dfrac{G_1}{\lambda_1^{-2} - \lambda.^{-2}} + \dfrac{G_2}{\lambda_2^{-2} - \lambda.^{-2}} + \dfrac{G_3}{\lambda_3^{-2} - \lambda.^{-2}} \\ \lambda_1 = 0.4431307 + 0.000050564 * t \\ \lambda_2 = 1746453 + 0.0001913 * t - 4.882e - 07 * t^2 \\ \lambda_3 = 36.9166 - 0.011622 * t \\ G_0 = 5.372514 \\ G_1 = 27.83972 \\ G_2 = 0.031764 + 4.350e - 05 * t + 4.664e - 07 * t^2 \\ G_3 = 0.00143636 \end{cases} \tag{16.4}$$

GaAs thermal and chromatic powers are displayed in Figure 16.8 for both MWIR and LWIR bands:

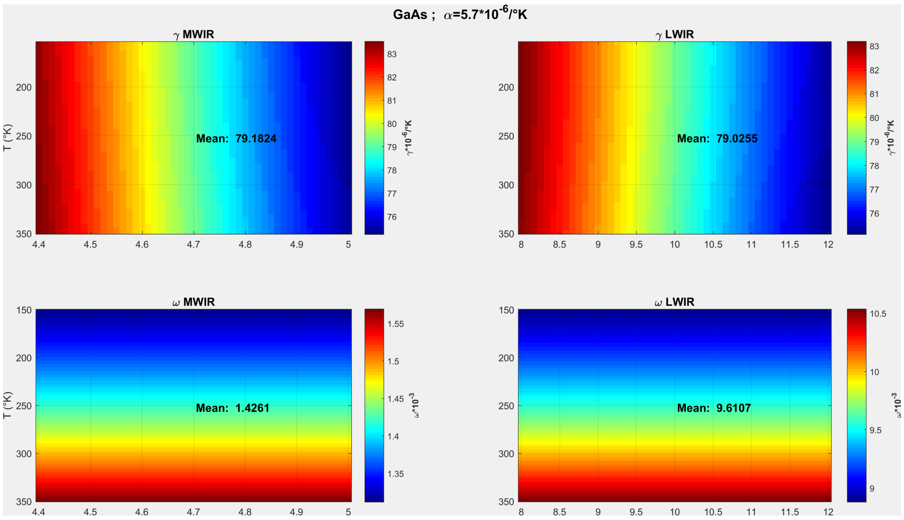


Figure 16.8: GaAs thermal and chromatic powers in MWIR and LWIR wavebands.

16.5 Cadmium Telluride CdTe

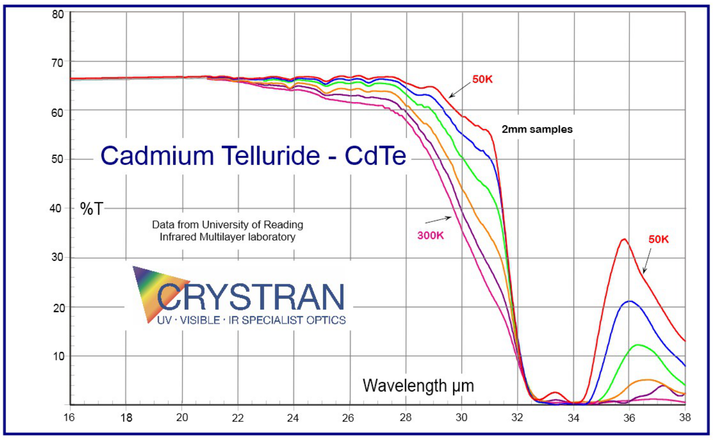


Figure 16.9: IR transmission of uncoated CdTe.

CdTe refractive index model comes from [104]:

$$\begin{cases} n^2(\lambda,T) = A + \dfrac{B\lambda^2}{\lambda^2 - C} + \dfrac{D\lambda^2}{\lambda^2 - E} \\ A = -2.973e-04 * T + 3.8466 \\ B = 8.057e-04 * T + 3.2215 \\ C = -1.10e-04 * T + 0.1866 \\ D = -2.160e-02 * T + 12.718 \\ E = -31.60 * T + 18753 \end{cases} \tag{16.5}$$

CdTe thermal and chromatic powers are displayed in Figure 16.10 for both MWIR and LWIR bands:

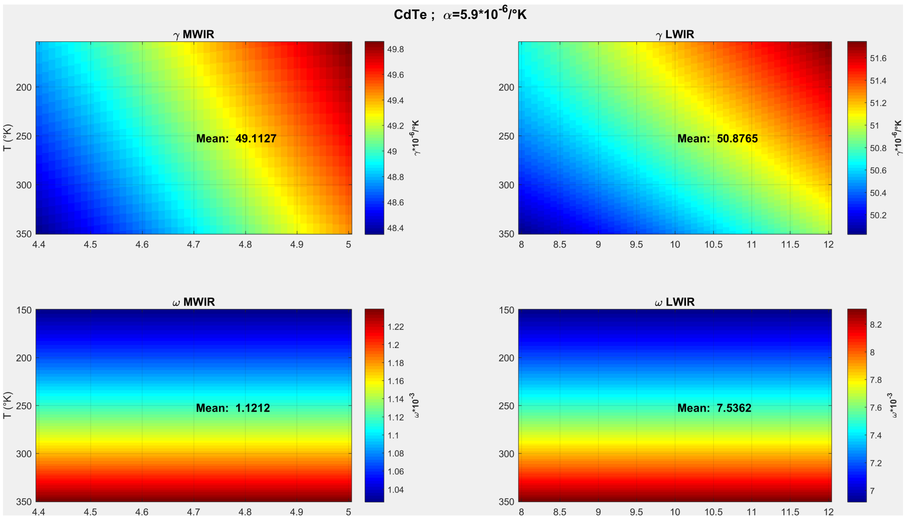


Figure 16.10: CdTe thermal and chromatic powers in MWIR and LWIR wavebands.

16.6 Silver Chloride AgCl

AgCl is not a very hard material, so it is easy to polish. However, it can be deformed under heat and pressure. AgCl transmission is 80% in both wavebands. It darkens when exposed to sunlight, but it doesn't affect its IR transmission.

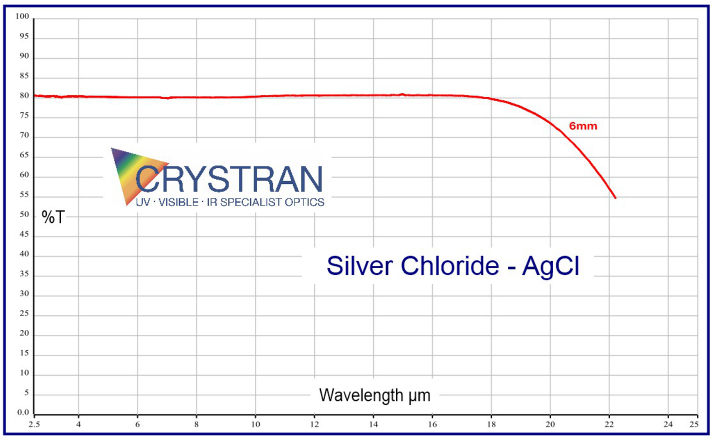


Figure 16.11: IR transmission of uncoated AgCl.

AgCl refractive index model comes from [105] for the chromatic dispersion. We obtain AgCl thermal dispersion (constant) using [106].

$$\begin{cases} \dfrac{dn}{dT} = -61 \\ n^2(\lambda, 296.9) = A + \dfrac{B}{\lambda^2 - C} - D\lambda^2 - E\lambda^4 \\ \Rightarrow n(\lambda, T) = \dfrac{dn}{dT}(T - (296.9)) + \sqrt{n^2(\lambda, 296.9)} \\ A = 4.00804 \\ B = 0.079086 \\ C = 0.04584 \\ D = 0.00085111 \\ E = 0.00000019762 \end{cases} \tag{16.6}$$

AgCl thermal and chromatic powers are displayed in Figure 16.12 for both MWIR and LWIR bands:

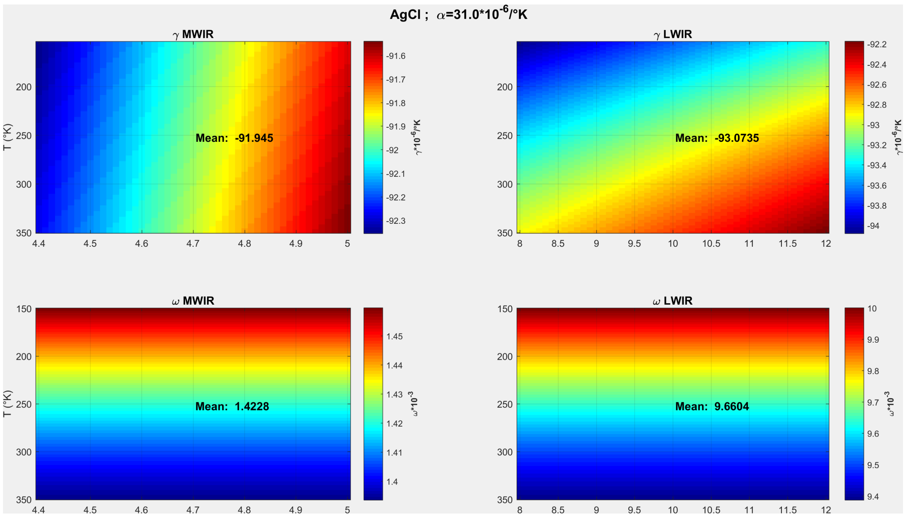


Figure 16.12: AgCl thermal and chromatic powers in MWIR and LWIR wavebands.

16.7 Chalcogenide glasses (IRG family)

The refractive index of all chalcogenide materials follows the same model, coming from SCHOTT datasheets [53]. It is valid for any temperature between 1 µm and 12 µm.

$$\begin{cases} n(\lambda, 295K) = \sqrt{1 + \dfrac{B_1\lambda^2}{\lambda^2 - C_1} + \dfrac{B_2\lambda^2}{\lambda^2 - C_2} + \dfrac{B_3\lambda^2}{\lambda^2 - C_3}} \\ \dfrac{dn}{dT}(\lambda, T) = \dfrac{n^2(\lambda, 295) - 1}{2n(\lambda, 295)}[D_0 + \dfrac{E_0}{\lambda^2 - \lambda_{TK}^2}] \end{cases} \tag{16.7}$$

By simple integration we get: $n(\lambda, T) = n(\lambda, 295) + \dfrac{dn}{dT}(T - 295)$. Table 16.1 displays the above mentioned parameters for each chalcogenide material [53]:

	IRG22	IRG23	IRG24	IRG25	IRG26	IRG27
B_1	2.4834	3.7107	2.8965	2.7574	3.1934	1.4005
B_2	2.8203	3.1091	2.9567	3.099	3.5854	3.4429
B_3	0.9773	0.8202	0.9461	1.666	2.2337	1.0394
C_1	0	0	0	0	0	0
C_2	0.1352	0.2459	0.162	0.1596	0.1852	0.0871
C_3	1420.7	1451.5	1939.1	2045.5	3398.08	805.5
D_0	$6.41.10^{-5}$	$8.44.10^{-5}$	$1.8.10^{-5}$	$5.5.10^{-5}$	$2.63.10^{-5}$	$-3.89.10^{-6}$
E_0	$1.68.10^{-5}$	$1.51.10^{-5}$	$2.1.10^{-5}$	$1.21.10^{-5}$	$2.93.10^{-5}$	$1.77.10^{-5}$
λ_{TK}	0.996	0.736	$1.67.10^{-3}$	$7.59.10^{-5}$	0.315	0.240

Table 16.1: Refractive index parameters (Equation 16.7) for each chalcogenide material [53].

Chalcogenide materials thermal and chromatic power are displayed in Figures 16.13, 16.14, 16.15, 16.16, 16.17, 16.18 for both MWIR and LWIR wavebands:

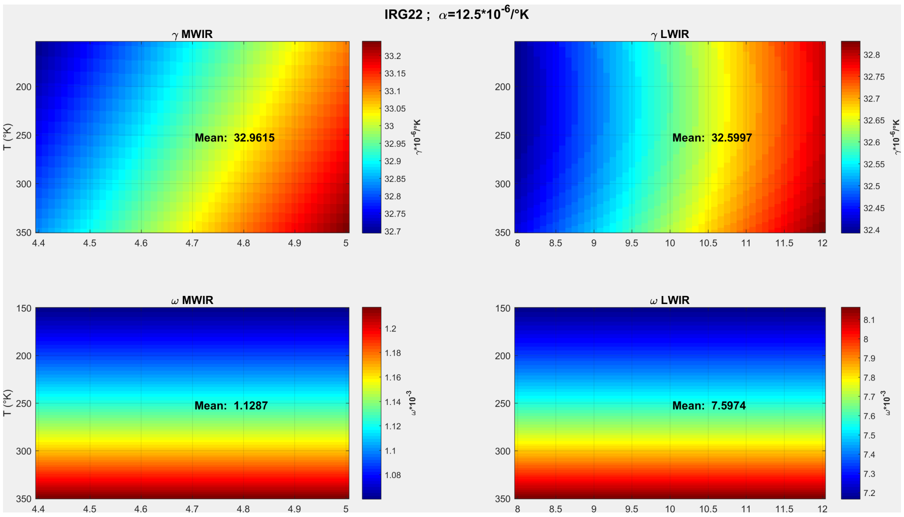


Figure 16.13: IRG22 thermal and chromatic powers in MWIR and LWIR wavebands.

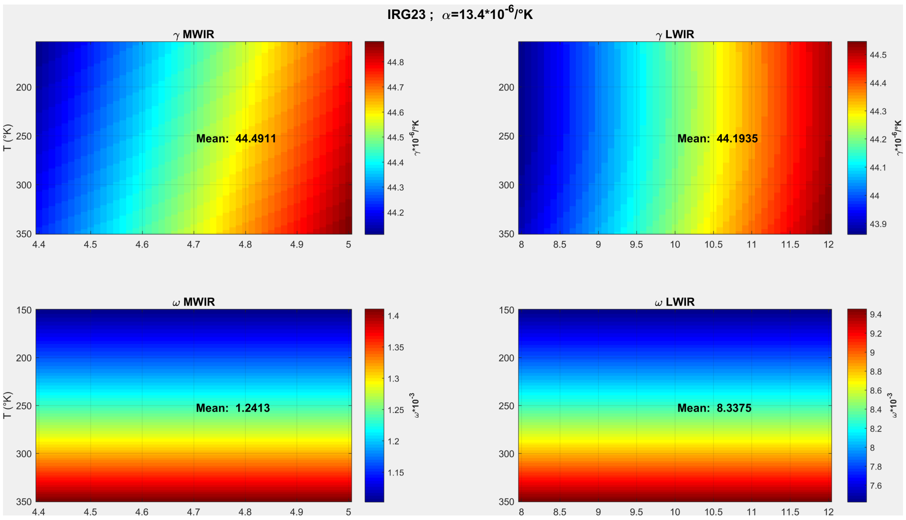


Figure 16.14: IRG23 thermal and chromatic powers in MWIR and LWIR wavebands.

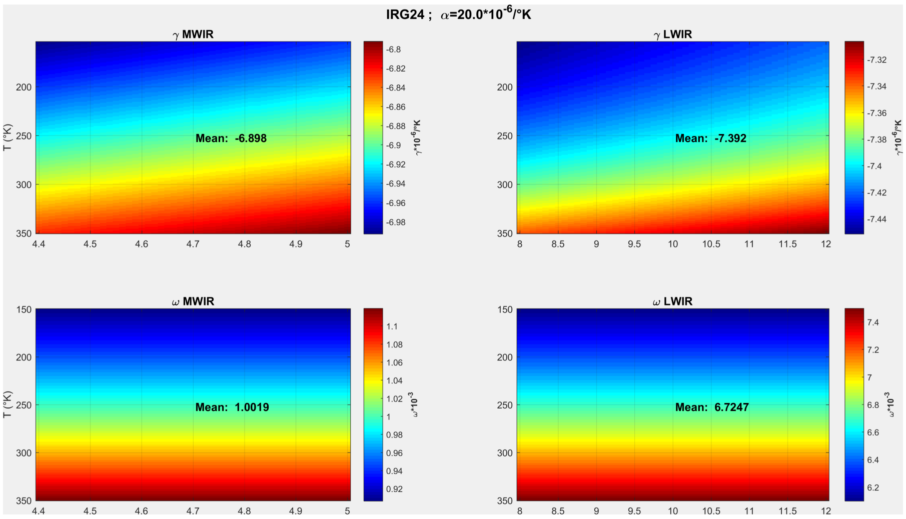


Figure 16.15: IRG24 thermal and chromatic powers in MWIR and LWIR wavebands.

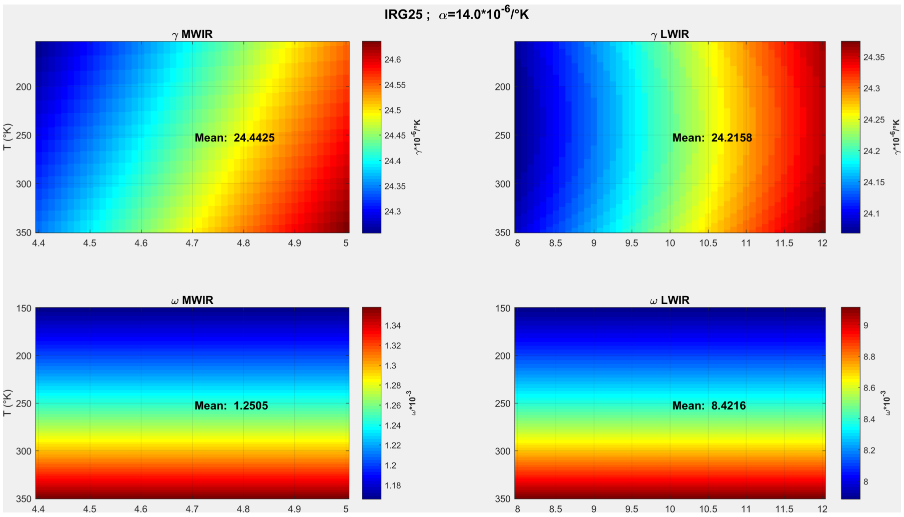


Figure 16.16: IRG25 thermal and chromatic powers in MWIR and LWIR wavebands.

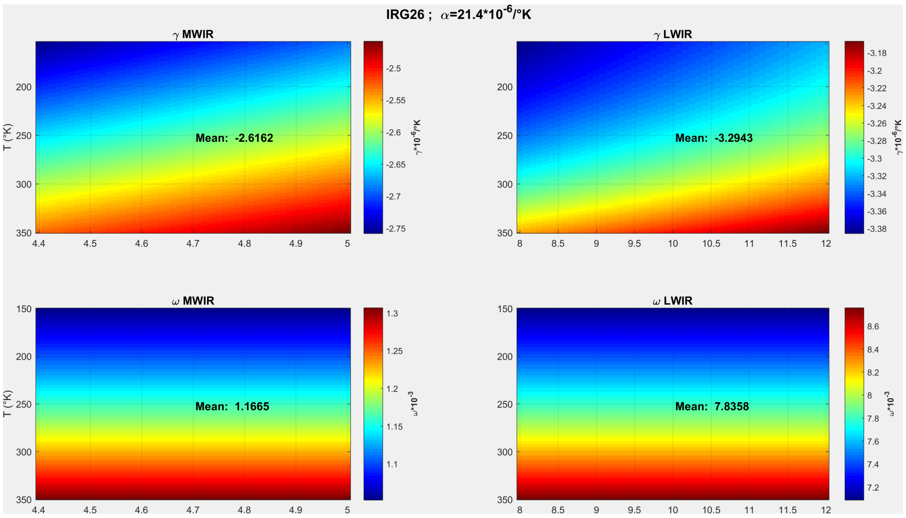


Figure 16.17: IRG26 thermal and chromatic powers in MWIR and LWIR wavebands.

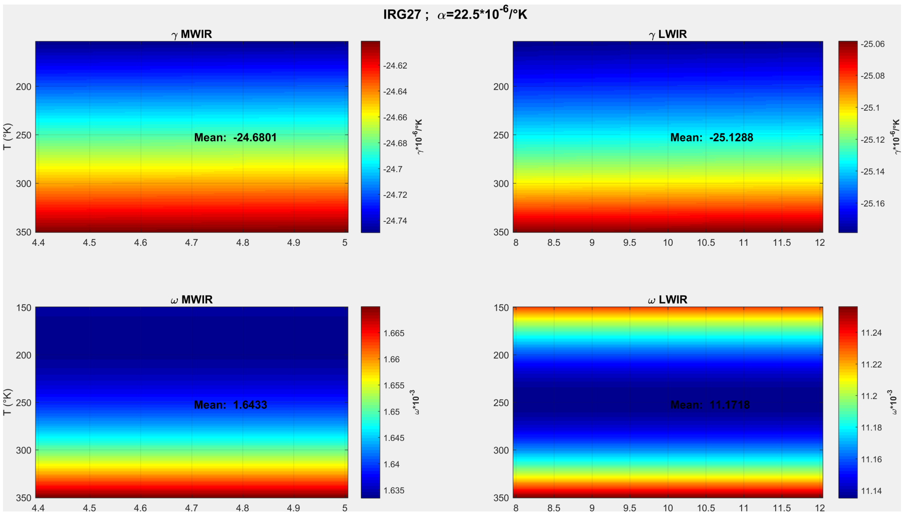


Figure 16.18: IRG27 thermal and chromatic powers in MWIR and LWIR wavebands.

Chapter 17

Appendix B: Resolution of MLDOE intersection

System 5.4 is solved for two types of profiles:

- Parabola: $z_m = -H_m \left(m - \frac{r^2}{r_1^2} \right) + S_m$
- Edge: $r = r_m$ (or $r_{m\pm1}$ depending on layer configuration and ray direction)

17.1 Parabola intersection

The solution of System 5.4 is rewritten:

$$\begin{cases} x = \frac{a_i}{b_i}(y - y_0) + x_0 \\ z = \frac{c_i}{b_i}(y - y_0) + z_0 \\ z = -H_m \left(m - \frac{r^2}{r_1^2} \right) + S_m \\ r^2 = x^2 + y^2 \end{cases} \tag{17.1}$$

S_m is the profile shift along the optical axis z and allows defining the profile z position. Once more, depending on the value of the direction vector parameter b: two cases arise:

17.2 Case: $b_i \neq 0$

In this case, for most rays, the expressions of r^2 and z^2 are rather "bulky" y^2 and y polynomials. The second-degree equation to solve is given by:

$$\begin{cases} x = \dfrac{a_i}{b_i}(y - y_0) + x_0 \\ z = \dfrac{c_i}{b_i}(y - y_0) + z_0 \\ Ay^2 + By + C = 0 \\ \Rightarrow A = 1 + \dfrac{a_i^2}{b_i^2} \\ \Rightarrow B = 2x_0\dfrac{a_i}{b_i} - 2y_0\dfrac{a_i^2}{b_i^2} - \dfrac{r_1^2 c_i}{H_m b_i} \\ \Rightarrow C = y_0^2\dfrac{a_i^2}{b_i^2} - 2x_0 y_0\dfrac{a_i}{b_i} + x_0^2 - r_1^2 m + \dfrac{r_1^2}{H_m}\left[-(z_0 - S_m) + \dfrac{c_i}{b_i}y_0\right] \end{cases} \tag{17.2}$$

17.3 Case: $b_i = 0$

In this case, using Systems 5.4 and 17.1, we have a second degree equation in z to solve:

$$\begin{cases} y = y_0 \\ x = \dfrac{a_i}{c_i}(z - z_0) + x_0 \\ A(z - z_0)^2 + B(z - z_0) + C = 0 \\ \Rightarrow A = \dfrac{a_i^2}{c_i^2} \\ \Rightarrow B = 2x_0\dfrac{a_i}{c_i} - r_1^2 \\ \Rightarrow C = -mH_m r_1^2 + y_0^2 + x_0^2 - r_1^2 z_0 + S_m r_1^2 \end{cases} \tag{17.3}$$

Finally, the interface partial derivatives defining the director coefficient of the normal **N** are expressed. The MLDOE interface profile $f(x, y, z) = 0$ is first multiplied by the curvature radius $\mathcal{R} = r_1^2/(2H_m)$ to simplify the normal director vector **N**:

$$\begin{cases} \dfrac{df}{dx} = x \\ \dfrac{df}{dy} = y \\ \dfrac{df}{dz} = -\dfrac{r_1^2}{2H_m} \end{cases} \tag{17.4}$$

17.4 Edge intersection

An MLDOE edge is defined as a cylinder with axis Z, centered on $x = y = 0$, with constant radius r_m (or $r_{m\pm1}$ depending on layer configuration and ray direction). The solution of System 5.4 is rewritten:

$$\begin{cases} y = \dfrac{b_i}{a_i}(x - x_0) + y_0 \\ z = \dfrac{c_i}{a_i}(x - x_0) + z_0 \\ r_m^2 = \sqrt{y^2 + x^2} \end{cases} \tag{17.5}$$

System 17.5 is a second degree equation in x:

$$\begin{cases} y = \dfrac{b_i}{a_i}(x - x_0) + y_0 \\ z = \dfrac{c_i}{a_i}(x - x_0) + z_0 \\ Ax^2 + Bx + C = 0 \\ \Rightarrow A = 1 + \dfrac{b_i^2}{a_i^2} \\ \Rightarrow B = 2y_0\dfrac{b_i}{a_i} - 2x_0\dfrac{b_i^2}{a_i^2} \\ \Rightarrow C = x_0^2\dfrac{b_i^2}{a_i^2} - 2\dfrac{b_i}{a_i}x_0y_0 + y_0^2 - r_m^2 \end{cases} \tag{17.6}$$

In the case where $a_i = 0$, the solution trivially becomes $x = x_0$, $y = \sqrt{r_m^2 - x^2}$ and $z = \dfrac{c_i}{b_i}(y - y_0) + z_0$.

Finally, the cylinder interface partial derivatives defining the director coefficient of the normal are trivially expressed:

$$\begin{cases} \dfrac{df}{dx} = 1 \\ \dfrac{df}{dy} = 1 \\ \dfrac{df}{dz} = 0 \end{cases} \tag{17.7}$$