OPEN ACCESS

JEOS-RP
Journal of the EOS
European Optical Society

The Journal of the European Optical Society – Rapid Publications



Topical Issue

EOSAM 2025

edp sciences

Table of Contents

Editorial 1-3
Kurcsics M and Eberhard P, **Versatile absorption modeling for transmissive optical elements using ray tracing and finite element analysis** 4-12
Jimenez C and Herkommer A, **3D printed freeform micro-optics meets generated Jacobian equations based inverse designs** 13-18
Bociort F, **Quasi-powers and primary aberrations of thin lenses in contact** 19-27
Chen W-J, **Zernike by one Pascal triangle: A high performance, low memory cost and flexible computation scheme for Zernike polynomials** 28-42
Reichelt S, **Visualization of wavefront aberrations by Zernike polynomials** 43-49
Cho H, Park H, Jee MJ and Kim D, **Large resolution-weighted étendue space telescope designs categorized by tertiary mirror location diversity** 50-56
Santarsiero M, Martínez-Herrero R, González de Sande JC, Piquero G, Korotkova O and Gori F, **Propagation of partially coherent fields radiated by sources with univariable cross-spectral density** 57-72
Mukundakumar B and Matsuura S, **Coherence control and wavelength correction in multi-wavelength digital holography** 73-78
Dubois A, Aloïs Baudry and Sylvie Lebrun, **Study of tapered microfibers by line-field confocal optical coherence tomography** 79-84
Piaia V, Robalinho P, Silva S and Frazão O, **Fano-like resonance in optical fiber** 85-96
Tainta S, Salinas A, Janices I, San-Miguel U, Sanchez-Gonzalez A, Erro MJ and Perez-Herrera RA, **L-band dual-wavelength fiber ring laser with automated polarization stabilization and switching for remote sensing applications** 97-103
Doval A, Suárez-Fernández L, de la Fuente R and Arosa Y, **Anomalous reflection coefficient underlying experimentally validated outstanding transmission through metal–dielectric–metal microcavities** 104-107
Tok Ç, Satılmış M, Keleş H, Oğuz F, Sarı H and Özbay E, **Numerical modeling of InGaAs/AlGaAs heterostructure for short-wave infrared nBn photodetectors** 108-116
Bogucki O, Kaczkan M, Kozłowska A, Pantoł-Boczon A and Malinowska A, **Luminescent thermal history sensing potential in Pr3+-activated YAG matrix** 117-121
Anăstăsoaie V, Mihalache I, Brincoveanu O, Dinescu A, Kusko C, Cristea D and Tomescu R, **Enhancing the fluorescence of low quantum yield fluorophores using plasmonic metasurfaces** 122-128
Siefke T, Gerold K, Shestaeva S, Paul P, Alam S, Franta D, Szeghalmi A, Schröder S and Kroker S, **Atomic layer deposition for hafnium oxide-based meta-optics in the ultraviolet spectral range** 129-133
Grimpe C-F, Lüßmann-Sorokina A, Du G, Sah P, Sauer S, Jordan E, Thomas R, Gehrmann P, Lipkin M, Suckow S, Lemme MC, Kroker S and Mehlstäubler TE, **Benchmarking dual-polarization silicon nitride photonic integrated circuits for trapped-ion quantum technologies** 134-143

Kroneberger M, Stanik E and Döhring T, **Verifying optical simulations for infrared curing of powder coatings** 144-147
Wong CL, **Monkeypox virus (MPXV) neutralizing antibodies diagnosis with spectral-phase surface plasmon resonance 3D imaging biosensor** 148-152
Martins Garcia M, Bessif B, Erni D and Jalali M, **Determining the health state of a single exosome via nano-coated silica microresonator** 153-165
Takino H, Aizawa H and Kanaoka M, **Evaluation of surface topography and microstructure of electroplated copper ultraprecision-machined with a single-crystal diamond tool** 166-174
Frazão O, Silva S, Corela C, Loureiro A, Gonçalves S, Robalinho P, Sousa R, Martins HF, Carrilho F, Omira R, Niehus M and Matias L, **Comparison of seismic records obtained by distributed acoustic sensing and ocean bottom seismometers** 175-178
Martins P, Van Golde I, Silva S, Frazão O and Sousa R, **Modeling of surface vessels using distributed acoustic sensing data and physics-based optimization** 179-183
Albert K, Klein J, Ligges M and Grabmaier A, **Eye-safe non-line-of-sight localization using compact nanosecond laser diodes and single-photon-avalanche-diode arrays** 184-195
MV Imperatore, Trespidi F, Minuto A and Timò G, **In-situ monitoring of a flashed solar simulator spectral characteristics for multijunction solar cells performance assessment** 196-200

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, E1
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026066
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

EDITORIAL

OPEN ACCESS

Editorial

Omar El Gawhary[1], Stefan Witte[2], and Ignacio Moreno[3,*]

[1] Imaging Physics Department, Faculty Applied Sciences. Delft University of Technology, and ASML Research, The Netherlands
[2] Imaging Physics Department, Faculty Applied Sciences. Delft University of Technology, The Netherlands
[3] Instituto de Bioingeniería, Universidad Miguel Hernández, Avda. Universidad s/n, 03202, Elche, Spain

Received 14 September 2026 / Accepted 14 September 2026

The 13th edition of the European Optical Society Annual Meeting, EOSAM 2025, was held onsite from 24 to 28 of August 2025 on Delft, a picturesque, historic canal city in the Netherlands, famous for its distinctive blue-and-white pottery (Delftware), its rich history as the birthplace of painter Johannes Vermeer, and its vibrant student energy.

Hosted by TU Delft, the oldest and largest public technical university in the Netherlands, EOSAM 2025 brought together nearly 440 participants from more than 30 countries, including researchers, students, industry professionals, and leading experts. With a program that featured 10 plenary sessions, 10 tutorial sessions, 54 parallel sessions of 90 min each, 290 oral presentations, 82 posters, and 19 commercial exhibitors, EOSAM 2025 offered a comprehensive view of current developments in Optics and Photonics, arranged in nine Topical Meetings (TOM) covering adaptive and freeform optics, biophotonics and biosensors, nanophotonics, optical materials, ultrafast phenomena, optical fiber technology, quantum polaritons, nonlinear and quantum optics and applications of optics and photonics, and four Focused Sessions covering visual optics and imaging, optics of complex wavefronts, mid-infrared photonics and photonics for cultural heritage. In addition, EOSAM 2025 included the Face2Phase meeting, centered around the use of phase information in optics.

The scientific program was complemented by high-level plenary talks and tutorial presentations, an EU project session, a joint session between EOS and the Chinese Optical Society (COS), a poster session, a vibrant exhibition, and special industry events. The meeting confirmed the growing impact of optics and photonics as key enabling technologies for scientific progress, sustainable development, and innovation in Europe and beyond.

Following previous EOSAM conferences, this issue of the Journal of the European Optical Society–Rapid Publications (JEOS-RP) gathers 25 articles that expands related communications presented during the conference. They cover different topics which reflect the scientific diversity and vitality of the conference.

One first group of articles present advanced research in optical engineering, wavefront analysis and optical system optimization, spanning from cutting-edge fabrication techniques to grand-scale astronomical applications. Kurcsics and P. Eberhard model thermal loading in optical elements by mapping ray-traced heat sources directly onto a finite element mesh [1]. At the micro-scale, Jimenez and Herkommer [2] explore advanced fabrication by merging 3D-printed freeform micro-optics with inverse design methods based on Jacobian equations to precisely mold light. Bociort [3] revisits fundamental analytical optical theory, establishing mathematical tools to analyze the quasi-powers and primary aberrations of classic thin lenses in contact. Moving into computational methods, Chen [4] introduces a highly efficient, low-memory mathematical scheme to compute Zernike polynomials using Pascal's triangle, while Reichelt [5] utilizes Zernike polynomials for the practical visualization of wavefront aberrations. Finally, at the macro-scale, Cho et al. [6] apply design principles to aerospace engineering, categorizing large, high-étendue space telescope designs based on variations in tertiary mirror locations.

Other articles approach the domain of optical coherence theory and its applications in digital imaging. Santarsiero et al. [7] examine how partially coherent fields propagate when radiated by specific sources with univariable cross-spectral density. Mukundakumar and Matsuura [8] introduce methods for coherence control and precise wavelength correction to reduce noise and errors in multi-wavelength digital holography. Dubois et al. [9] implement an advanced metrology approach using line-field confocal optical coherence tomography to non-destructively study and map the profile of tapered glass microfibers.

Fiber-optic devices are also covered by Piaia et al. [10], who explore wave phenomena by generating and analyzing a sharp, asymmetric Fano-like resonance within an optical fiber, and by Tainta et al. [11], who develop an L-band dual-wavelength fiber ring laser integrating automated

* Corresponding author: i.moreno@umh.es

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

polarization stabilization and switching mechanism specifically optimized for remote sensing.

Another group of papers covers optical materials and light-matter interactions for diverse advancements in nanophotonics and optoelectronics. Doval et al. [12] investigate anomalous reflection coefficients to explain and experimentally validate high light transmission through metal–dielectric–metal microcavities. Tok et al. [13] utilize computational engineering to perform numerical modeling on InGaAs/AlGaAs heterostructures, optimizing it for high-performance short-wave infrared photodetectors. Bogucki et al. [14] explores the optical properties of rare-earth ions, evaluating the luminescent thermal history sensing potential of a Pr^{3+}-activated YAG matrix for temperature tracking. Anăstăsoaie et al. [15] utilize plasmonic metasurfaces to enhance the emission and fluorescence of low quantum yield fluorophores for biochemical or imaging applications. Siefke et al. [16] address layer deposition (ALD) to manufacture hafnium oxide-based meta-optics useful to efficiently manipulate ultraviolet light. Grimpe et al demonstrates dual-polarization silicon nitride photonic circuits to expand beam delivery capabilities in trapped-ion quantum systems [17]. Finally, Kroneberger et al. [18] study the accuracy of optical simulations used for the infrared curing of industrial powder coatings.

Moving to applications of optics and photonics, two articles address issues in biophotonics: Wong [19] describes a spectral-phase surface plasmon resonance 3D imaging biosensor to diagnose and track Monkeypox virus (MPXV), and Martins Garcia et al. [20] implement a nano-coated silica microresonator to precisely evaluate and determine the health state of an individual exosome.

Other diverse engineering advancements centered on optical instrumentation and optical technologies are presented. Takino et al. [21] evaluated the microscopic surface topography of electroplated copper treated with ultraprecision diamond machining tools. Frazão et al. [22] study the efficacy of distributed acoustic sensing (DAS) against traditional ocean-bottom seismometers for capturing seismic records, while Martins et al. [23] build upon DAS technology by combining its acoustic data with physics-based optimization to accurately model and track surface vessels. In the area of active optical scanning, Albert et al. [24] develops an eye-safe, non-line-of-sight localization system utilizing compact nanosecond laser diodes paired with single-photon avalanche diode arrays. Finally, Imperatore et al. [25] implement an optical monitoring setup to track the spectral characteristics of a flashed solar simulator in real-time for multijunction solar cells.

This diverse collection of research articles serves as an excellent compendium of the multidisciplinary advancements driving the fields of optics and photonics, testifying to the high scientific quality of the work presented at Delft during EOSAM 2025. We thank all the authors for their valuable contributions and the reviewers for their thoughtful assessments.

References

1 Kurcsics M and Eberhard P, Versatile absorption modeling for transmissive optical elements using ray tracing and finite element analysis, *J. Eur. Opt. Society-Rapid Publ.* **21**(2), 47 (2025). https://doi.org/10.1051/jeos/2025043.

2 Jimenez C and Herkommer A, 3D printed freeform micro-optics meets generated Jacobian equations based inverse designs, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 3 (2026). https://doi.org/10.1051/jeos/2025055

3 Bociort F, Quasi-powers and primary aberrations of thin lenses in contact, *JEOS-RP* **22**(1), 30 (2026). https://doi.org/10.1051/jeos/2026025.

4 Chen W-J, Zernike by one Pascal triangle: A high performance, low memory cost and flexible computation scheme for Zernike polynomials, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 31 (2026). https://doi.org/10.1051/jeos/2026023.

5 Reichelt S, Visualization of wavefront aberrations by Zernike polynomials, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 50 (2026). https://doi.org/10.1051/jeos/2026034.

6 Cho H, Park H, Jee MJ and Kim D, Large resolution-weighted étendue space telescope designs categorized by tertiary mirror location diversity, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 35 (2026). https://doi.org/10.1051/jeos/2026035.

7 Santarsiero M, Martínez-Herrero R, González de Sande JC, Piquero G, Korotkova O and Gori F, Propagation of partially coherent fields radiated by sources with univariable cross-spectral density, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 19 (2026). https://doi.org/10.1051/jeos/2026011.

8 Mukundakumar B and Matsuura S, Coherence control and wavelength correction in multi-wavelength digital holography, *J. Eur. Opt. Society-Rapid Publ.* **22**(2), 58 (2026). https://doi.org/10.1051/jeos/2026051.

9 Dubois A, Aloïs Baudry and Sylvie Lebrun, Study of tapered microfibers by line-field confocal optical coherence tomography, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 26 (2026). https://doi.org/10.1051/jeos/2026029.

10 Piaia V, Robalinho P, Silva S and Frazão O, Fano-like resonance in optical fiber, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 23 (2026). https://doi.org/10.1051/jeos/2026016.

11 Tainta S, Salinas A, Janices I, San-Miguel U, Sanchez-Gonzalez A, Erro MJ and Perez-Herrera RA, L-band dual-wavelength fiber ring laser with automated polarization stabilization and switching for remote sensing applications, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 54 (2026). https://doi.org/10.1051/jeos/2026042.

12 Doval A, Suárez-Fernández L, de la Fuente R and Arosa Y, Anomalous reflection coefficient underlying experimentally validated outstanding transmission through metal–dielectric–metal microcavities, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 5 (2026). https://doi.org/10.1051/jeos/2025056.

13 Tok Ç, Satılmış M, Keleş H, Oğuz F, Sarı H and Özbay E, Numerical modeling of InGaAs/AlGaAs heterostructure for short-wave infrared nBn photodetectors, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 39 (2026). https://doi.org/10.1051/jeos/2026038.

14 Bogucki O, Kaczkan M, Kozłowska A, Pantoł-Boczon A and Malinowska A, Luminescent thermal history sensing potential in Pr3+-activated YAG matrix, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 18 (2026). https://doi.org/10.1051/jeos/2026010.

15 Anăstăsoaie V, Mihalache I, Brincoveanu O, Dinescu A, Kusko C, Cristea D and Tomescu R, Enhancing fluorescence of low quantum yield fluorophores with plasmonic metasurfaces, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 28 (2026). https://doi.org/10.1051/jeos/2026022.

16 Siefke T, Gerold K, Shestaeva S, Paul P, Alam S, Franta D, Szeghalmi A, Schröder S and Kroker S, Atomic layer deposition for hafnium oxide-based meta-optics in the ultraviolet spectral range, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 33 (2026). https://doi.org/10.1051/jeos/2026036.
17 Grimpe C-F, Lüßmann-Sorokina A, Du G, Sah P, Sauer S, Jordan E, Thomas R, Gehrmann P, Lipkin M, Suckow S, Lemme MC, Kroker S and Mehlstäubler TE, Benchmarking dual-polarization silicon nitride photonic integrated circuits for trapped-ion quantum technologies, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 56 (2026). https://doi.org/10.1051/jeos/2026049.
18 Kroneberger M, Stanik E and Döhring T, Verifying optical simulations for infrared curing of powder coatings, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 41 (2026). https://doi.org/10.1051/jeos/2026033.
19 Wong CL, Monkeypox virus (MPXV) neutralizing antibodies diagnosis with spectral-phase surface plasmon resonance 3D imaging biosensor, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 34 (2026). https://doi.org/10.1051/jeos/2026039.
20 Martins Garcia M, Bessif B, Erni D and Jalali M, Determining the health state of a single exosome via nano-coated silica microresonator, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 53 (2026). https://doi.org/10.1051/jeos/2026046.
21 Takino H, Aizawa H and Kanaoka M, Evaluation of surface topography and microstructure of electroplated copper ultraprecision-machined with a single-crystal diamond tool, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 29 (2026). https://doi.org/10.1051/jeos/2026030.
22 Frazão O, Silva S, Corela C, Loureiro A, Gonçalves S, Robalinho P, Sousa R, Martins HF, Carrilho F, Omira R, Niehus M and Matias L, Comparison of seismic records obtained by distributed acoustic sensing and ocean bottom seismometers, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 11 (2026). https://doi.org/10.1051/jeos/2026007.
23 Martins P, Van Golde I, Silva S, Frazão O and Sousa R, Modeling of surface vessels using distributed acoustic sensing data and physics-based optimization, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 13 (2026). https://doi.org/10.1051/jeos/2026008.
24 Albert K, Klein J, Ligges M and Grabmaier A, Eye-safe non-line-of-sight localization using compact nanosecond laser diodes and single-photon-avalanche-diode arrays, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 40 (2026). https://doi.org/10.1051/jeos/2026019.
25 MV Imperatore, Trespidi F, Minuto A and Timò G, In-situ monitoring of a flashed solar simulator spectral characteristics for multijunction solar cells performance assessment, *J. Eur. Opt. Society-Rapid Publ.* **22**(1), 52 (2026). https://doi.org/10.1051/jeos/2026045.

J. Eur. Opt. Society-Rapid Publ. 2025, **21**, 47
© The Author(s), published by EDP Sciences, 2025
https://doi.org/10.1051/jeos/2025043
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE OPEN ACCESS

Versatile absorption modeling for transmissive optical elements using ray tracing and finite element analysis

Mark Kurcsics* and Peter Eberhard

Institute of Engineering and Computational Mechanics, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Received 23 June 2025 / Accepted 30 September 2025

Abstract. Emerging quality requirements in modern optical systems increase the need for accurate simulation and compensation of disturbances. One key disturbance in high-power laser applications results from thermal loading due to absorption. This paper presents a flexible modeling approach for absorption based on ray tracing, where point heat sources are defined along ray paths. In addition, several methods for mapping these point heat sources along the ray path to the finite element mesh are introduced and compared. A major advantage of the proposed approach lies in its flexibility, as it enables the simulation of transient phenomena such as dynamic beam positioning and time-dependent variations in beam shape.

Keywords: Absorption, High-energy laser, Finite element method, Ray tracing, Heat source mapping.

1 Introduction and motivation

Modern optical systems, such as those used in high-power laser applications, are highly complex and demand exceptional process stability. Various factors can impair their performance, with thermal disturbances being particularly significant. These disturbances may arise from ambient temperature changes or, in the case of high-power lasers, from partial absorption of the laser beam within optical elements, which leads to temperature gradients inside the material [1]. These gradients result in changes to the refractive index and induce thermal deformations, both of which degrade optical performance [2].

Although analytical models exist to describe these effects [3–6], they rely on strong simplifications, limiting their applicability in complex, real-world systems [1, 2]. At the same time, numerical simulation has become an essential and effective tool. Simulations enable design evaluations early in the development process, helping to reduce costs by minimizing the need for late-stage adjustments.

Ray tracing has become the standard technique for simulating optical behavior, while the finite element method is widely used for thermal analysis. Coupling these methods provides a robust foundation for modeling absorption effects in optical elements. Ideally, a numerical algorithm for computing absorption should be flexible enough to account for arbitrary beam shapes, distortions induced by optical components, and time-dependent variations in beam shape and position.

Several approaches have been proposed in the literature to address this challenge. In [7] an approach is presented based on computing the heat flux on an auxiliary mesh, which is then mapped to the finite element mesh. The method supports both rotationally symmetric and non-symmetric static beam profiles. However, it is not suitable for modeling time-dependent variations. Another method is proposed by [8] that accounts for intensity distributions by using ray paths to define individual heat sources, which are mapped to the finite element mesh. However, heat distribution within a lens is calculated only through linear interpolation, and time-dependent effects are not considered. The algorithm described by [9] is developed for laser welding. It applies ray tracing with each ray's power being proportional to the local beam intensity. However, since the method is designed for modeling absorption on metal surfaces, it is not applicable to optical elements. The method presented by Lyu and Zhan [10] is a coupled ray tracing-finite element approach. Heat sources are calculated and transferred to the finite element mesh, but the contribution to each element is averaged. However, this averaging requires a fine thermal mesh to be able to resolve gradients in the heat flux. A similar strategy is presented by Genberg et al. [11], where heat sources are computed along ray paths. An additional voxel mesh is introduced, and heat deposition per voxel is mapped to the finite element mesh. This extra additional mesh introduces additional the complexity to the absorption method. The method introduced by Cosson et al. [12] computes absorbed energy within volume elements. The power of each ray is determined based on its initial position relative to a Gaussian beam

* Corresponding author: mark.kurcsics@itm.uni-stuttgart.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

profile. However, their focus is on inhomogeneous materials which necessitates to introduce macroscopic and microscopic meshes. An absorption model for selective laser melting is implemented by Dorussen et al. [13]. Rays are spatially distributed according to a normal distribution, and each ray receives a fraction of the total laser energy. As the algorithm is tailored to absorption in metallic particles, it is not applicable to optical elements. The method proposed by Rall et al. [14] incorporates an auxiliary mesh for intensity distribution. However, the mesh is reconstructed after tracing rays through the optical system, with absorption effects taken into account. The approach presented by Hahn and Eberhard [15] prescribes a heat flux for each ray on the surface intersection point and performs linear interpolation for obtaining the heat flux within the volume. However, no physical law is involved for calculating the heat flux.

In this paper, we present a novel approach for modeling light absorption in transmissive optical elements that directly couples ray tracing with the finite element method. Our method offers high flexibility in handling complex and time-dependent beam profiles enabling simulations of thermally induced effects, such as thermal deformation and spatially varying refractive index, in modern optical systems.

In Section 2 the methods are presented, implying the ray tracing method (Sect. 2.2), the finite element method (Sect. 2.3), the absorption model (Sect. 2.4) and the methods for mapping the point heat sources (Sect. 2.5). In Section 3 the simulation setup and the results are presented. The paper closes with Section 4 where the conclusions are presented.

2 Methods

2.1 Simulative procedure

The overarching simulation approach involves a sequential, multi-step process designed to model the thermal response of the optical element to absorption effects. The first step is to perform a transient ray tracing of the thermally undisturbed system for the complete simulation duration, as described in Section 2.2, which simulates the propagation of rays through the optical system and determines their interactions with surfaces and materials. The spatial distribution of ray positions and ray paths obtained from this step serve as the basis for calculating the heat sources resulting from energy absorption, detailed in Section 2.4. These heat sources are computed by evaluating the energy deposited as the rays interact with surface and bulk material, taking into account heat absorption.

Once the absorption-induced heat sources have been identified and the heat originating form each of these is determined, the next step is to map these discrete heat values onto the nodes of the finite element mesh (see Sect. 2.5). This mapping process is crucial for ensuring that the energy input is represented within the finite element framework. With the heat sources distributed across the mesh, the final stage of the simulation involves solving for the resulting temperature field for each timestep of the simulation. This is accomplished using the thermal finite element method (Sect. 2.3), which computes the temperature distribution based on the imposed heat flux and the governing equations of heat transfer. Together, these steps form an integrated simulation pipeline that enables dynamic thermal analysis based on ray-based energy deposition.

2.2 Ray tracing

A commonly used technique for optical analysis is ray tracing, which models the propagation of light through an optical system by discretizing the light source into a large number of individual rays. These rays are then traced through the system using principles of vector algebra and geometrical optics [16], allowing for a detailed simulation of light behavior such as reflection, refraction, and absorption.

The mathematical basis for ray propagation in general form is described by a differential equation

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(n(\boldsymbol{r})\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\right)=\frac{\partial n}{\partial s}, \tag{1}$$

where $n(\boldsymbol{r})$ denotes the refractive index at a given position $\boldsymbol{r}$, and s is the geometric path length, which represents the physical distance traveled by the ray [17]. This formulation accounts for the effects of spatial inhomogeneity in the refractive index, which becomes particularly important in gradient-index (GRIN) media. In such cases, the light path is curved due to variations in refractive index, for example, as a result of temperature gradients or material composition changes.

In contrast, for homogeneous media where the refractive index is constant throughout the domain, the equation simplifies significantly, and rays propagate in straight lines.

There are various implementations of ray tracing. In sequential ray tracing, for instance, the interaction of each ray with the optical surfaces is computed in the specific order in which the elements are arranged in the system. At each interface, the ray is typically refracted or reflected based on the vector form of Snell's law, which determines the change in direction according to the refractive indices of the two media and the angle of incidence.

As discussed in Section 2.4, the ray tracing algorithm must also account for the spatial intensity distribution of the light source. This is particularly important when modeling physically realistic sources. For example, if the light source emits a Gaussian beam, the rays should be distributed in space according to a Gaussian probability density function. This ensures that the ray density, and thus the simulated light intensity, faithfully represents the actual beam profile.

2.3 Thermal finite element method

The finite element method is a widely adopted numerical technique used for solving partial differential equations across various physical domains, including structural mechanics, fluid dynamics, and heat transfer. In the context of this study, the focus lies on thermal phenomena, and thus the thermal finite element method is employed to compute temperature fields resulting from localized heat sources.

At its core, the finite element method involves discretizing a physical body into a finite number of elements, which are connected at nodes. These elements can take various shapes, such as triangles, quadrilaterals, or tetrahedra, depending on the geometry and dimensionality of the domain. The nodes, which form the vertices of the elements, play a central role in the numerical solution. The primary unknown, the temperature in this case, is computed at these nodes, and the field within each element is approximated using interpolation functions (shape functions) based on the nodal values.

The governing heat conduction equation, typically a second-order partial differential equation in space and first-order in time, is transformed into a system of ordinary differential equations through spatial discretization into n_{el} elements, each with n_{en} nodes, and a total of n_{tn} nodes

$$\boldsymbol{D}_t \dot{\boldsymbol{T}}(t) + \boldsymbol{K}_t \boldsymbol{T}(t) = \boldsymbol{f}(t) \tag{2}$$

with the capacitance matrix $\boldsymbol{D}_t \in \mathbb{R}^{n_{\mathrm{tn}} \times n_{\mathrm{tn}}}$, the conductance matrix $\boldsymbol{K}_t \in \mathbb{R}^{n_{\mathrm{tn}} \times n_{\mathrm{tn}}}$, the time-dependent flux vector $\boldsymbol{f}(t) \in \mathbb{R}^{n_{\mathrm{tn}} \times 1}$, and the time-dependent temperature vector $\boldsymbol{T}(t) \in \mathbb{R}^{n_{\mathrm{tn}} \times 1}$ [18, 19]. This differential equation is then solved for the temperature vector $\boldsymbol{T}(t)$, which provides the temporal evolution of nodal temperatures.

In detail, the flux vector $\boldsymbol{f}(t)$ is the excitation term of the differential equation. It includes the effects of volumetric heat sources and surface heat sources and defines the heat input for each node. Here, the flux is governed by the volumetric absorption as well as the absorption due to anti-reflection coatings on the surface. The global heat flux vector $\boldsymbol{f}$ is composed through summation over the element heat flux vectors $\boldsymbol{f}^e$. Thereby, the element heat flux $\boldsymbol{f}^e$ is mapped to the corresponding globally numbered nodes through the element gather matrices $\boldsymbol{L}^e \in \mathbb{R}^{n_{\mathrm{tn}} \times n_{\mathrm{en}}}$ [19]

$$\boldsymbol{f} = \sum_{e=1}^{n_{\mathrm{el}}} \boldsymbol{L}^e \boldsymbol{f}^e \,. \tag{3}$$

While gather matrices provide a good conceptual basis for understanding the summation process and the assembly of global vectors and matrices, it is numerically more efficient in implementation to use lookup tables, where for each local node in an element the corresponding global node number is stored. These lookup tables are typically used in computational frameworks to speed up the matrix assembly process and minimize overhead, especially in large-scale simulations.

Overall, the thermal finite element method offers a robust and flexible framework for modeling transient and steady-state heat transfer problems. Its ability to handle complex geometries and heterogeneous material properties makes it particularly suitable for simulations involving spatially distributed heat sources, such as those arising from optical absorption in this study.

2.4 Absorption modeling

In optical systems incorporating lenses with anti-reflective coatings, two primary absorption mechanisms contribute to heat generation. The first is surface absorption due to the absorptive properties of the coating, while the second is volumetric absorption within the lens material itself [20]. The latter is typically modeled using Beer-Lambert's law [21, 22], which describes the exponential attenuation of an initial light intensity I_{init} as it traverses a medium of thickness z with a material-specific absorption coefficient α_V

$$I(x, y, z) = I_{\mathrm{init}}(x, y) e^{-\alpha_{\mathrm{V}} z} \,. \tag{4}$$

The goal of the proposed absorption algorithm is to bridge ray tracing and thermo-mechanical finite element method in a flexible and generalizable manner. Since the Beer-Lambert's law is inherently tied to the local intensity of the light field, the strategy involves distributing rays spatially to represent the beam's intensity profile. For instance, in the case of a Gaussian beam, this means sampling ray positions such that the ensemble reflects the Gaussian distribution of intensity.

To each individual ray an initial power $P_{\mathrm{ray,init}}$ is assigned such that the total beam power P_{tot} is distributed across n_{rays} rays

$$P_{\mathrm{ray,init}} = \frac{P_{\mathrm{tot}}}{n_{\mathrm{rays}}} \,. \tag{5}$$

By preserving the intensity distribution through spatial ray placement and equal power assignment, the algorithm remains adaptable to arbitrary initial beam shapes and accounts for beam reshaping by refractive elements.

After initialization, each ray is traced through the optical system to determine its propagation path and interaction points (see Sect. 2.2). Subsequently, the absorbed heat along each ray is computed individually.

In the following, the absorption process is described in the context of a single lens. When a ray strikes the front surface of the lens, a fraction of its power is absorbed by the anti-reflective coating. This surface absorption is modeled using a scalar surface absorption factor α_S, which serves as a lumped parameter representing complex physical dependencies that are difficult to model explicitly [7, 21]. The resulting point heat source on the surface $q_{\mathrm{frontsurf}}$ is computed as

$$q_{\mathrm{frontsurf}} = \alpha_{\mathrm{S}} P_{\mathrm{ray,inc}} \,. \tag{6}$$

The incoming ray power $P_{\mathrm{ray,inc}}$ equals the initial ray power $P_{\mathrm{ray,init}}$ for the first lens; for subsequent optical elements, $P_{\mathrm{ray,inc}}$ is updated to account for prior absorptive losses. The surface heat source $q_{\mathrm{frontsurf}}$ is then mapped to the surrounding finite element nodes using the method described in Section 2.5.

After traversing the front surface, the remaining ray power undergoes volumetric attenuation according to Beer-Lambert's law. The residual power after passing through a lens is

$$P_{\mathrm{ray}}(l) = (P_{\mathrm{ray,inc}} - q_{\mathrm{frontsurf}}) e^{-\alpha_{\mathrm{V}} l}, \tag{7}$$

where l is the geometric path length within the lens. Because the absorption occurs continuously along the ray path, it must be discretized for compatibility with the finite element model. To this end, the ray path between its front entry point $\boldsymbol{r}_{\mathrm{front}}$ and back exit point $\boldsymbol{r}_{\mathrm{back}}$ is subdivided into n_{seg} segments of equal length l_{seg}

$$l_{\text{seg}} = \frac{||\boldsymbol{r}_{\text{back}} - \boldsymbol{r}_{\text{front}}||_2}{n_{\text{seg}}} \,. \tag{8}$$

The absorbed heat in each segment $i \in [1, n_{\text{seg}}]$ is computed recursively. Let P_{ray_i} denote the ray power entering segment i, and $q_{\text{vol}\,i}$ the volumetric heat absorbed in that segment. The computational scheme is then

$$\begin{aligned} P_{\text{ray}_i} &= P_{\text{ray}_{i-1}} - q_{\text{vol}\,i-1}, \\ q_{\text{vol}\,i} &= P_{\text{ray}_i}(1 - \mathrm{e}^{-\alpha_{\text{V}} l_{\text{seg}}}), \\ P_{\text{ray}_0} &= P_{\text{ray,inc}} - q_{\text{frontsurf}}, \quad q_{\text{vol}0} = 0 \,. \end{aligned} \tag{9}$$

Each volumetric heat flux contribution is modeled as a point heat source located at the segment center

$$\begin{aligned} \boldsymbol{r}_{\text{vol}\,i} &= \boldsymbol{r}_{\text{front}} + \frac{2i-1}{2} l_{\text{seg}} \boldsymbol{d} \quad \text{with} \\ \boldsymbol{d} &= \frac{\boldsymbol{r}_{\text{back}} - \boldsymbol{r}_{\text{front}}}{\|\boldsymbol{r}_{\text{back}} - \boldsymbol{r}_{\text{front}}\|_2} \,. \end{aligned} \tag{10}$$

These discrete point heat sources are then mapped to the finite element mesh using appropriate spatial weighting schemes as presented in Section 2.5.

At the back surface due to the coating, energy is absorbed again. The resulting point heat source is given by

$$q_{\text{backsurf}} = \alpha_{\text{S}} \left(P_{\text{ray}_{n_{\text{seg}}}} - q_{\text{vol}\,n_{\text{seg}}} \right) . \tag{11}$$

2.5 Heat source mapping

2.5.1 Mapping based on Shape functions

The mapping of heat sources into the finite element context is essential for transferring ray-tracing-based energy deposition into the thermal simulation domain. This section describes how discrete point heat sources, derived from optical ray tracing and optical absorption, are projected onto the finite element mesh using an isoparametric formulation.

In the isoparametric approach, both the geometry and the field variables within an element e with n_{en} nodes are interpolated using the same shape functions $\boldsymbol{N}(\boldsymbol{\xi}) = [N_1(\boldsymbol{\xi}) \ \ldots \ N_{n_{\text{en}}}(\boldsymbol{\xi})] \in \mathbb{R}^{1 \times n_{\text{en}}}$, which are defined in terms of local coordinates $\boldsymbol{\xi} = [\xi \, \eta \, \zeta] \in \mathbb{R}^{3\times 1}$ with $-1 \leq \xi, \eta, \zeta \leq 1$ [19]. The temperature field T^e and spatial coordinates $\boldsymbol{x}$ are thus represented as

$$T^e(\boldsymbol{\xi}) = \boldsymbol{N}(\boldsymbol{\xi}) \boldsymbol{T}^e \in \mathbb{R}, \tag{12}$$

$$\boldsymbol{x}(\boldsymbol{\xi}) = \begin{bmatrix} \boldsymbol{N}(\boldsymbol{\xi}) \boldsymbol{x}^e \\ \boldsymbol{N}(\boldsymbol{\xi}) \boldsymbol{y}^e \\ \boldsymbol{N}(\boldsymbol{\xi}) \boldsymbol{z}^e \end{bmatrix} \in \mathbb{R}^{3\times 1}, \tag{13}$$

where $\boldsymbol{x}^e, \boldsymbol{y}^e, \boldsymbol{z}^e \in \mathbb{R}^{n_{\text{en}}\times 1}$ are the global coordinates of the element's nodes in the x, y, and z directions and $\boldsymbol{T}^e \in \mathbb{R}^{n_{\text{en}}\times 1}$ is the vector of the nodal temperatures.

Given a point heat source q_{vol} located at position $\boldsymbol{r}_{\text{vol}}$, its contribution to the nodal heat flux must be distributed across the finite element mesh. Since the finite element formulation is inherently element-wise, it is first necessary to determine which element contains the point source. A simple strategy is to locate the closest node to $\boldsymbol{r}_{\text{vol}}$ and check the adjacent elements. For each candidate element, the global coordinates $\boldsymbol{r}_{\text{vol}}$ are mapped to the local coordinates $\boldsymbol{\xi}_{\text{vol}} = \boldsymbol{\xi}(\boldsymbol{r}_{\text{vol}})$, and containment is confirmed if each component of $\boldsymbol{\xi}_{\text{vol}}$ lies within the standard domain $[-1, 1]$.

The forward mapping from local to global coordinates, $\boldsymbol{x}(\boldsymbol{\xi})$, is well-defined, straightforward to evaluate numerically but in most cases nonlinear. Therefore, the inverse mapping $\boldsymbol{\xi}(\boldsymbol{x})$ is typically not analytically available [23]. To compute $\boldsymbol{\xi}_{\text{vol}}$, a Newton-Raphson iteration is employed by solving the nonlinear residual equation

$$\boldsymbol{g}(\boldsymbol{\xi}) = \boldsymbol{x}(\boldsymbol{\xi}) - \boldsymbol{r}_{\text{vol}} \tag{14}$$

using the iterative update

$$\begin{aligned} &\boldsymbol{J}(\xi_n)\Delta\boldsymbol{\xi}_n = -\boldsymbol{g}(\boldsymbol{\xi}_n), \text{ where } \boldsymbol{J}(\boldsymbol{\xi}) = \frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \frac{\partial \boldsymbol{x}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}, \\ &\boldsymbol{\xi}_{n+1} = \boldsymbol{\xi}_n + \Delta\boldsymbol{\xi}_n, \ \boldsymbol{\xi}_0 = \boldsymbol{0} \,. \end{aligned} \tag{15}$$

The iteration continues until the residual norm satisfies the convergence criterion

$$||\boldsymbol{g}(\boldsymbol{\xi}_{n+1})||_2 < \epsilon \,. \tag{16}$$

Once the containing element is identified, the heat source q_{vol} is modeled as a volumetric heat source density acting over an infinitesimally small volume, represented using the Dirac delta function centered at the source location

$$s = q_{\text{vol}} \delta(\boldsymbol{x} - \boldsymbol{r}_{\text{vol}}) \,. \tag{17}$$

The elemental heat flux vector $\boldsymbol{f}^e$ is given by a volume integral over the element volume $\boldsymbol{\Omega}^e$ in global coordinates

$$\boldsymbol{f}^e = \int_{\Omega^e} \boldsymbol{N}^\top s \mathrm{d}\boldsymbol{\Omega} = \int_{\Omega^e} \boldsymbol{N}^\top q_{\text{vol}} \delta(\boldsymbol{x} - \boldsymbol{r}_{\text{vol}}) \mathrm{d}\boldsymbol{\Omega} \,. \tag{18}$$

Applying the isoparametric transformation to switch from global to local coordinates results in

$$\boldsymbol{f}^e = \int_{\tilde{\Omega}^e} \boldsymbol{N}^\top(\boldsymbol{\xi}) q_{\text{vol}} \delta(\boldsymbol{x}(\boldsymbol{\xi}) - \boldsymbol{r}_{\text{vol}}) |J(\boldsymbol{\xi})| \mathrm{d}\tilde{\boldsymbol{\Omega}}, \tag{19}$$

where the Jacobian determinant $J(\boldsymbol{\xi}) = \det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\xi}}\right)$ accounts for the transformation of volume. As outlined by [24], the delta function can be reformulated in terms of local coordinates

$$\delta(\boldsymbol{x}(\boldsymbol{\xi}) - \boldsymbol{r}_{\text{vol}}) = \frac{\delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{r}_{\text{vol}}))}{|J(\boldsymbol{\xi})|} = \frac{\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{\text{vol}})}{|J(\boldsymbol{\xi})|} \,. \tag{20}$$

The local coordinates $\boldsymbol{\xi}_{\text{vol}} = \boldsymbol{\xi}(\boldsymbol{r}_{\text{vol}})$ are reused from the determination of the element membership as described earlier. Then, the integral results in

$$\boldsymbol{f}^e = \int_{\tilde{\Omega}^e} \boldsymbol{N}^\top(\boldsymbol{\xi}) q_{\text{vol}} \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{\text{vol}}) \mathrm{d}\tilde{\boldsymbol{\Omega}} \,. \tag{21}$$

Then, the sifting property of the Dirac delta yields the heat flux vector

$$\boldsymbol{f}^e = \boldsymbol{N}^\top(\boldsymbol{\xi}_{\mathrm{vol}})\, q_{\mathrm{vol}}\,. \tag{22}$$

This provides the nodal distribution of the point heat source q_{vol} within the containing element. The resulting local heat flux vector $\boldsymbol{f}^e$ is then assembled into the global heat flux vector $\boldsymbol{f}$ based on the global node indices of the element nodes as explained in Section 2.3.

2.5.2 Mapping based on global inverse distances

Since the procedure involving shape functions requires the use of the Newton-Raphson method to compute the local coordinates, it introduces iterative computations that can be computationally expensive, especially when handling a large number of point heat sources. To reduce the computational effort, an alternative approach is proposed that bypasses the need for local coordinate mapping. This method is based on a weighted inverse distance mapping of point heat sources directly to nearby nodes of the finite element mesh. It saves computation time at the expense of being less accurate.

In this approach, for each point heat source located at position $\boldsymbol{r}_{\mathrm{vol}}$, the Euclidean distance to the closest mesh node is first computed. If this distance is exactly zero, meaning the heat source coincides with a mesh node, the full heat contribution q_{vol} is assigned directly to that node, eliminating the need for further distribution. However, in the more general case where the heat source lies between mesh nodes, the heat must be distributed across several nearby nodes to reflect its spatial influence.

To achieve this, the Euclidean distances d_j from the heat source to a selected number n_{dist} of neighboring nodes with $j \in [1, 2, \ldots, n_{\mathrm{dist}}]$ are computed. The heat is then distributed to these nodes inversely proportional to their distance from the source, such that closer nodes receive a larger share of the heat. The heat flux contribution to each node is given by

$$f_j = \frac{d_j^{-1}\, q_{\mathrm{vol}}}{\sum_{k=1}^{n_{\mathrm{dist}}} d_k^{-1}}\,. \tag{23}$$

This formulation ensures conservation of energy, as the total distributed heat sums to q_{vol}. Additionally, the inverse-distance weighting yields a smooth and physically plausible distribution without requiring knowledge of the element shape functions or iterative solving.

Finally, the computed nodal contributions f_j are assembled into the global heat flux vector $\boldsymbol{f}$ by adding them to the corresponding global node indices. This approach significantly simplifies the implementation and provides, therefore, a practical compromise between computational simplicity and spatial accuracy.

2.5.3 Mapping based on element-wise inverse distances

A third possibility is to combine both previously discussed approaches by mapping the point heat source to the nodes of an enclosing finite element using inverse-distance weighting. This hybrid method aims to retain the local resolution of shape-function-based mapping while avoiding the complexity of iterative schemes. For each point heat source, the closest node is identified first. If the heat source position exactly coincides with a node, the corresponding heat q_{vol} is directly assigned to that node. Otherwise, the elements are determined that include the closest node. To identify the potential enclosing element, Delaunay triangulation is used to connect the nodes for each neighboring element. Each candidate element is then tested to check whether it contains the heat source position. Once the enclosing element is found, the heat is distributed to its n_{en} nodes using inverse-distance weights

$$f_j = \frac{d_j^{-1}\, q_{\mathrm{vol}}}{\sum_{k=1}^{n_{\mathrm{en}}} d_k^{-1}}, \qquad j \in [1, 2, \ldots, n_{\mathrm{en}}], \tag{24}$$

where d_j denotes the Euclidean distance from the heat source to node j. As in the previous strategies, the individual nodal heat loads f_j are assembled into the global flux vector $\boldsymbol{f}$ based on the global node numbering.

3 Simulation setup and results

In the following, the behavior of the absorption algorithm and associated heat mapping procedures is evaluated using a simplified scenario involving a single plano-convex N-BK7 lens. The focus of the simulations is on examining the qualitative behavior. However, realistic material parameters are chosen. Exemplarily, for N-BK7 at a wavelenght of $\lambda = 1.064$ µm the volumetric absorptivity is $\alpha_{\mathrm{V}} = 0.1286\ \frac{1}{\mathrm{m}}$ [25]. For the surface absorptivity $\alpha_S = 50 \times 10^{-6}$ [7] is selected. For the finite element mesh isoparametric hexahedral elements are chosen. As initial temperature for the lens 20 °C is preset as well as a fixed temperature boundary condition of 20 °C at the outer cylindrical surface of the lens. Heat generation is computed for the beam path and mapped to the mesh using different strategies (using shape functions, global inverse-distances and element-wise inverse distances). The implementation of the whole simulation procedure is done in Matlab.

In the first simulation setup, spatial variation in energy deposition is introduced by illuminating the lens with two inclined Gaussian laser beams, each intersecting the lens at a different incident angle. An exemplary simulation result is shown in Figure 1, illustrating the computed temperature fields resulting from the three heat mapping strategies: (a) shape function-based projection, (b) global inverse distance weighting, and (c) element-wise inverse distance weighting. Despite of the use of a relatively coarse finite element mesh, all three methods qualitatively reproduce the expected thermal behavior. Specifically, the temperature field exhibits a gradient from the front (left) to the back (right) surface of the lens, in accordance with the attenuation of laser intensity through the material. Additionally, a localized temperature increase is observed in the center area of the beams.

A closer comparison reveals that the shape function-based method and the global inverse distance method yield similar temperature distributions, both in terms of spatial

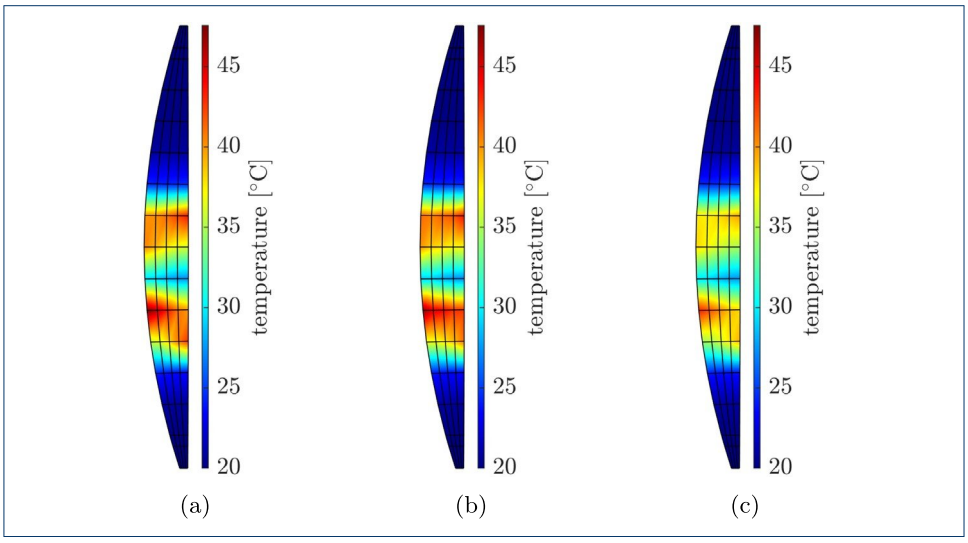


Figure 1. Exemplary simulation of two transmitted Gaussian laser beams for different mapping methods while $n_{\text{seg}} = 30$ point heat sources along each ray within the lens are considered. (a) Mapping with shape function. (b) Mapping with global inverse distances. (c) Mapping with element-wise inverse distances.

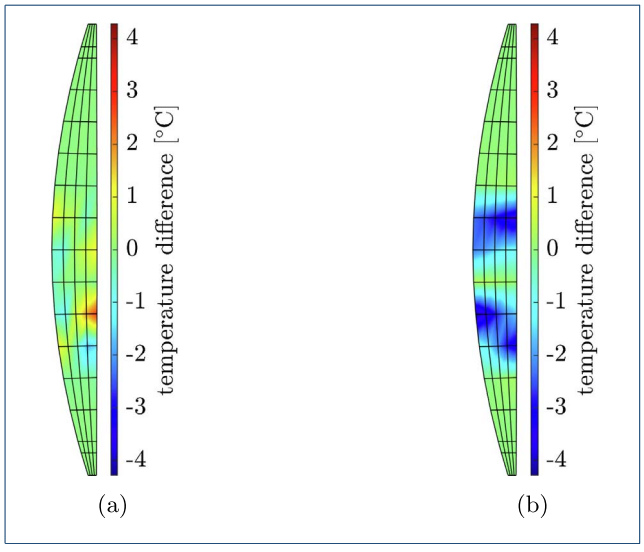


Figure 2. Comparing the mapping methods with inverse distances to the method with shape functions. (a) Mapping with shape fusnctions vs. mapping with global inverse distances. (b) Mapping with shape functions vs. mapping with element-wise inverse distances.

shape and peak values. This indicates that both approaches are capable of distributing the absorbed energy over the mesh nodes in a consistent and physically plausible manner. In contrast, the element-wise inverse distance method results in noticeably lower peak temperatures.

A more detailed comparison of the heat mapping strategies is provided in Figure 2, where the two inverse distance-based methods are directly compared to the shape function-based mapping. The plot illustrates how each method distributes the heat from the heat source over the finite element mesh and reveals differences in how localized or diffuse the resulting temperature profiles become.

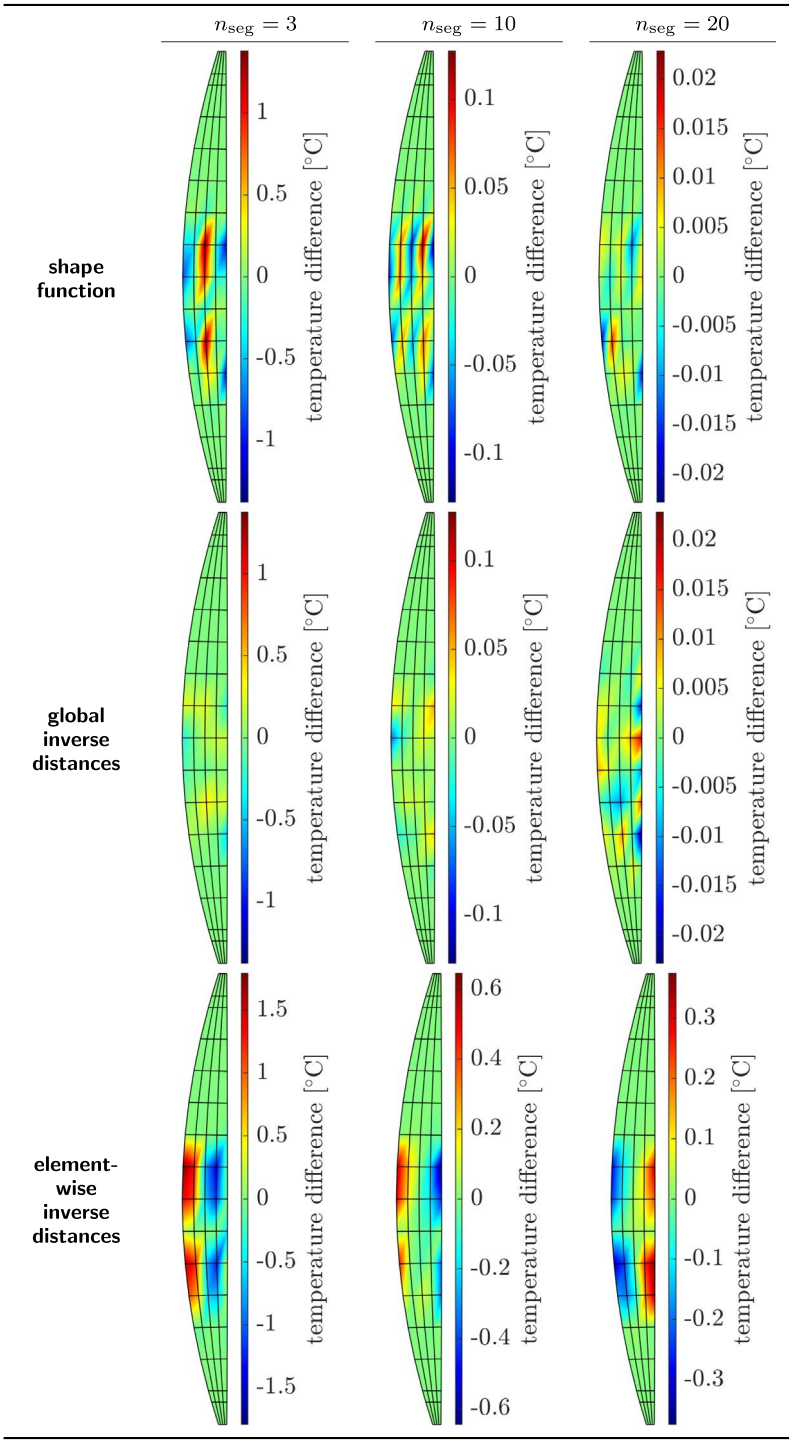
It can be observed, that the mapping method using global inverse distance weighting exhibits only minor deviations from the reference solution obtained with shape functions. This suggests that global weighting provides a sufficiently smooth and consistent heat distribution across the mesh. In contrast, the element-wise inverse distance method shows more pronounced discrepancies.

The influence of the number of point heat sources n_{seg} along each ray within the lens on the resulting temperature distribution for the three mapping procedures is analyzed in Table 1. For this purpose, n_{seg} is varied across several values for each mapping method, with the case of $n_{\text{seg}} = 30$ serving as the reference solution.

Across all mapping strategies, it is observed that increasing the number of point heat sources improves the accuracy of the resulting temperature distribution. The temperature profiles become more consistent with the reference solution as the discretization of the heat input becomes finer. For this reason it can be concluded that it is not necessary to have a very fine discretization of the volumetric absorption, which reduces the computational effort.

Among the three methods, the mapping approaches using shape functions and global inverse distances exhibit a faster convergence towards the reference. This indicates that these methods are capable of capturing the spatial characteristics of the heat input more effectively, even with a relatively coarse segmentation of the ray. In contrast, the element-wise inverse distance method shows slower convergence and does not reach the same level of accuracy as the other two methods. Notably, the most significant deviations occur near the entry and exit zones of the laser beams. This behavior is attributed to how the volumetric absorption is discretized and mapped within individual elements. Specifically, due to the segment-based sampling of the beam path, it can happen that a disproportionate

Table 1. Convergence study varying the number n_{seg} of volumetric point heat sources along a single ray for each mapping method with $n_{\text{seg}} = 30$ serving as the reference solution.



number of point heat sources fall into the entry element compared to the exit element. This causes an imbalance in heat input, resulting in artificially elevated temperatures at the lens front and suppressed temperatures at the rear. Furthermore, in the element-wise mapping approach, heat contributions from each point source are distributed to all nodes of the enclosing element based on inverse distances. Since all distances are finite, even for nodes that are physically further away, every node within the element receives a non-negligible share of the heat input. This results in a spatially smoother heat distribution and lower local peak temperatures compared to the other methods.

In summary, the element-wise method tends to diffuse the heat more broadly and can introduce asymmetries at

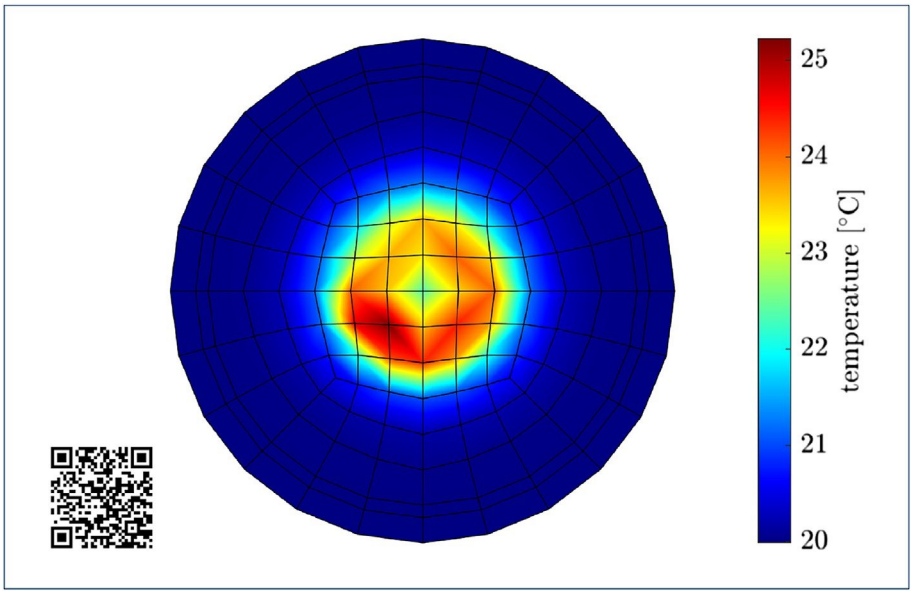


Figure 3. Transient temperature field for laser beam following a circular path. The QR code leads to a video presented in [26].

the boundaries of the beam path due to localized oversampling or undersampling of segments. However, in intermediate regions of the beam path, where the number of heat sources per element tends to average out, this effect is less pronounced.

In addition to accuracy, computational performance is also a critical factor. All mapping strategies require, at a minimum, a nearest-node search for each point heat source. The global inverse distance method is computationally the least expensive, as it does not require element-specific computations beyond this initial search. On the other hand, the shape function method requires a more complex element search, which involves solving for local coordinates using the Newton-Raphson iteration. Therefore, the computation time is 217.13% longer compared to the global inverse distance method. Similarly, the element-wise inverse distance method relies on identifying the enclosing element, for which a Delaunay triangulation is used. The computation time is 1137.87% longer compared to the global inverse distance method.

From the simulation results, it is found that the Newton-Raphson method for shape function mapping is computationally less expensive than the Delaunay triangulation used for element-wise mapping. However, among all tested methods, the global inverse distance mapping demonstrated the lowest overall computational effort, as it only requires identifying the nearest nodes without the need for any element search or iterative procedures. While being slightly less accurate than the shape function method in certain cases, its simplicity and speed make it particularly attractive for applications where computational resources are limited or where a fast approximation is sufficient. Therefore, depending on the required trade-off between accuracy and performance, the global inverse distance method can be a very effective alternative for efficiently modeling volumetric heat deposition in finite element simulations.

An exemplary simulation involving a time-varying laser beam is presented in Figure 3. In this setup, the laser beam is prescribed to move along a circular trajectory on the lens surface, introducing continuous spatial and temporal variation in energy deposition. The temperature field is computed using the previously described algorithm with the mapping method based on shape functions selected for its higher accuracy and robustness. As seen in the resulting temperature distribution, the trajectory of the laser beam is clearly reflected in the thermal footprint, showing a distinct circular heating pattern. This confirms the method's capability to resolve dynamic beam motion and transient heat input. The simulation highlights not only the spatial resolution of the approach, but also its ability to maintain stability and physical accuracy over time, making it well-suited for applications involving scanning or modulated laser sources in thermal-optical simulations.

4 Conclusion

This paper presents an algorithm for modeling both surface and volumetric laser absorption in optical elements, with an emphasis on methods for mapping point heat sources along each ray to finite element nodes. The proposed framework offers high flexibility in handling spatial and temporal variations in beam parameters and element positioning, making it well-suited for simulating realistic optical systems under complex illumination conditions.

Among the mapping methods evaluated, the global inverse distance approach demonstrates the lowest computational cost while producing temperature distributions that closely resemble those obtained from the more accurate shape function-based method. This makes it a promising alternative for time-sensitive simulations where computational efficiency is a higher priority than peak accuracy.

Funding

This research did not receive any specific funding.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Data availability statement

Data underlying the results presented in this paper are not publicly available but may be obtained from the authors upon reasonable request.

Author contribution statement

The authors are the sole contributors to this work.

References

1 Gorajek L, et al., Characterization of absorption losses and transient thermo-optic effects in a high-power laser system, *Photonics* **7**, 94 (2020). https://doi.org/10.3390/photonics7040094.
2 Rall PL, Simulation and compensation of thermally influenced optical systems, *Dissertation*, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen-Nürnberg, 2023. Available from: https://open.fau.de/handle/openfau/21789.
3 Bisson JF, Suppression of the focal shift of single-mode laser with a miniature laser processing head, *J. Laser Micro/Nanoeng.* **4**(3), 170–176 (2009). https://doi.org/10.2961/jlmn.2009.03.0005.
4 Baesso ML, Shen J, Snook RD, Mode-Mismatched thermal lens determination of temperature coefficient of optical path length in soda lime glass at different wavelengths, *J. Appl. Phys.* **75**(8), 3732–3737 (1994). https://doi.org/10.1063/1.356045.
5 Glur H, Lavi R, Graf T, Reduction of thermally induced lenses in Nd:YAG with low temperatures, *IEEE J. Quantum Electron.* **40**(5), 499–504 (2004). https://doi.org/10.1109/jqe.2004.826448.
6 Koechner W, Thermal Lensing in a Nd:YAG Laser Rod, *Appl. Opt.* **9**(11), 2548–2553 (1970). https://doi.org/10.1364/ao.9.002548
7 Bonhoff T, Multiphysikalische Simulation und Kompensation thermooptischer Effekte in Optiken für Laseranwendungen (in German), *Dissertation, RWTH Aachen University*, Apprimus Verlag, Aachen, 2009. https://doi.org/10.18154/RWTH-2019-06707.
8 Yu X, Ni M, Rui D, Qu Y, Zhang W, Computational method for simulation of thermal load distribution in a lithographic lens, *Appl. Opt.* **55**(15), 4186–4191 (2016). https://doi.org/10.1364/ao.55.004186.
9 Qin Y, Michalowski A, Weber R, Yang S, Graf T, Ni X, Comparison between ray-tracing and physical optics for the computation of light absorption in capillaries – the influence of diffraction and interference, *Opt. Express.* **24**, 26606–26617 (2012). https://doi.org/10.1364/oe.20.026606.
10 Lyu C, Zhan R, STOP model development and analysis of an optical collimation system for a tactical high-energy laser weapon, *Appl. Opt.* **60**(13), 3596–3603 (2021). https://doi.org/10.1364/ao.419554.
11 Genberg VL, Michels G, Bisson G, Obtaining finite element thermal loads from fluence maps and voxels, in *Optomechanical Engineering 2019*, edited by Doyle KB, Ellis JD (SPIE, 2019), paper 111000A. https://doi.org/10.1117/12.2531176.
12 Cosson B, Akué Asséko AC, Lagardère M, Dauphin M, 3D modeling of thermoplastic composites laser welding process – a ray tracing method coupled with finite element method, *Optics Laser Technol.* **119**, 105585 (2019). https://doi.org/10.1016/j.optlastec.2019.105585.
13 Dorussen BJA, Geers MGD, Remmers JJC, An efficient ray tracing methodology for the numerical analysis of powder bed additive manufacturing processes, *Add. Manuf.* **73**, 103706 (2023). https://doi.org/10.1016/j.addma.2023.103706.
14 Rall PL, Förster D, Graf T, Pflaum C, Simulation and compensation of thermal lensing in optical systems, *Opt. Express* **30**(21), 38643–38662 (2022). https://doi.org/10.1364/oe.467813.
15 Hahn L, Eberhard P, Transient simulation of high-power dynamical-thermoelastic-optical systems, in *Optical Modeling and Performance Predictions XII*, edited by Kahan MA (SPIE, 2022), paper 1221508. https://doi.org/10.1117/12.2632227.
16 Gross H, *Handbook of optical systems: Volume 1: Fundamentals of technical optics* (Wiley, 2005). https://doi.org/10.1002/9783527699223.
17 Störkle J, Dynamic simulation and control of optical systems, *Dissertation, Schriften aus dem Institut für Technische und Numerische Mechanik der Universität Stuttgart*, Vol. **58** (Shaker Verlag, Aachen, 2018). https://doi.org/10.2370/9783844063011.
18 Nicholson DW, *Finite element analysis: Thermomechanics of Solids*, 2nd ed. (CRC Press, Boca Raton, 2008). https://doi.org/10.1201/9781420050974.
19 Fish J, Belytschko T, *A first course in finite elements*, (Wiley, 2007). https://doi.org/10.1002/9780470510858.
20 Gatej A, Modeling and compensation of thermally induced optical effects in highly loaded optical systems, *Dissertation* (Aachen, RWTH Aachen University, 2014). Available from: https://publications.rwth-aachen.de/record/444604.
21 Piehler S, Thiel C, Voss A, Abdou Ahmed M, Graf T, *Self-compensation of thermal lensing in optics for high-brightness solid-state lasers*, in *High power laser materials processing: lasers, beam delivery, diagnostics, and applications*, vol. **8239**, edited by Beyer E, Morris T, (SPIE, 2012), paper 82390Z. https://doi.org/10.1117/12.905879.
22 Hügel H, Graf T, *Laser in der Fertigung (in German)*, (Vieweg+Teubner, 2009). https://doi.org/10.1007/978-3-8348-9570-7.
23 Schrader KN, Subia SR, Myre JW, Summers KL, Ray tracing in a finite-element domain using nodal basis functions, *Appl. Opt.* **53**(24), F10–F20 (2014). https://doi.org/10.1364/ao.53.000f10.
24 Saichev AI, Woyczyński WA, Distributions in the physical and engineering sciences, in *Volume 3: Random and anomalous fractional dynamics in continuous media*, (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-319-92586-8.
25 Polyanskiy MN, Refractiveindex.info databas of optical constants, *Scientific Data* **11**(1), 94 (2024). https://doi.org/10.1038/s41597-023-02898-2.
26 Kurcsics M, Eberhard P, *Simulation results for versatile absorption modeling for transmissive optical elements using ray tracing and finite element analysis* (DaRUS, 2025). https://doi.org/10.18419/DARUS-5386.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 3
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2025055
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

3D printed freeform micro-optics meets generated Jacobian equations based inverse designs

Carlos Jimenez[1,2,*] and Alois Herkommer[1,2]

[1] Institute for Applied Optics (ITO), University of Stuttgart, 70569 Stuttgart, Germany
[2] Research Center SCoPE, University of Stuttgart, 70569 Stuttgart, Germany

Received 1 September 2025 / Accepted 10 December 2025

Abstract. In this work, we explore the use of Generated Jacobian Equations (GJEs) for the inverse design of freeform micro-optical surfaces. We demonstrate the integration of this design technique into an applied pipeline targeting 3D printed freeform micro-optics fabricated via two-photon polymerization (2PP). This work establishes a bridge between alternative design tools and cutting-edge micro-fabrication technologies, opening new opportunities in the field of engineered micro-optics. We demonstrate this via two representative examples, a round top-hat and a fringe pattern projection freeform lens.

Keywords: Inverse design, Freeform micro-optics, Two-photon polymerization.

1 Introduction

Freeform micro-optical components are central to the next generation of compact optical systems, enabling precise control over light propagation in ways that conventional optics cannot achieve. To this end, GJEs based inverse designs represent an interesting alternative to the realization of such freeform based micro-optical systems. However, such a design strategy has not been explored in combination with 2PP based micro-fabrication technologies. By combining the best from both fields, we highlight the potential for creating novel micro-optical elements tailored for specific optical transformations. In the following sections, we provide details behind the employed numerical implementation, the followed micro-fabrication process and the methodology used to extract relevant information on the smoothness associated to the computed freeform surface solutions.

2 Numerical implementation

Our implementation relies on the work presented in [1] for which an iterative algorithm is used to obtain freefrom optical surfaces via the use of a GJE. Such a design algorithm exploits the relationship between a generating function G and its unique inverse H, also referred to as a Hamilton's characteristic function. As in [1] we rely on an angular characteristic formulation to our problem, for which the source (x) and target ($\mathbf{m}$) coordinates are both expressed in terms of stereographic values. For a system consisting of a point source with N interfaces prior to the sought freeform surface, the associated Hamilton's characteristic function can be expressed as

$$H(x,\mathbf{m},u_{ff}) = u_1(x)\frac{2|x-\mathbf{m}|^2}{(1+|x|^2)(1+|\mathbf{m}|^2)} + \sum_{i=2}^{N-1}\left[u_i(x)\left(n_i - 1 + \frac{2|y_{i-1}(x)-\mathbf{m}|^2}{(1+|y_{i-1}(x)|^2)(1+|\mathbf{m}|^2)}\right)\right] + u_{ff}\left(n_N - 1 + \frac{2|y_N(x)-\mathbf{m}|^2}{(1+|y_N(x)|^2)(1+|\mathbf{m}|^2)}\right) \quad (1)$$

with u_1, u_i being the Euclidean distance functions between the N interfaces located before the freeform surface u_{ff}, $\{y_{i-1}(x)\}_{i=2}^{N-1}$ represent the set of stereographic coordinates associated to the outgoing ray-directions (I) at interfaces $i = 2,\ldots, N-1$ while $y_N(x)$ corresponds to the stereographic coordinates acting as local source ones to the last term in equation (1). Finally, u_{ff} represents the Euclidean distance between the Nth interface and the freeform surface. To implement the algorithm, we make use of an open source finite element method (FEM) solver [2] which represents two main advantages. First, it allows the use of an unstructured grid to solve the required numerical problems enabling higher flexibility in terms of the source domain shape. Second, it allows the use of local basis functions to describe the solutions in terms of polynomials within each element. Complementary to this, we employ a

* Corresponding author: jimenez@ito.uni-stuttgart.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

hybrid numerical scheme to obtain all gradient information required as part of the iterative routine. In this case, for all quantities in equation (1) excepting m and u_{ff}, we rely on automatic differentiation tools via an open source implementation of a differentiable ray-tracing framework [3]. On the other hand, for $\mathbf{m}$ and u_{ff} we make use of the polynomial preserving recovery(PPR) technique [4] which allows obtaining least squares approximations to the gradients at the domain's degrees of freedom. Importantly, by making use of this approach, we obtain a sparse matrix operator which can be directly applied on each iteration. Once a set of solution functions m and u_{ff} are obtained, these are used to reconstruct a three-dimensional freeform surface using

$$S = u_1(x)I_0(x) + \sum_{i=2}^{N-1}[u_i(x)I_{i-1}(x)] + u_{ff}I_N(x). \qquad (2)$$

Finally, as a means of comparison and to extract information on the smoothness associated with the obtained solutions, we have also chosen to use Forbes Q freeform polynomials [5–6], with these being commonly used to express freeform surfaces in terms of global orthonormal basis functions.

3 Fabrication details

The chosen freeform lens was fabricated by means of a commercially available 2PP system (Photonic Professional GT2 by Nanoscribe GmbH, Karlsruhe, Germany) in combination with a high numerical aperture objective (Plan-Apochromat 25×/1.40 Oil DIC, Zeiss). This objective was chosen given the obtained freeform lens dimensions, which has an elliptical shape with a major axis half-diameter equal to 400 µm. At the same time, a slicing and hatching distance of 100 nm were chosen, in order to have close fidelity to the original surface design and minimize discretization artifacts. The lens was fabricated directly on top of an AlGaInP visible laser diode (LD) with a central emission wavelength at 650 nm. These LDs consist of an emission surface protected by a window glass with a thickness of 400 µm. In order to align the freeform lens center to the emission point, an additional alignment step was performed. For this, a 10×/0.35 NA objective was used to find the lateral location of the LD's emission center point, which is found 700 µm below the LD's upper window interface. Consequently, two reference markers were fabricated, these being used in the subsequent freeform lens fabrication step to align the structure. In Figure 1, a schematic representation of the geometrical characteristics of an LD is provided, in addition to representative rays traced from the emission point location towards the freeform optical surface direction and indications of all involved terms directly contributing to equation (1).

4 Results

In this section we present two freeform surface design examples obtained following the methodology introduced in Section 2. For this, two different characteristic target functions have been chosen. The first one consists of an uniform target irradiance distribution defined over a circular aperture, with a diameter of 5 cm, at a target distance of 4.5 cm from the origin. The second scenario consists of a modulated cosine-squared function, given by $cos(\frac{2\pi x_t}{\beta})^2$, where x_t denotes the x cartesian coordinates at the target plane and β the desired fringe periodicity. Both distribution are defined over the same spatial extent. As in [1], we make use of a standard transformation to obtain intensity functions from both irradiance distributions. For the source, we use the model

$$I(\theta_x, \theta_y) = I_0 \exp^{-2((\theta_x/\alpha_x)^2 + (\theta_y/\alpha_y)^2)} \qquad (3)$$

with θ_x and θ_y being perpendicular emission angles associated with a Gaussian-like far-field intensity profile, while α_x and α_y represent half angles of divergence along both x and y axes. These were computed from the full-width at half-maximum angles provided in the manufacturer's datasheet (28° and 9°, respectively) using the relationship $\alpha = \frac{\theta_{fwhm}}{\sqrt{2\ln 2}}$. Following, in Figure 2 we compare the Euclidean distance function solutions u_{ff} obtained for both target functions. Additionally, we contrast the extracted convergence curves in both cases, using the same definition for J_I as given in [1].

From a visual comparison between the functions shown in (a) and (b), a clear similarity can be observed. Interestingly, the overall global function shapes and values extend over the same range. To better contrast both u_{ff} functions, on (c) we display the pointwise function difference Δu_{ff} which reveals the local distinctions between both functions, with Δu_{ff} not exceeding values beyond 1 µm. Consequently,

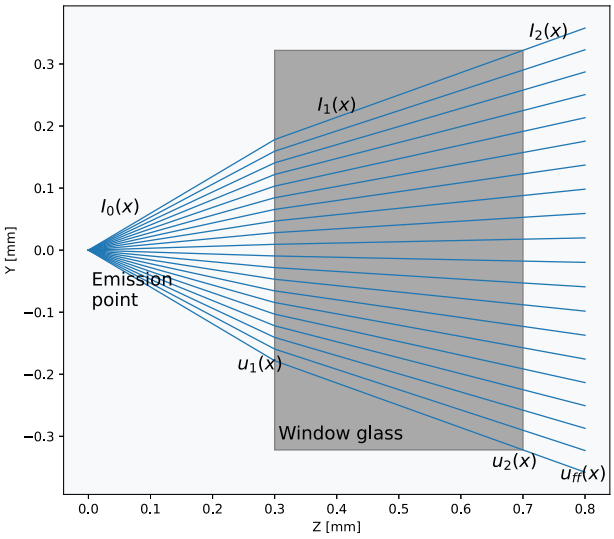


Figure 1. Schematic representation for the laser diode system considered in this case. The emission region is modeled as a point emitter, positioned at the origin. 300 µm above this point, a window glass with a thickness of 400 µm is found. All of the interfaces $u(x)$ and the associated intermediate ray-directions $I(x)$ are directly included in the Hamilton's characteristic function H shown in equation (1).

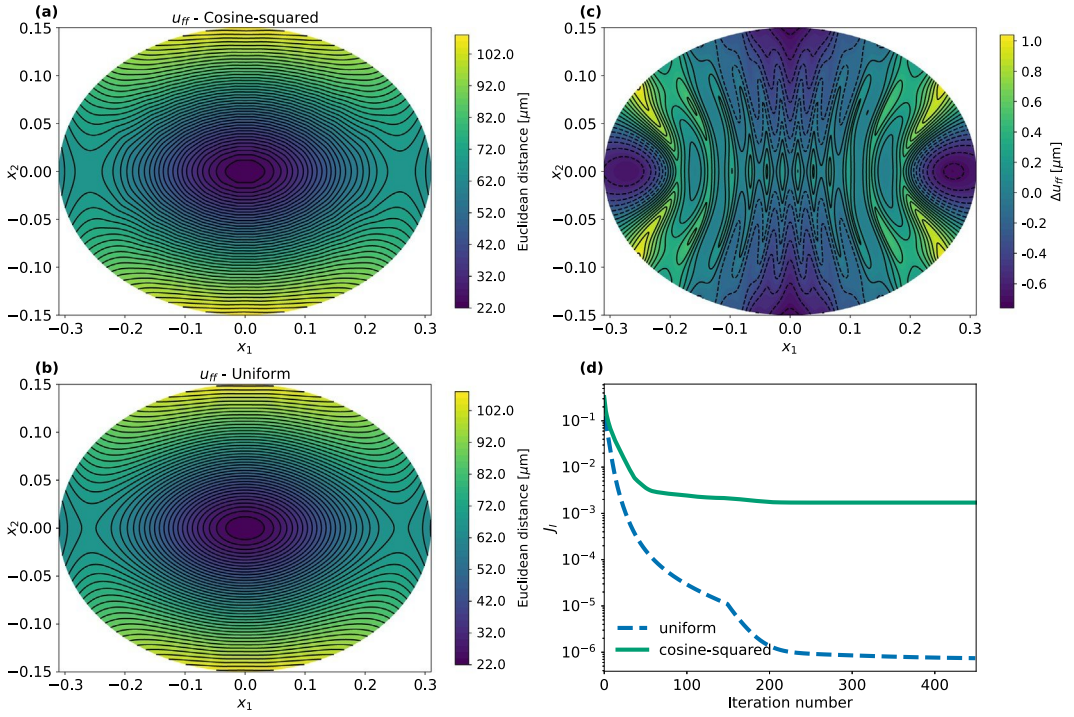


Figure 2. (a) and (b) provide the obtained u_{ff} function values after the iterative evaluations as described in Section 2. (c) Pointwise function difference Δu_{ff} between maps shown in (a) and (b) which encodes the local difference between both surface solutions. (d) provides convergence curves using the quantity J_I, according to the definition provided in [1].

we performed non-sequential ray-tracing analysis with the reconstructed three-dimensional freeform lens models via equation (2). These results are presented in Figure 3 with, the original target distributions, the obtained mapping functions **m** and the irradiance information extracted from the non-sequential analysis, being shown. As described in Section 2, our numerical implementation is based on standard local polynomial basis functions. This results in C_0-only continuous solutions, with potential detrimental effects induced in the reconstructed three dimensional surfaces. To further evaluate this, we make use of an alternative surface representation based on Forbes Q freeform polynomials as briefly introduced in Section 2. We employ least-squares approximations defined directly on the original surface domains, by projecting the solutions into basis functions characterized by indices n and m, with all terms satisfying $2m + n \leq T$ being included in the expansion. From the obtained projections, we calculate residue function maps ΔS_z by subtracting the polynomial fits from the FEM solutions. These residue maps, which are shown in Figures 4a and 4b were obtained with a T value of 54, resulting in a total of 1592 basis terms. From the chosen T, we estimate the minimum radial feature period being captured by the polynomial fit, by considering the shortest axis radius, yielding an approximate value of $\Delta r \approx \frac{200\mu\text{m}}{T/2} = 7.40\mu\text{m}$. At this minimum radial period, a clear distinction in terms of the residue surface height can be observed.

For the uniform target case, a peak residue of 67 nm was obtained, with the large majority of the surface height structure being captured by the smooth polynomial fit. Differently, for the modulated-cosine target, the obtained fit residue contains relevant spatial features with height contributions in the order of 160 nm. In other words, the unresolved surface structure consist of spatial features bounded by an extension of approximately Δr with height amplitude contributions not larger than a value of $\lambda/4$.

Additionally, power spectral density (PSD) analysis was performed from the extracted ΔS_z residue maps. These results are presented in Figure 4c as a comparison between cross-sectional views of the obtained PSDs. Both curves reveal that spatial features located at frequencies larger than 1 μm^{-1} are highly suppressed. Complementary to this, in Figure 4d we compare root-mean-square (RMS) estimates for ΔS_z as functions of the expansion index T. In this case, clear RMS reduction trends can be observed, with values of 3.38 and 7.75 nm achieved at $T = 106$, for the uniform and cosine-squared targets, respectively. Finally, the freeform lens corresponding to the cosine-squared target pattern was fabricated following the methodology described in Section 3. The obtained freeform lens was characterized using a white-light interferometer(WLI, Zygo, Nexview NX2, 10X, Δxy: 750 nm) for which the recorded intensity map is presented in Figure 5a.

As performed for the design surface case, we make use of Forbes Q freeform polynomials ($T = 54$) to fit the measured surface height profile. In (b), a portion of the obtained height fit residue map ΔS_z with contributions bounded by a value of λ is presented. From the full extracted ΔS_z, a total residue RMS of 110 nm, in addition to a peak-to-valley value of 4.67 μm were estimated, which compare to values of 37 and 263 nm for the designed surface case at the same T value, respectively. These parameter differences can be

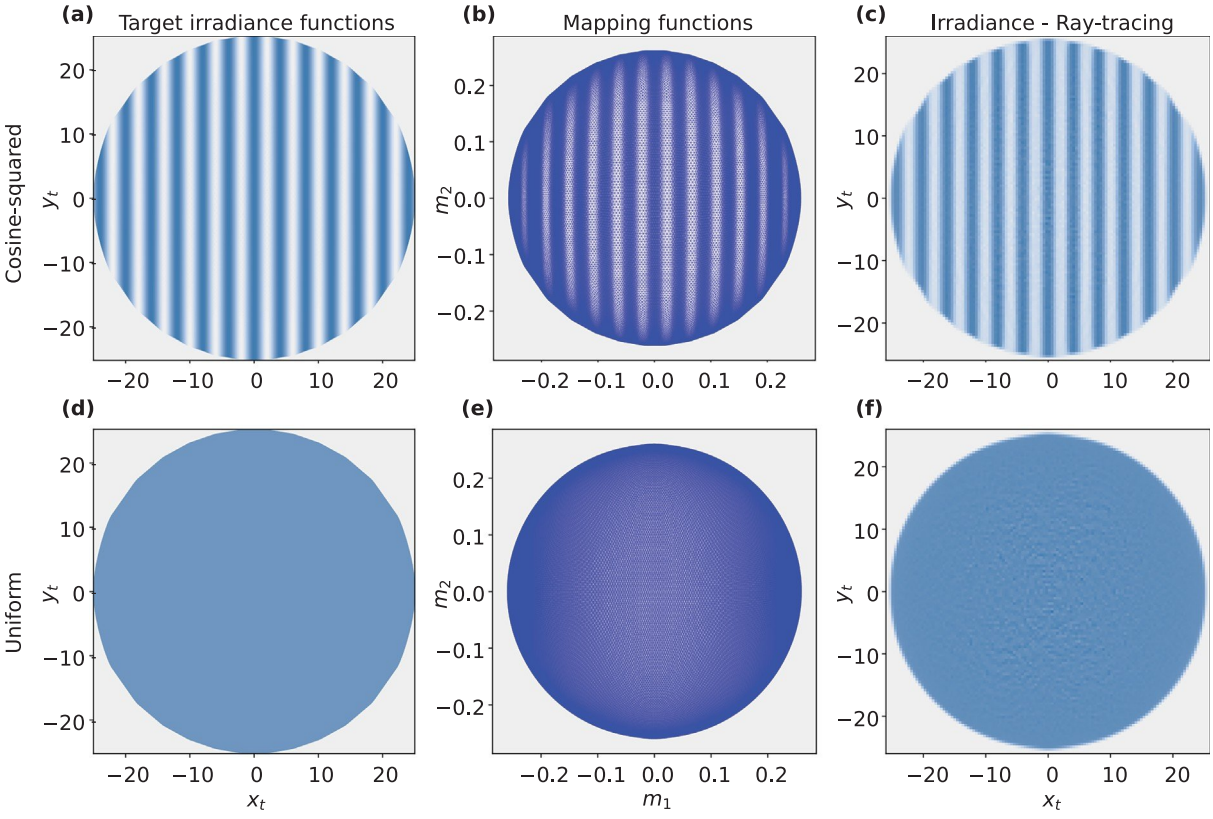


Figure 3. (a), (b) and (c) show the target irradiance, obtained mapping functions m and extracted non-sequential irradiance distribution, respectively, associated with the freeform surface generated for the cosine-squared case (u_{ff} shown in Fig. 2a). Similarly, in (d), (e) and (f) the results associated with the uniform target case are presented.

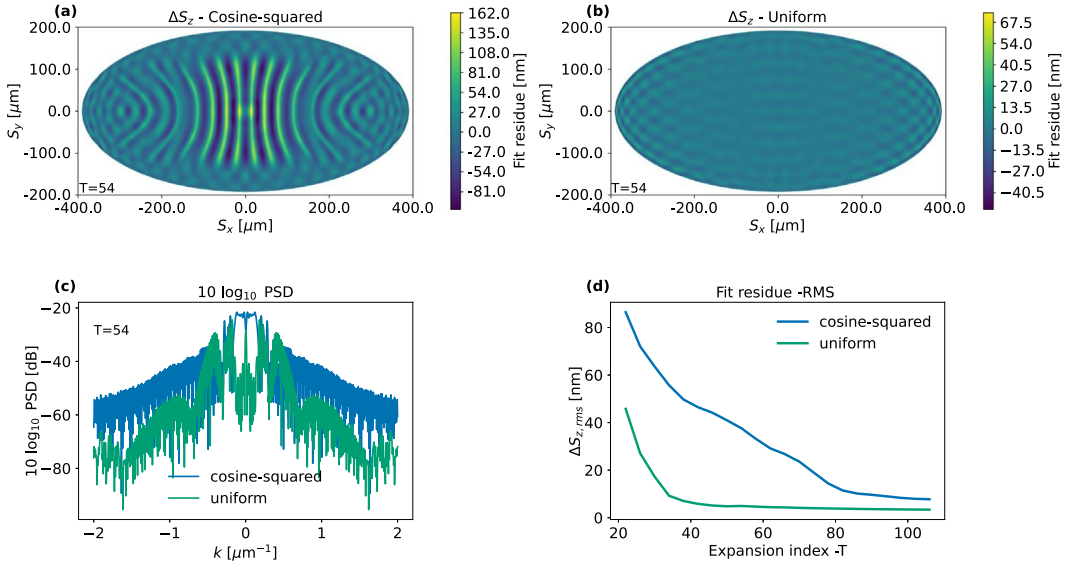


Figure 4. (a) and (b) present fit residue maps ΔS_z obtained after using least squares projections based on Forbes Q freeform polynomials to represent both FEM solutions. (c) presents power spectral density (PSD) cross-sections extracted from both residue maps shown in (a) and (b). (d) provides a comparison between ΔS_z RMS curves as a function of the expansion index T.

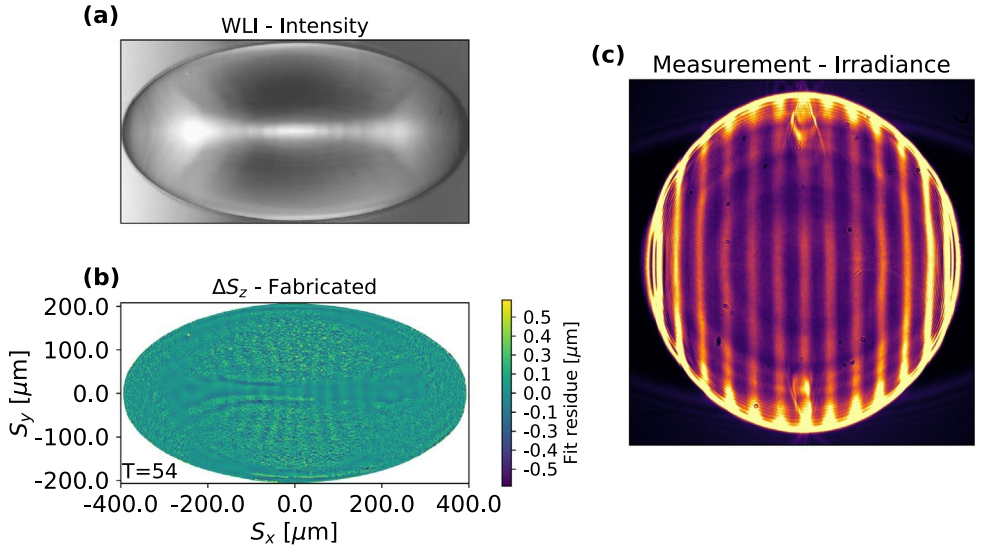


Figure 5. (a) Intensity record by employed white-light interferometer system. (b) Surface residue fit ΔS_z from extracted topography via WLI system. Only features with a magnitude bounded by λ are presented. (c) Recorded irradiance distribution from freeform lens fabricated directly on top of LD system in accordance to the process described in Section 3.

attributed to the utilized fabrication parameters and development process, with material shrinkage representing a significant portion of these contribution. Following, we measured the irradiance distribution generated by the manufactured freeform lens. This was recorded with a CMOS camera in close proximity to the LD sample, for which the obtained irradiance map is presented in Figure 5c. From the recorded map, a distorted fringe pattern in comparison to the target one can be recognized, with higher intensity values obtained towards the domain's boundary. Such a distortion is caused by the employed α_x and α_y values which were directly taken from the manufacturer datasheet and were not extracted from intensity measurements as part of the followed process. Nevertheless, despite all mentioned factors, the pattern presented in (c) still preserves most of the fringe-like structure from the intended target distribution. From the Euclidean distance function difference Δu_{ff} shown in Figure 2c with an RMS value of 345 nm and the results presented above in Figure 5, it can be concluded that the generated freeform surface design is well resolved by the employed 2PP fabrication process. Based on these results, we believe that the presented design and fabrication pipeline represents potential advantages for the realization of compact micro-optical freeform lens systems, benefiting from the flexibility offered by both, the employed GJE based design technique and the two photon polymerization based fabrication process.

5 Conclusions

In this work we have demonstrated the feasibility of combining novel inverse design algorithms with the micro fabrication capabilities offered by two photon polymerization. The presented integrated engineering pipeline benefits from the developments in both fields and serves as an important bridge between advanced design techniques and cutting edge micro fabrication technologies, enabling in this form, the further development of complex freeform surface based micro-optical systems.

Funding

This work was funded by the German Research Foundation (DFG) – project number: 537519988.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Author contribution statement

All authors contributed equally to this work.

References

1 Romijn LB, ten Thije Boonkkamp JHM, Anthonissen MJH, IJzerman WL, An iterative least-squares method for generated Jacobian equations in freeform optical design, *SIAM J. Sci. Comput.* **43**, B298 (2021). https://doi.org/10.1137/20M1338940.

2 Baratta IA, Dean JP, Dokken JS, Habera M, Hale JS, Richardson CN, Rognes ME, Scroggs MW, Sime N, Wells GN. *Dolfinx: The next generation FEniCS problem solving environment*. Zenodo; 2023. https://doi.org/10.5281/zenodo.10447666.

3 Baratta IA, Dean JP, Dokken JS, Habera M, Hale JS, Richardson CN, Rognes ME, Scroggs MW, Sime N,

Harrison K, *Optiland*; 2025. https://github.com/HarrisonKramer/optiland.

4 Naga A, Zhang Z, The polynomial-preserving recovery for higher order finite element methods in 2d and 3d, *Discrete Contin. Dyn. Syste. B* **5**, 769 (2005). https://doi.org/10.3934/dcdsb.2005.5.769.

5 Forbes GW, Characterizing the shape of freeform optics, *Opt. Express* **20**, 2483 (2012). https://doi.org/10.1364/OE.20.002483.

6 Forbes GW, Fitting freeform shapes with orthogonal bases, *Opt. Express* **21**, 19061 (2013). https://doi.org/10.1364/OE.21.019061.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 30
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026025
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Quasi-powers and primary aberrations of thin lenses in contact

Florian Bociort*

Department of Imaging Physics, Faculty of Applied Sciences, TU Delft, 2628CJ Delft, Netherlands

Received 28 January 2026 / Accepted 14 March 2026

Abstract. This paper introduces a novel framework for analysing the aberrations of thin lenses, based on the concept of surface quasi-power. Using these surface variables, remarkably simple expressions have been derived for all primary aberrations of systems of thin lenses in contact. Apart from a constant term, primary aberrations become essentially sums of powers of the new variables. When the emphasis is on qualitative properties rather than on quantitative ones, then even in complex optical systems groups of lenses can be modelled as thin lenses in contact. Especially for spherical aberration, the simplicity of the new formalism helps explaining significant properties of the lens design landscape.

Keywords: Lens design, Aberration theory, Geometrical optics, Thin lenses.

1 Introduction

Thin-lens aberration theory is a foundational element of optical design, covered extensively in standard textbooks [1, 2]. The assertion that this venerable theory still has potential for significant new insights may therefore surprise many lens designers. By introducing for each lens surface a new variable, the quasi-power, we derive primary aberration expressions for multi-lens systems that are significantly simpler than traditional formulations. Simplicity facilitates insight, and the novel formalism can provide clear explanations for both established – but perhaps insufficiently understood – and recent findings.

In Section 2 we derive the new expression for the 3rd-order spherical aberration and show how the definition of "quasi-powers" results naturally from the goal of simplifying the formalism. Section 3 is dedicated to the new type of surface variable, the quasi-power. In Section 4 other aberrations are discussed, and in Section 4.3 it is shown that all Seidel aberrations of thin lenses follow the same remarkably simple polynomial pattern when expressed in terms of quasi-powers. Section 5 provides several examples, one of which sheds light on a fundamental question in lens design – namely, why the lens design landscape exhibits such a large number of local minima.

2 Spherical aberration

Consider a rotationally symmetric lens group consisting of L thin lenses, all with the same refractive index n, in air and in contact with each other, i.e. all axial distances between surfaces within the group are set to zero. The L thin lenses in contact either form a separate optical system or are part of a larger system. In this section we derive a new expression for the 3rd-order spherical aberration of this group of thin lenses.

2.1 Framework

The starting point of the derivation is the well-known formula for total Seidel spherical aberration S computed, using the heights h_k and angles u_k of the paraxially traced marginal ray, as a sum of the surface contributions S_k of the $2L$ surfaces [1]

$$S = \sum_{k=1}^{2L} S_k = \sum_{k=1}^{2L} \left[-A_k^2 h_k \left(\frac{u_{k+1}}{n_{k+1}} - \frac{u_k}{n_k} \right) + 8 G_k h_k^4 (n_{k+1} - n_k) \right] \tag{1}$$

Here, the refraction invariant A_k is given by

$$A_k = n_k (h_k c_k + u_k) \tag{2}$$

The angles u_k are related to the surface powers P_k and curvatures c_k by the paraxial refraction formula

$$n_{k+1} u_{k+1} = n_k u_k - h_k P_k = n_k u_k - h_k c_k (n_{k+1} - n_k). \tag{3}$$

If surface k is aspheric and is described as a spherical surface plus a polynomial, then G_k is the fourth-order radial coefficient appearing in the polynomial.

In the above formulas the angle u_k and the refractive index n_k are those before refraction at surface k, whereas the index $k+1$ denotes the corresponding values after refrac-

* Corresponding author: f.bociort@tudelft.nl

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

tion (see Fig. 1). For each lens m, with $m = 1...L$, the first surface has index $k = 2m-1$ and the second surface has $k = 2m$. Outside of the lenses, the refractive indices are $n_{2m-1} = 1$, and inside the lenses we have $n_{2m} = n$. The marginal ray angles before the first and after the last lens of the thin lens group will be denoted by $u_1 = \alpha$ and $u_{2L+1} = \beta$ respectively.

2.2 Derivation of the simple spherical aberration expression

Readers primarily interested in the results may skip directly to Section 2.3.

To obtain S expressed entirely in terms of the angles u_k, we write A_k as

$$A_k = \frac{u_{k+1} - u_k}{n_{k+1}^{-1} - n_k^{-1}}. \tag{4}$$

If equation (3) is used to eliminate u_{k+1} in equation (4), then after simple algebra the refraction invariant A_k becomes the one given by the more familiar equation (2).

To shorten the formulas, consider first only spherical surfaces, i.e. we have $G_k = 0$ for all surfaces (the G_k terms will be included later). Because in the thin-lens approximation all distances between the surfaces shown in Figure 1 are considered to be zero, the height of the marginal ray does not change inside this group and we have $h_k = h = const$ for all $k = 1...2L$. For thin lenses in contact, the contributions S_k for odd and even surface numbers result from equations (1) and (4) after simple algebra as

$$\begin{gathered} S_{2m-1} = t(u_{2m-1} - u_{2m})^2(nu_{2m-1} - u_{2m}) \\ S_{2m} = t(u_{2m} - u_{2m+1})^2(u_{2m} - nu_{2m+1}) \\ t = \tfrac{hn}{(n-1)^2}. \end{gathered} \tag{5}$$

These two surface contributions can be combined as

$$S_k = (-1)^{k-1} t(\tilde{u}_k - u_{2m})^2(n\tilde{u}_k - u_{2m}), \tag{6}$$

where for odd-index angles we introduce the notation

$$\tilde{u}_k = \begin{cases} u_{2m-1} & \text{for } k = 2m-1 \\ u_{2m+1} & \text{for } k = 2m \end{cases} \tag{7}$$

To derive the new spherical aberration formula for thin lenses in contact, we start by observing that the surface contributions S_k in equation (6) contain an almost perfect cube of the angle difference $\tilde{u}_k - u_{2m}$, the obstacle being the refractive index appearing in the last bracket. As a first step, we show below that the surface contributions can be written as a perfect cube plus correction terms that give the departure from the cube, such that most of the correction terms cancel each other out during summation over surfaces. To facilitate the construction of these expressions, we also introduce temporary variables μ_k, that in air are equal to the corresponding angle u_k, and inside the lens differ from the angle by a factor q that needs to be determined,

$$\mu_{2m-1} = u_{2m-1},\ \mu_{2m} = qu_{2m}. \tag{8}$$

We see from equations (6) and (7) that if we expand S_k we obtain four terms of total power 3 in the angles u, e.g. for the first surface the result will contain terms corresponding to $u_1^3, u_1^2u_2, u_1u_2^2, u_2^3$. In the new variables given by equation (8) S_k will also contain four terms, with coefficients $c_{S1}, c_{S2}, c_{S3}, c_{S4}$ that need to be determined,

$$S_k = (-1)^{k-1}\left[c_{S1}(\mu_{2m} - \tilde{\mu}_k)^3 + c_{S2}\tilde{\mu}_k^3 + c_{S3}\mu_{2m}^3 + c_{S4}\tilde{\mu}_k^2\mu_{2m}\right], \tag{9}$$

where for odd indices $\tilde{\mu}_k$ is defined in the same way as $\tilde{u}_k$ in equation (7). Because the goal of this approach is to construct an expression containing a perfect cube, we use the perfect cube in equation (9) instead of the term $\mu_{2m}^2\tilde{\mu}_k$. Consider first the odd surfaces, for which we have $(-1)^{2m-2} = 1$.

By substituting equations (8) and (7) into equation (9) we obtain an expression for S_{2m-1} in terms of the angles u that must be equal to that of S_{2m-1} in the first of equation (5). By subtracting the two equivalent expressions for S_{2m-1}, and by using e.g. $m = 1$, we obtain after elementary algebra the zero polynomial

$$\begin{aligned} 0 &= t(u_1 - u_2)^2(nu_1 - u_2) - c_{S1}(qu_2 - u_1)^3 - c_{S2}u_1^3 \\ &\quad - c_{S3}q^3u_2^3 - c_{S4}qu_2u_1^2 \\ &= u_1^3(c_{S1} - c_{S2} + tn) - u_2^3\left(q^3(c_{S1} + c_{S3}) + t\right) \\ &\quad - u_2u_1^2(q(3c_{S1} + c_{S4}) + (2n+1)t) \\ &\quad + u_2^2u_1\left(3c_{S1}q^2 + (n+2)t\right) \end{aligned} \tag{10}$$

By annulling the coefficients of the four terms of total power 3 in the angles u we obtain four equations with five unknowns. We can freely choose one of these unknowns, which are the four coefficients c_S and q. To simplify the construction of the new expression in equation (9), we choose $c_{S4} = 0$. After substituting for t the value given in equation (5) we obtain the system of equations

$$\begin{aligned} c_{S1} - c_{S2} + \tfrac{hn^2}{(n-1)^2} &= 0 \\ 3c_{S1}q + \tfrac{hn(2n+1)}{(n-1)^2} &= 0 \\ 3c_{S1}q^2 + \tfrac{hn(n+2)}{(n-1)^2} &= 0 \\ q^3(c_{S1} + c_{S3}) + \tfrac{hn}{(n-1)^2} &= 0 \end{aligned} \tag{11}$$

The 2nd and 3rd equation give, after moving one term to the other side, followed by division

$$q = \frac{n+2}{2n+1}. \tag{12}$$

Note that if we set $n = 1$ in equation (12) we obtain $q = 1$. Because we have $n_{2m-1} = 1$, and $n_{2m} = n$, equation (8) can also be written as $\mu_k = (n_k + 2)u_k/(2n_k + 1)$.

From the 2nd and first equation in equation (11) we obtain immediately

$$\begin{aligned} c_{S1} &= -\tfrac{hn(2n+1)^2}{3(n-1)^2(n+2)} \\ c_{S2} &= -\tfrac{hn}{3(n+2)}. \end{aligned} \tag{13}$$

The coefficient c_{S3} results from the last of equation (11) but, as will be seen below, it is not important.

Because for S_{2m} we have in equation (9) the factor $(-1)^{2m-1} = -1$, the coefficients c_{S1}, c_{S2}, c_{S3} appear with a sign opposite to that in S_{2m-1}. Using equations (7) and (9) we can then write

$$\begin{aligned} S_{2m-1} &= c_{S1}(\mu_{2m} - \mu_{2m-1})^3 + c_{S2}\mu_{2m-1}^3 + c_{S3}\mu_{2m}^3 \\ S_{2m} &= c_{S1}(\mu_{2m+1} - \mu_{2m})^3 - c_{S2}\mu_{2m+1}^3 - c_{S3}\mu_{2m}^3 \end{aligned} \quad (14)$$

Note that in the first term of S_{2m} we have changed the order of μ_{2m+1} and μ_{2m} compared to equation (9), therefore c_{S1} appears here with the plus sign.

We now replace the temporary variables μ_k by the new variables $z_k = (\mu_{k+1} - \mu_k)/(\beta - \alpha)$. The total spherical aberration S is then obtained in terms of the new variables z_k by summing up the odd and even surface contributions in equation (14) over all lenses, with $m = 1 \ldots L$. Note that in this sum all terms with coefficients c_{S3} cancel each other out, as well as the terms with coefficients c_{S2}, excepting those with $\mu_1^3 = u_1^3 = \alpha^3$, and $\mu_{2L+1}^3 = u_{2L+1}^3 = \beta^3$.

2.3 The simple spherical aberration expression

By using equations (8) and (12) the new variables can be rewritten as

$$z_k = \frac{1}{\beta - \alpha}\left(\frac{n_{k+1} + 2}{2n_{k+1} + 1}u_{k+1} - \frac{n_k + 2}{2n_k + 1}u_k\right) \quad (15)$$

and, in the absence of aspheres, spherical aberration becomes

$$S = c_{S1}(\beta - \alpha)^3 \sum_{k=1}^{2L} z_k^3 + c_{S2}(\alpha^3 - \beta^3), \quad (16)$$

where c_{S1} and c_{S2} are given by equation (13).

The marginal ray angles α and β, before and after the thin-lens group, are related to the total power K of the group via

$$\beta = \alpha - hK. \quad (17)$$

Including in equation (16) the aspheric contributions appearing in equation (1) is straightforward because, apart from an alternating sign, the refractive index difference is the same for all surfaces. Using equation (17), we obtain for the spherical aberration of the thin-lens group the final expression

$$\begin{aligned} S = \frac{h^4K^3n}{3(n+2)}\left[\frac{(2n+1)^2}{(n-1)^2}\sum_{k=1}^{2L} {z_k}^3 - 3\bar{\alpha}^2 + 3\bar{\alpha} - 1\right] \\ + 8h^4(n-1)\sum_{k=1}^{2L}(-1)^{k-1}G_k, \end{aligned} \quad (18)$$

where we have used the abbreviation

$$\bar{\alpha} = \frac{\alpha}{hK}. \quad (19)$$

If the group of L thin lenses is part of a larger system, then the 3rd-order spherical aberration of the entire system is

$$S_{tot} = S + S^*, \quad (20)$$

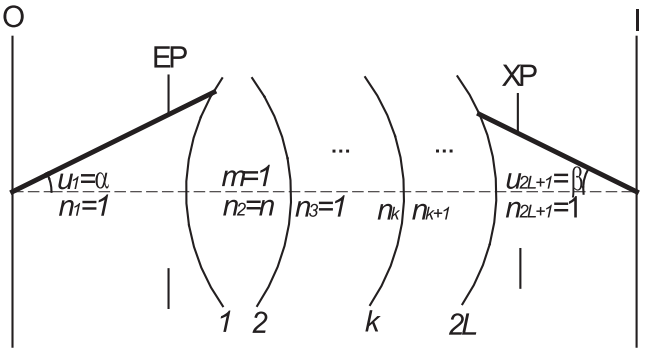


Figure 1. The paraxially traced marginal ray (thick line) has before the first surface of the group of L lenses the angle $u_1 = \alpha$ with the optical axis and after the last surface the angle $u_{2L+1} = \beta$. The refractive index before surface k is n_k, after the surface it is n_{k+1} .The surface numbering for the marginal ray angles u_k is the same. Inside each lens (i.e. for even k values) the refractive index is n, as shown here for the first lens ($m=1$) with surfaces 1 and 2. Outside the lenses we have $n_1 = n_3 = \ldots = n_{2L+1} = 1$. In this figure, the L lenses of interest form the entire optical system, but the same notation is used when these lenses are part of a larger system. The planes of the object, image, entrance pupil and exit pupil are denoted by O, I, EP and XP respectively. In the thin-lens approximation, all axial distances between surfaces 1 and $2L$ will be set equal to zero in the aberration formulas.

where S^* denotes the contribution of the other lenses in the larger system.

3 Quasi-powers and surface powers

The new variables z_k defined by equation (15) are essential for simplifying the entire thin-lens formalism and provide a new framework for analysing aberrations. In this section we discuss their properties as well as their relationship with the surface powers and curvatures.

When the angles u_k inside the lens group are known, the corresponding z_k values can be determined using equation (15). However, as shown in the Examples section, it is sometimes possible to determine the z_k values first. Then, the surface curvatures c_k result from the z_k values as follows. We find from equations (15) and (17)

$$\begin{aligned} \mu_{k+1} &= \mu_k + (\beta - \alpha)z_k = \mu_k - hKz_k \\ &= \mu_{k-1} - hK(z_k + z_{k-1}) = \ldots = \alpha - hK\sum_{i=1}^{k} z_i. \end{aligned} \quad (21)$$

For $k = 2L$ we expect to have, because of equation (17), $u_{2L+1} = \beta = \alpha - hK$. The denominator in equation (15) was therefore chosen such that the sum of all z-variables is normalized to unity,

$$\sum_{k=1}^{2L} z_k = 1. \quad (22)$$

The surface powers result from equations (3) and (8):

$$\begin{aligned} P_k &= (n_k u_k - n_{k+1} u_{k+1})/h \\ P_{2m} &= \left(n\mu_{2m}/q - \mu_{2m+1}\right)/h = (n\mu_{2m}/q - \mu_{2m} + hKz_{2m})/h \\ &= Kz_{2m} + (n/q-1)\mu_{2m}/h \\ P_{2m-1} &= (\mu_{2m-1} - n\mu_{2m}/q)/h = (\mu_{2m} + hKz_{2m-1} - n\mu_{2m}/q)/h \\ &= Kz_{2m-1} - (n/q-1)\mu_{2m}/h \end{aligned} \tag{23}$$

equation (12) gives

$$\frac{n}{q} - 1 = \frac{2(n^2-1)}{n+2} \tag{24}$$

and from equations (21) and (19) we obtain

$$\mu_{2m}/h = -K\left(\sum_{i=1}^{2m-1} z_i - \bar{\alpha}\right) \tag{25}$$

The surface powers are then

$$\begin{aligned} P_{2m-1} &= K\left[z_{2m-1} + \tfrac{2\left(n^2-1\right)}{(n+2)}\left(\sum_i^{2m-1} z_i - \bar{\alpha}\right)\right] \\ P_{2m} &= K\left[z_{2m} - \tfrac{2\left(n^2-1\right)}{(n+2)}\left(\sum_i^{2m-1} z_i - \bar{\alpha}\right)\right] \end{aligned} \tag{26}$$

and the corresponding surface curvatures result then from equation (3) as

$$c_k = P_k/(n_{k+1} - n_k). \tag{27}$$

It follows from equation (26) that the power of lens m, $\bar{P}_m = P_{2m-1} + P_{2m}$ is simply

$$\bar{P}_m = K(z_{2m-1} + z_{2m}). \tag{28}$$

Note from equations (26) and (28) that, for each lens surface, z_k has a term proportional to the surface power, plus a correction term that is exactly compensated by a correction term of equal magnitude and opposite sign coming from the other surface of the same lens. The power of each lens is then proportional to the sum of the z-values of its two surfaces. Because they can be viewed intuitively as power-like quantities, we refer to the variables z_k as "quasi-powers".

When the group of thin lenses forms the entire system, the position of the object s_o and that of the image s_i with respect to the lens group and the transverse magnification M_T are determined by the angles $u_1 = \alpha$ and $u_{2L+1} = \beta$,

$$s_o = -h/\alpha,\ s_i = -h/\beta,\ M_T = s_i/s_o = \alpha/\beta. \tag{29}$$

Using equations (29), equation (17) becomes after division by h the well-known Lensmaker's Formula (1) $/s_i - 1/s_o = K$.

4 Other aberrations

4.1 Axial colour, astigmatism and Petzval sum

The simple relation (28) between the power $\bar{P}_m$ of a lens and the two quasi-powers leads immediately to the expression for the total axial colour of the thin lens group expressed in terms of quasi-powers. As well known, the axial colour contribution of each lens in the group is proportional to its lens power [1]. The total axial colour of the thin lens group is then the sum of the contributions of the individual lenses

$$A = -h^2 K \sum\nolimits_{m=1}^{L} V_m^{-1}(z_{2m-1} + z_{2m}), \tag{30}$$

where V_m is the Abbe number for lens m. When equation (30) is used, the Abbe numbers can be different, but the refractive index n needs to be the same for all lenses in the thin group.

For the thin lens group, several primary aberrations do not depend on the quasi-powers and have well-known expressions [1]. If the aperture stop is placed at the thin lens group, the 3rd-order distortion and lateral colour vanish. The total astigmatism T_1 and Petzval sum T_1' of the lens group are

$$T_1 = H^2 K \tag{31}$$

and

$$T'_1 = T_1/n, \tag{32}$$

where H is the Lagrange invariant of the entire system.

The last Seidel aberration that remains to be expressed in terms of the quasi-powers is coma. The same approach as for spherical aberration can be used to obtain a simple formula for the coma contribution of thin lenses in contact. The Seidel sum for the 3rd-order coma is [1]

$$C = \sum_{k=1}^{2L} C_k = -\sum_{k=1}^{2L} \bar{A}_k A_k h_k \left(\frac{u_{k+1}}{n_{k+1}} - \frac{u_k}{n_k}\right) \tag{33}$$

When the aperture stop is placed at the group of thin lenses, the chief-ray height at the group is zero. The paraxial refraction invariant $\bar{A}_k$ for the chief ray (which has a formula similar to equation (2), but using the chief-ray height and angle) is then given by the chief ray angle $\bar{u}$ before and after the group, $\bar{A}_k = \bar{u}_1 = \bar{u}_{2L+1} = \bar{u}$.

4.2 Derivation of the simple coma expression

Readers primarily interested in the results may skip directly to Section 4.3.

For odd and even surfaces we have the surface contributions

$$\begin{aligned} C_{2m-1} &= t'(u_{2m-1} - u_{2m})(nu_{2m-1} - u_{2m}) \\ C_{2m} &= t'(u_{2m+1} - u_{2m})(u_{2m} - nu_{2m+1}) \\ t' &= \frac{h\bar{u}}{n-1} \end{aligned} \tag{34}$$

or, using equation (7)

$$C_k = (-1)^{k-1} t'(\tilde{u}_k - u_{2m})(n\tilde{u}_k - u_{2m}). \tag{35}$$

When expanded, the coma surface coefficient contains three quadratic terms in the angles u. We look for new forms of equation (34) as a perfect square plus two correction terms. For odd surfaces we look for a form

$$C_{2m-1} = c_{C1}(\mu_{2m} - \mu_{2m-1})^2 + c_{C2}\mu_{2m-1}^2 + c_{C3}\mu_{2m}^2. \tag{36}$$

If in equation (36) we substitute equation (8) and expand the square, the resulting expression must be equal to the expanded form of the first of equation (34). Subtracting these two expressions gives for $m = 1$ the zero polynomial

$$\begin{aligned} 0 &= t'(u_1 - u_2)(nu_1 - u_2) - c_{C1}(qu_2 - u_1)^2 \\ &\quad - c_{C2}u_1^2 - c_{C3}q^2u_2^2 \\ &= u_2^2\left(-c_{C1}q^2 - c_{C3}q^2 + t'\right) - \left(u_1^2(c_{C1} + c_{C2} - nt')\right) \\ &\quad + u_2u_1(2c_{C1}q - (n+1)t') \end{aligned} \tag{37}$$

After substituting q and t' using equations (12) and (34) we obtain by annulling the three coefficients of the quadratic terms the system of equations

$$\begin{aligned} c_{C1} + c_{C2} &= \tfrac{h n \bar{u}}{n-1} \\ \tfrac{2c_{C1}(n+2)}{2n+1} &= \tfrac{h\bar{u}(n+1)}{n-1} \\ \tfrac{h\bar{u}}{n-1} &= \tfrac{(n+2)^2(c_{C1}+c_{C3})}{(2n+1)^2} \end{aligned} \tag{38}$$

that gives for c_{C1} and c_{C2} (c_{C3} will not be needed)

$$\begin{aligned} c_{C1} &= h\bar{u}\tfrac{(n+1)(2n+1)}{2(n-1)(n+2)}, \\ c_{C2} &= \tfrac{h\bar{u}}{2(n+2)}. \end{aligned} \tag{39}$$

Because of the factor $(-1)^{k-1}$ in equation (35), for the even surfaces the three coefficients are exactly the opposite of those in equation (36) and we have

$$C_{2m} = -c_{C1}\left(\mu_{2m+1} - \mu_{2m}\right)^2 - c_{C2}\mu_{2m+1}^2 - c_{C3}\mu_{2m}^2. \tag{40}$$

When we sum up the surface contributions (36) and (40) over all lenses, all c_{C3} terms cancel each other out, as well as the c_{C2} terms, excepting those with $\mu_1^2 = u_1^2 = \alpha^2$, and $\mu_{2L+1}^2 = u_{2L+1}^2 = \beta^2$.

4.3 Polynomial pattern

With the new variables defined by equation (15), the quadratic terms appear in the coma expression C with alternating signs,

$$C = c_{C1}(\beta - \alpha)^2 \sum_{k=1}^{2L} (-1)^{k-1} z_k^2 + c_{C2}\left(\alpha^2 - \beta^2\right), \tag{41}$$

where c_{C1} and c_{C2} are given by equation (39). Alternatively, the coma contribution of the thin lens group, with the stop at the lens group, can be written as

$$C = \frac{-\bar{u}h^3K^2}{2(n+2)}\left[\frac{(n+1)(2n+1)}{n-1}\sum_{k=1}^{2L}(-1)^k {z_k}^2 - 2\bar{\alpha} + 1\right]. \tag{42}$$

For an arbitrary stop position, the contribution of the thin lens group to the primary aberrations can be computed by using the well-known stop-shift formulas [1, 2].

Note that, when all surfaces are spherical, all Seidel (monochromatic) aberration formulas have the same structure

$$T_j = a_j(\beta - \alpha)^j\left(\sum_{k=1}^{2L}(-1)^{(j+1)(k-1)} z_k^j\right) + b_j\left(\alpha^j - \beta^j\right). \tag{43}$$

The exponent of -1 was chosen such that for odd indices j all terms z_k^j have the same sign for all values of k, and that for even j the signs of z_k^j are alternating. For spherical aberration and coma we have $j = 3$ and $j = 2$, respectively, with $T_3 = S$ given by equation (16) and $T_2 = C$ given by equation (41), and the coefficients are $a_3 = c_{S1}$, $b_3 = c_{S2}$, $a_2 = c_{C1}$, $b_2 = c_{C2}$. The aberrations that have simple expressions also fit into this pattern. For astigmatism and Petzval sum (equations (31) and (32)) we have $j = 1$. The sum in equation (43) is then 1 because it becomes the constraint (22), therefore both aberrations are constant. According to equation (17), for *j=1* both factors $(\beta - \alpha)^j$ and $\alpha^j - \beta^j$ are proportional to the power K, a property that is in agreement with equation (31). The distortion T_0 is zero as expected, because for $j = 0$ the sum with alternating terms in equation (43) is $L - L = 0$ and we have $\alpha^0 - \beta^0 = 1 - 1 = 0$.

5 Examples

The thin-lens formulas for primary aberrations derived in this paper have been verified using the lens design programs CODE V and Zemax OpticStudio. For lens systems where distances between surfaces have been set to zero, the quasi-powers are computed using paraxial ray-tracing data and equation (15). Then, as shown in the supplementary data, implementing the new aberration formulas in the macro languages leads to numerical values that are identical with the corresponding coefficients listed by these programs (see the link in the Data availability statement).

5.1 Equal quasi-powers

In the examples below we consider only systems having spherical surfaces. We first focus on the spherical aberration S. We denote the sum of cubes that appears in equations (16) and (18) by $s = \sum_{k=1}^{2L} {z_k}^3$. If for a system consisting of L lenses the quasi-powers z_k are considered to be variables that satisfy the constraint (22), then it can be seen that, because of the perfect symmetry, a system having equal quasi-powers, i.e.

$$z_k = 1/(2L), \tag{44}$$

for all k-values, must be an extremum of s. By slightly perturbing this system by a small quantity ε, $z_1 = \frac{1}{(2L)} + \varepsilon$ and (to satisfy the constraint) $z_2 = \frac{1}{2L} - \varepsilon$, $z_k = \frac{1}{2L}$ for $k > 2$, we obtain $s(\varepsilon) = 1/(4L^2) + 3\varepsilon^2/L$ which is always larger than

$$s_{\min} = \frac{1}{(4L^2)}. \tag{45}$$

Several known results can be easily derived by starting from systems with equal quasi-powers, that are minima of s, with the minimum value $s = s_{\min}$.

For $L = 1$, we recover familiar results of traditional thin-lens theory. The system with $z_1 = z_2 = 1/2$ corresponds then to the well-known singlet with optimal bending that has minimal spherical aberration. Traditional thin-lens theory uses the magnification variable $\bar{C} = (\alpha + \beta)/(\alpha - \beta)$. Inserting in equation (18) $s_{\min} = 1/4$ and $\bar{\alpha} = (1 + \bar{C})/2$ leads to the well-known minimal spherical aberration formula [1]

$$S_{\min} = \frac{1}{4} h^4 K^3 n \left(\frac{n}{(n-1)^2} - \frac{\bar{C}^2}{n+2} \right). \tag{46}$$

For larger L, an interesting result that has a rather complex derivation in the literature follows easily from the present model. Fulcher has shown that 3rd-order spherical aberration can be corrected with thin lenses having the same power, but different bendings. In his telescope objective, four lenses with a refractive index close to $n = 1.5$ are used to achieve this goal [3]. For $L=4$ we have for all k $z_k = 1/8$ and $s_{\min} = 1/64$. With $\bar{\alpha} = 0$ (object at infinity) equation (18) leads to

$$S_{\min} = -\frac{h^4 K^3 n(2n-3)(10n-7)}{64(n-1)^2(n+2)}. \tag{47}$$

Spherical aberration vanishes for $n = 1.5$ because of the first parenthesis in the numerator. It follows from equation (28) that all four lens powers are equal, $\bar{P}_m = K/4$, despite of the fact that the four lenses have different curvatures (the surface powers resulting from equation (26) are the same as those listed in Table 1 of Ref. [3]). As shown by Shafer, Fulcher systems are good starting points for further design and lead to relaxed designs that have an axial imaging of excellent quality even at large apertures [4]. For $L = 2$, converting the four equal quasi-powers in equation (44) into curvatures using equations (26) and (27) leads to a doublet configuration that is also given as a typical example of a relaxed design (see Fig. 1 of Ref. [4]). Figure 2 shows a Fulcher quartet appearing as a lens group in a lithographic objective having only spherical surfaces and a numerical aperture of 0.56. This system is closely related to a system in [5, 6]. The similarity with the lens shapes in the blue box supports the interpretation of the lenses in the red box as essentially a Fulcher group.

5.2 Permutation symmetry for spherical aberration

In many imaging systems, including the one shown in Figure 2, we encounter groups of lenses having reasonably small thicknesses and air spaces between them. Simplified models, including the thin-lens approximation used here, rarely yield accurate quantitative results, but the deliberate neglect of distracting complexities can reveal qualitative properties that are otherwise obscured. The principal motivation behind deriving the thin-lens formulas was to provide a simplified framework for gaining insight into the properties of the lens design landscape. Because of the extensive derivations involved, detailed examples will be presented in a separate paper. Here we show an example that helps answering a fundamental question in optical system optimization: why are there so many local minima in the design landscape?

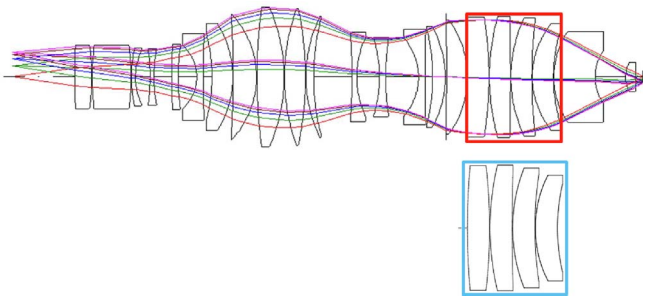


Figure 2. Red box: Fulcher group in an optimized design in which all lenses have the same material. Blue box: the shapes of the same four lenses resulting from the present thin-lens model using $z_k = 1/8$ and the marginal ray angles α and β before and after the four-lens group, extracted from the optimized design. In the drawing, the lens thicknesses, which are assumed to be zero in the calculations, are kept the same as in the red box.

The existence of certain local minima in the optimization landscape can already be explained using 3rd-order aberrations. If the surrounding landscape is not flat, higher-order aberrations only determine how deep these minima are. In an optimization landscape with specifications that make spherical aberration the most significant aberration, consider for simplicity a rough approximation of the error function, $E = S_{tot}{}^2$. Because in S given by equation (18) the quasi-powers appear in the sum of cubes s, the mathematical property of commutativity leads to permutation symmetry: if a certain set of variables z_k corresponds to a local minimum, then any permutation of these variables will have the same values of s, S, S_{tot} (given by equation (20)) and E. Any such permutation will then correspond to a different minimum, a property that increases the number of existing minima in the landscape significantly. This permutation symmetry was not visible in earlier formalisms, because of the sequential character of ray propagation (rays pass first through surface 1, then through surface 2 etc.). However, the quasi-power formalism reveals this symmetry because the sequential character of ray propagation is now absorbed in equation (26) and is therefore separated from the more important aberration properties resulting from equation (18), or, more generally, from equation (43).

Figure 3 shows an example of the effect of the permutation symmetry in S on the number of local minima. As shown previously [7], local minima in the optimization landscape surrounding the system in Figure 2 generally have localized changes in the corresponding system drawings. The red boxes in Figure 3 contain local minima for which the most significant changes occur within the group of four lenses considered in Figure 2. These systems have been obtained with CODE V, with lens curvatures as optimization variables, and distortion control added to the default error function. For this study, telecentricity was not controlled and edge thickness control inside this lens group was disabled.

In Figure 3 we consider the same group of four lenses as in Figure 2. For the upper left local minimum, the model

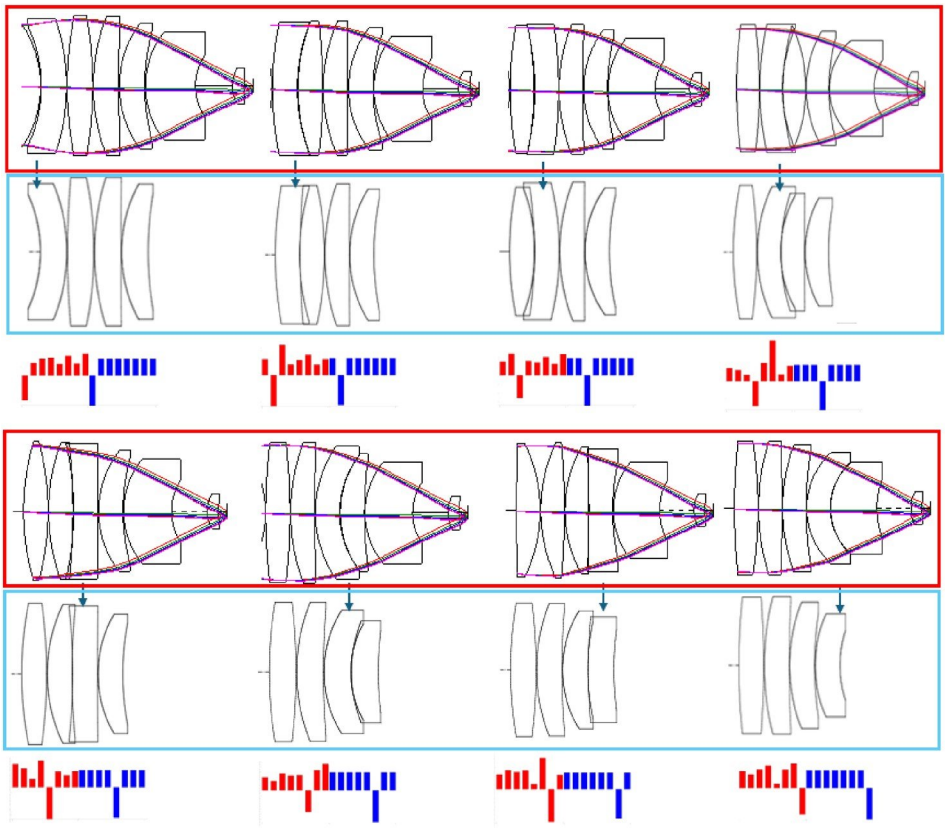


Figure 3. Eight local minima in the vicinity of the system in Figure 2 are shown in the red boxes. Only the last six lenses are shown, which include the four lenses of interest. The lenses with the most significant change compared to Figure 2 are marked with an arrow. For the four lenses of interest the blue bar charts show the z-values that result from theory, one negative z-value and seven equal positive z-values. When these z-values are translated into surface curvatures, the lenses in the blue boxes are obtained. For comparison, the red bar charts show the z-values obtained from data extracted from the optimized systems.

assigns one negative and seven identical positive z-values to the four lenses, as illustrated by the corresponding blue bar chart. (Theory – partly already developed in [8] and to be presented in detail in a separate work – predicts the existence of such minima, characterized by one negative and seven equal positive z-values, in the vicinity of a Fulcher-like group having eight equal z-values.) The permutation symmetry in S then implies the existence of seven additional local minima, in which the negative z-value of the first minimum appears at each of the other positions within the group. The permuted z-values are shown by the blue bar charts. The lenses enclosed by blue boxes are obtained by translating each of the eight permuted sets of z-values into surface curvatures using equations (26) and (27).

The systems in red boxes are candidates for the predicted local minima in the optimization landscape. The red bar charts show for these systems the z-values obtained using equation (15) and marginal-ray data from the optimized systems. Some discrepancy between the red and blue bars is expected, because the approximate error function E neglects many aberrations and because the model assumes zero-thickness lenses, whereas the optimized systems contain lenses of finite thickness. (Also, the red bars show seemingly larger discrepancies because they show relative rather than absolute differences, and because the z-values for the four surfaces of interest are significantly smaller than those of more strongly curved surfaces elsewhere in the system.) However, for the systems in red boxes the shapes of the four lenses of interest agree reasonably well with the corresponding lenses in blue boxes. This agreement supports the interpretation of these systems as the eight minima resulting from permutation symmetry.

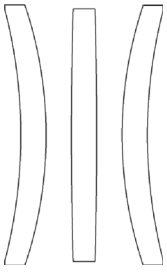


Figure 4. Aplanatic thin triplet for the infrared region ($n = 4$), with an effective focal length of 1, and transverse magnification $M_T = -1$. The zero distances between surfaces are drawn for clarity as finite. The stop is at the thin lens. All six quasi-powers are equal to 1/6. Despite of the different bendings, the three spherical lenses have the same lens power 1/3 (see equation (28)).

Table 1. The quasi-power surface contributions for spherical aberration (QPS) and coma (QPC) differ significantly from the corresponding traditional surface contributions for spherical aberration (Trad. S) and coma (Trad. C). The constant terms (Const.), which are absent (–) in the traditional approach, are also listed in the QPS and QPC columns. The values in the columns Trad. S and Trad. C are identical with the corresponding Seidel coefficients SPHA S1 and COMA S2 listed by Zemax.

k	z_k	R_k	Trad. S	QPS	Trad. C	QPC
1	0.166666	−2	0.000000	0.000579	0.000000	0.000765
2	0.166666	−1.636363	−0.000514	0.000579	0.000408	−0.000765
3	0.166666	18	0.000514	0.000579	0.000816	0.000765
4	0.166666	−18	0.000514	0.000579	−0.000816	−0.000765
5	0.166666	1.636363	−0.000514	0.000579	−0.000408	0.000765
6	0.166666	2	0.000000	0.000579	0.000000	−0.000765
Const.	–	–	–	−0.003472	–	0.000000
Total	1		0.000000	0.000000	0.000000	0.000000

5.3 Aplanatic correction

In the special case of Fulcher-like thin-lens systems it can be easily seen that the 3rd-order coma formula (equation (42)) is also consistent with traditional aberration theory. For equal quasi-powers (as in equation (44)) the alternating sum of squares in the coma formula vanishes. Coma itself then vanishes for $\bar{\alpha} = 1/2$, which corresponds to the case of equal conjugates (i.e. transverse magnification $M_T = -1$ in equation (29)). However, if the stop is at the lens, the system is symmetric with respect to the stop, and the zero-coma value can also be derived from the traditional symmetry principle [9].

While in the traditional approach the total values of the Seidel aberrations result only from sums over surfaces, in the present approach the corresponding totals in e.g. equation (43) include constant terms in addition to the sums of quasi-power terms over the surfaces. In equations (16) for spherical aberration and (41) for coma we can consider the terms $c_{S1}(\beta-\alpha)^3 z_k^3$ and $(-1)^{k-1} c_{C1}(\beta-\alpha)^2 z_k^2$ to be the "quasi-power surface contributions". The constant terms are then $c_{S2}(\alpha^3-\beta^3)$ and $c_{C2}(\alpha^2-\beta^2)$, respectively. The example below shows that, numerically, the quasi-power surface contributions can differ significantly from the corresponding traditional ones. Using the Fulcher approach to annul spherical aberration and symmetry to annul coma, the thin triplet with equal conjugates shown in Figure 4 can achieve 3rd-order aplanatic correction in the infrared region. For spherical aberration, the equivalent of equation (47) for $L = 3$ and $\bar{\alpha} = \frac{1}{2}$ (which corresponds to $M_T = -1$) is

$$S_{\min} = -\frac{h^4 K^3 (n-4) n (5n-2)}{108 (n-1)^2 (n+2)}, \tag{48}$$

which becomes zero for $n = 4$ (germanium in infrared).

For the triplet shown in Figure 4 the quasi-powers and the corresponding surface radii $R_k = \frac{1}{c_k}$ resulting from equations (26) and (27) are listed in Table 1, together with aberration coefficients computed using an entrance pupil diameter of 1, and a field angle of 10 degrees. We then have $\alpha = 1/4, \beta = -1/4$. While for spherical aberration the traditional surface contributions vary significantly (note for instance that surfaces 1 and 6 are aplanatic), all quasi-power surface contributions are identical (because the z-values are identical, their cubes are also identical). The zero total spherical aberration is achieved due to the constant term, which has the opposite sign and six times the magnitude of the surface contributions. For coma, all quasi-power surface contributions have the same magnitude, but their total vanishes because of their alternating signs.

In the system shown in Figure 4 three lenses have been used to correct 3rd-order spherical aberration and coma. It is well-known that in fact only two thin lenses are sufficient for annulling, not only these two aberrations, but axial colour as well, while keeping the desired value of the focal length [2]. (Four solutions can be found, and it will be shown in a future publication that the quasi-power approach can explain the reason why the number of possible solutions is precisely four.) However, the approach used above, called by Shafer the "relaxation design method" [4], achieves, as in the Fulcher case, more than just annulling spherical aberration for appropriate values of α, β, and n. The fact that spherical aberration given by equation (48) is also an extremum with respect to small changes of quasi-power leads to a flat design landscape around the solution in Figure 4. Pioneered by Glatzel, the "relaxation design method" is especially useful for systems having a high numerical aperture, due to better tolerances and reduced high-order aberrations, but often at the cost of an increased element count [4]. It is therefore unsurprising that Fulcher-like groups are often encountered as building blocks in lithographic objectives like the one shown in Figure 2. (There, a 2nd Fulcher-like building block can be found in the first wide group of lenses.)

Conclusion

This paper introduces a novel framework for analysing the aberrations of thin lenses, based on the concept of surface quasi-power. The equation (43) shows that in this framework all Seidel aberrations follow the same remarkably simple polynomial pattern. Apart from a constant

term, the aberrations become essentially sums over all surfaces of powers of the new variables. However, even in the zero-thickness limit the contribution of a surface in the present formalism is not the same as the corresponding classical Seidel surface contribution.

In the examples, several classical lens design results follow naturally from the new formalism, in some cases with a simpler derivation than the one found in the existing literature. Optimal singlet bending, relaxed doublets and Fulcher configurations can all be understood as a direct consequence of quasi-power equality. Although the thin-lens approximation employed here is not intended to yield quantitatively accurate predictions, it can serve as a qualitative model that separates the essential properties of primary aberrations from secondary factors.

When expressed in terms of quasi-powers, spherical aberration becomes independent of the internal surface ordering. This property explains the occurrence in the optimization landscape of large families of local minima through permutation symmetry – a property that is already present at the level of third-order theory. Extensions of this work, including more detailed analyses of the design landscape and higher-order effects, will be addressed in a future publication.

Acknowledgments

The author gratefully acknowledges the use of academic licenses for CODE V and Zemax OpticStudio. The author would also like to thank Kumar Rishav for his assistance with the data presented in Figure 3.

Funding

This research was funded by TU Delft.

Conflicts of interest

The author has nothing to disclose.

Data availability statement

Test lenses, a CODE V and a Zemax macro, the corresponding outputs and the Zemax lens for the system in Figure 4 are available at https://doi.org/10.4121/ecc198ad-889a-4ea6-aebb-302303f4e999.

References

1 Welford WT, *Aberrations of Optical Systems* (Adam Hilger, Bristol, 1986).
2 Sasian J, *Introduction to Aberrations in Optical Imaging Systems* (Cambridge University Press, Cambridge, 2013).
3 Fulcher GS, Telescope objective without spherical aberration for large apertures, consisting of four crown glass lenses, *J. Opt. Soc. Am.* **37**, 47 (1947). https://doi.org/10.1364/JOSA.37.000047.
4 Shafer D, Optical design and the relaxation response, *Proc. SPIE* **0766**, 2 (1987). https://doi.org/10.1117/12.940196.
5 Sasaya T et al., Projection optical system and projection exposure apparatus, U.S. Patent 5,805,344 (1998).
6 Caldwell JB, All-fused silica 248-nm lithographic projection lens, *Opt. Photon. News* **9**, 40 (1998). https://doi.org/10.1364/OPN.9.11.000040.
7 Marinescu O, Bociort F, Saddle-point construction in the design of lithographic objectives, part 1: method, *Opt. Eng.* **47**, 093002 (2008). https://doi.org/10.1117/1.2981512.
8 Bociort F, Why are there so many system shapes in lens design? *Proc. SPIE* **7849**, 78490D (2010). https://doi.org/10.1117/12.873880.
9 Gross H et al., *Handbook of Optical Systems*, Vol.**3** (Wiley-VCH, Weinheim, 2007).

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 31
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026023
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE OPEN ACCESS

Zernike by one *Pascal* triangle: A high performance, low memory cost and flexible computation scheme for *Zernike* polynomials

Wei-Jun Chen *

Carl Zeiss Meditec AG, Göschwitzer Str. 51–52, 07745 Jena, Germany

Received 21 December 2025 / Accepted 4 March 2026

Abstract. This work uncovers two hidden cases of block-wise recurrence in *Zernike* computations. Based on these findings, a new computation scheme for *Zernike* polynomials is proposed. It uses one *Pascal*'s triangle for all internal factors, thus avoiding the computation of factorials, cos/sin functions, and matrix inversions. It offers both a direct transformation method and a block-wise recursive method for calculating *Zernike* basis functions, thereby fulfilling the requirements for high accuracy, high speed, low memory footprint, and great application flexibility.

Keywords: *Zernike* polynomials, *Pascal* triangle, Homogeneous bivariate polynomials, Block-wise recurrence, Recursive method, Surface/wavefront reconstruction.

1 Introduction

Zernike polynomials play a key role in many optical applications [1–3] due to their two main features: 1. the orthogonality between individual polynomial components and 2. the direct correspondence with optical *Seidel* aberrations. *Zernike* polynomials are usually defined as

$$z_n^{\pm m}(\rho,\varphi) = R_n^m(\rho)\begin{cases}\cos(m\varphi) & \text{for} \quad +m\\ \sin(m\varphi) & \text{for} \quad -m\end{cases}, \tag{1}$$

$$R_n^m = \sum_{k=0}^{(n-m)/2}\left(\frac{(-1)^k(n-k)!}{k!\left(\frac{n+m}{2}-k\right)!\left(\frac{n-m}{2}-k\right)!}\rho^{n-2k}\right), \tag{2}$$

where R_n^m is called the *Zernike* radial polynomial. The above equations appear in the literature [4, 5] with different notations for the m-term. This work adopts the $\pm m$ notation since this notation provides an explicit and convenient way to divide *Zernike* components of order n into their even part ($\pm m = \pm 2l$), and their odd part $\pm m = \pm\,(2l+1)$, and to subdivide each part into distinct cos and sin subparts. Incidentally, all odd parts of the components for the even order ($n = 2j$) and all even parts of the components for the odd order ($n = 2j+1$) are discarded to avoid non-integer factorials in *Zernike* radial polynomials (equation (2)), and all index terms n, m, l, j, k in this work are non-negative integers, and $l \le j$ and $k \le (j-l)$.

If the polynomial order is not high, the calculation of *Zernike* polynomials is usually satisfied by the definition formulas. For *Zernike* radial polynomials a **three-term-recurrence-relation** (*TTRR*) exists and has been applied component-wise to higher-order recursive *Zernike* calculations [6, 7]:

$$R_n^m = \rho\left(R_{n-1}^{|m-1|}(\rho) + R_{n-1}^{m+1}(\rho)\right) - R_{n-2}^m(\rho). \tag{3}$$

On the other hand, *Zernike* polynomials are essentially one of the orthogonalized versions of the homogeneous bivariate polynomials in a *Cartesian* coordinate system with a unit circle and a coordinate center. In practice, the computation of *Zernike* polynomials is often based on a look-up-table (*LUT*) of the corresponding *Cartesian* forms of the individual *Zernike* components [3, 8, 9], although the complexity of such *Cartesian* forms increases sharply with polynomial order.

To achieve greater flexibility compared to component-wise recursive computations, and to lower the limits of polynomial order and complexity in *Cartesian* form based computations, this work uncovers two cases of block-wise recurrence in *Zernike* computations[1] and offers both a direct block-wise transformation method and a block-wise recursive method for matrix-form *Zernike* computations. The application of such block-wise methods promises higher computational speed and lower memory footprint without compromising accuracy.

The rest of this paper is organized as follows: The next section presents two hidden block-wise recurrences along

* Corresponding author: wei-jun.chen@zeiss.com

[1] To our knowledge, block-wise recurrence in *Zernike* calculations has hardly been addressed in the literature, although there is a long history of intensive studies on *Zernike* polynomials.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

with their close relationship to an n-row *Pascal*'s triangle; subsequently Section 3 explores a new computation scheme for *Zernike* polynomials, which covers basis function computation, coefficient determination, polar/*Cartesian* coordinate conversion, and application specific derivative analysis, among others; further technical discussions can be found in Section 4, where a preliminary performance evaluation is conducted in term of computational complexity, accuracy stability, memory requirements, and application flexibility; finally, Section 5 concludes this paper with an outlook on future works.

2 *Zernike* polynomials supported by one *Pascal*'s triangle

2.1 *Pascal* triangle supported block-wise recurrence

The *Zernike* radial polynomials of the first $n \leq 6$ orders can be expressed in matrix form, where all radial polynomials of an order are obtained by multiplying a coordinate value irrelevant factor matrix by a power vector of the radial coordinate values, as:

$$\begin{cases} R_0^0(\rho) = 1 \\ R_1^1(\rho) = \rho \\ R_2^0(\rho) = 2\rho^2 - 1 \\ R_2^2(\rho) = \rho^2 \\ R_3^1(\rho) = 3\rho^3 - 2\rho \\ R_3^3(\rho) = \rho^3 \\ R_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1 \\ R_4^2(\rho) = 4\rho^4 - 3\rho^2 \\ R_4^4(\rho) = \rho^4 \\ R_5^1(\rho) = 10\rho^5 - 12\rho^3 + 3\rho \\ R_5^3(\rho) = 5\rho^5 - 4\rho^3 \\ R_5^5(\rho) = \rho^5 \\ R_6^0(\rho) = 20\rho^6 - 30\rho^4 + 12\rho^2 - 1 \\ R_6^2(\rho) = 15\rho^6 - 20\rho^4 + 6\rho^2 \\ R_6^4(\rho) = 6\rho^6 - 5\rho^4 \\ R_6^6(\rho) = \rho^6 \end{cases}$$

$$\Longrightarrow \begin{cases} R_0^{\{m\}} = 1 \times \rho^0 \\ R_1^{\{m\}} = 1 \times \rho^1 \\ R_2^{\{m\}} = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} \rho^2 \\ \rho^0 \end{bmatrix} \\ R_3^{\{m\}} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} \rho^3 \\ \rho^1 \end{bmatrix} \\ R_4^{\{m\}} = \begin{bmatrix} 6 & -6 & 1 \\ 4 & -3 & 0 \\ 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \rho^4 \\ \rho^2 \\ \rho^0 \end{bmatrix} \\ R_5^{\{m\}} = \begin{bmatrix} 10 & -12 & 3 \\ 5 & -4 & 0 \\ 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \rho^5 \\ \rho^3 \\ \rho^1 \end{bmatrix} \\ R_6^{\{m\}} = \begin{bmatrix} 20 & -30 & 12 & -1 \\ 15 & -20 & 6 & 0 \\ 6 & -5 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \rho^6 \\ \rho^4 \\ \rho^2 \\ \rho^0 \end{bmatrix} \end{cases} \tag{4}$$

where $R_n^{\{m\}}$ denotes a column vector $\begin{bmatrix} R_{2j}^0 & R_{2j}^2 & \cdots & R_{2j}^{2j} \end{bmatrix}^T$ for even orders ($n = 2j$) or $\begin{bmatrix} R_{2j+1}^1 & R_{2j+1}^3 & \cdots & R_{2j+1}^{2j+1} \end{bmatrix}^T$ for odd orders ($n = 2j + 1$).

From the above equation seven weight vectors, $\{w_n^{\{k\}} | n = 0, 1, ..., 6\}$ with $0 \leq k \leq j$, can be extracted from the corresponding order-specific factor matrices (equation (5)). A block-wise recurrence among the de-weighted factor matrices ($\{T_n | n = 0, 1, \ldots, 6\}$) can be observed: the 0th order de-weighted factor matrix appears in the 2nd order de-weighted factor matrix, while the later subsequentially appears in the 4th order de-weighted factor matrix, which in turn appears in the 6th order de-weighted factor matrix. The same applies to odd orders from the 1st order to the 5th order, as follows:

$$\begin{cases} R_0^{\{m\}} = 1 \times \mathrm{diag}(1) \times \rho^0 \\ R_1^{\{m\}} = 1 \times \mathrm{diag}(1) \times \rho^1 \\ R_2^{\{m\}} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \times \mathrm{diag}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \right) \times \begin{bmatrix} \rho^2 \\ \rho^0 \end{bmatrix} \\ R_3^{\{m\}} = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} \times \mathrm{diag}\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}^T \right) \times \begin{bmatrix} \rho^3 \\ \rho^1 \end{bmatrix} \\ R_4^{\{m\}} = \begin{bmatrix} 6 & 2 & 1 \\ 4 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \times \mathrm{diag}\left(\begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}^T \right) \times \begin{bmatrix} \rho^4 \\ \rho^2 \\ \rho^0 \end{bmatrix} \\ R_5^{\{m\}} = \begin{bmatrix} 10 & 3 & 1 \\ 5 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \times \mathrm{diag}\left(\begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}^T \right) \times \begin{bmatrix} \rho^5 \\ \rho^3 \\ \rho^1 \end{bmatrix} \\ R_6^{\{m\}} = \begin{bmatrix} 20 & 6 & 2 & 1 \\ 15 & 4 & 1 & 0 \\ 6 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \times \mathrm{diag}\left(\begin{bmatrix} 1 \\ -5 \\ 6 \\ -1 \end{bmatrix}^T \right) \times \begin{bmatrix} \rho^6 \\ \rho^4 \\ \rho^2 \\ \rho^0 \end{bmatrix} \end{cases}$$

$$\Longrightarrow R_{n\leq 6}^{\{m\}} = T_n \times \mathrm{diag}(w_n^{\{k\}}) \times \begin{bmatrix} \rho^n \\ \rho^{n-2} \\ \vdots \\ \rho^{(n \bmod 2)} \end{bmatrix}, \tag{5}$$

where $\mathrm{diag}(w_n^{\{k\}})$ denotes the diagonalization of a vector to a diagonal matrix, as

$$\mathrm{diag}(w_n^{\{k\}}) = \begin{bmatrix} w_n^0 & 0 & \cdots & 0 \\ 0 & w_n^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^j \end{bmatrix}. \tag{6}$$

It appears that for order $n \geq 2$, the de-weighted factor matrix T_n can be interpreted as an extended T_{n-2}, with zero-padding at the bottom and a factor vector $t_n^{\{m\}}$ attached to its left side, as described in the following equation:

Table 1. A standard 6-row *Pascal* triangle and its right half vectors.

$n=0$										1			
$n=1$							1			1			
$n=2$					1					2		1	
$n=3$				1			3			3		1	
$n=4$			1		4					6	4		1
$n=5$		1		5			10			10		5	1
$n=6$	1		6		15					20	15	6	1

$$\begin{cases} T_0 = 1 \\ T_1 = 1 \\ T_{n\leq 6} = \begin{bmatrix} t_n^{\{m\}} & \begin{bmatrix} T_{n-2} \\ 0 \end{bmatrix} \end{bmatrix} \end{cases}. \tag{7}$$

Note that both the extracted weight vector $w_n^{\{k\}}$ and the de-weighted factor vector $t_n^{\{m\}}$ are closely related to an n-row *Pascal* triangle. In particular, $t_n^{\{m\}}$ corresponds to the right half of the nth row vector in a standard *Pascal* triangle (Table 1), i.e., here we have $t_6^{\{m\}} = [20 \quad 15 \quad 6 \quad 1]^T$, $t_5^{\{m\}} = [10 \quad 5 \quad 1]^T$, $t_4^{\{m\}} = [6 \quad 4 \quad 1]^T$, $t_3^{\{m\}} = [3 \quad 1]^T$, $t_2^{\{m\}} = [2 \quad 1]^T$, $t_1^{\{m\}} = 1$ and $t_0^{\{m\}} = 1$ for radial polynomials of orders $n \leq 6$; On the other hand, all extracted weight vectors $w_n^{\{k\}}$ can be found as corresponding anti-diagonal vectors in the left-aligned *Pascal* triangle (Table 2), as $w_{0,1}^{\{k\}} = 1$, $w_2^{\{k\}} = [1 \quad -1]$, $w_3^{\{k\}} = [1 \quad -2]$, $w_4^{\{k\}} = [1 \quad -3 \quad 1]$, $w_5^{\{k\}} = [1 \quad -4 \quad 3]$, $w_6^{\{k\}} = [1 \quad -5 \quad 6-1]$, where all odd elements (numbered from 0 onwards) are additionally assigned a negative sign.

To our knowledge, the above-mentioned discoveries are rarely discussed in the literature. Therefore, it is necessary to prove that these discoveries, i.e., the separation of coordinate-relevant and -irrelevant computations (equation (4)), the extraction of order-specific weight vectors (equation (5)), the block-wise recurrence within de-weighted factor matrices (equation (7)), and the close relationship to an n-row *Pascal* triangle (Tables 1 and 2), hold for *Zernike* radial polynomials without order limit (i.e., $\forall n > 6$); if possible and expected, this also holds for complete *Zernike* polynomials (i.e., radial × azimuth polynomials)

2.2 From canonical to block-wise *Zernike* calculation

The canonical definition of *Zernike* radial polynomials, equation (2), has its binomial form:

$$R_n^m = \sum_{k=0}^{(n-m)/2} (-1)^k \binom{n-k}{k} \binom{n-2k}{\frac{n-m}{2}-k} \rho^{n-2k}. \tag{8}$$

If we swap the terms m and k in the second binomial term, we obtain

$$R_n^m = \sum_{k=0}^{(n-m)/2} (-1)^k \binom{n-k}{k} \binom{n-2k}{\frac{n-2k}{2}-\frac{m}{2}} \rho^{n-2k}. \tag{9}$$

Table 2. A left-aligned 6-row *Pascal* triangle and its anti-diagonal vectors.

$n=0$	1					
$n=1$	1	1				
$n=2$	1	2	1			
$n=3$	1	3	3	1		
$n=4$	1	4	6	4	1	
$n=5$	1	5	10	10	5	1
$n=6$	1	6	15	20	15	6	1

In this equation, the term $\binom{n-k}{k}$ explicitly denotes the anti-diagonal vectors in a left-aligned *Pascal* triangle, and the term $\binom{n-2k}{\frac{n-2k}{2}-\frac{m}{2}}$ denotes the left and right halves of the $(n-2k)$th-row in a standard *Pascal* triangle, since *Pascal*'s triangle has its mirror symmetry with respect to its center position $\frac{n-2k}{2}$.

By Combining *de Moivre*'s formula $(\cos m\varphi + i\sin m\varphi) = (\cos\varphi + i\sin\varphi)^m$ with the radial polynomials and converting the coordinate system from polar to *Cartesian* coordinates, we obtain a direct block-wise transformation method for calculating *Zernike* basis functions in the *Cartesian* coordinate system. If we define $X_j^{\{q\}} = \left[x^{2j}y^0 \quad x^{2j-2}y^2 \cdots x^0y^{2j}\right]^T$, and set $w_n^k = (-1)^k \binom{n-k}{k}$ and $t_n^{\{l=m/2\}} = \binom{n-2k}{\frac{n-2k}{2}-\frac{\{m\}}{2}}$, where $\{l\}$ denotes the index vector as $\{l\} = [0 \quad 1 \quad \cdots \quad n/2]^T$ for even orders or $\{l\} = [0 \quad 1 \quad \cdots \quad (n-1)/2]^T$ for odd orders, then such a direct block-wise transformation is represented as a transformation of homogeneous bivariate (xy) polynomials into complete *Zernike* polynomials, as the following:

$$\begin{cases} z_{2j}^{-\{2l\}} = \sum_{k=0}^{j} w_{2j}^k \left(t_{2(j-k)}^{\{l\}} \odot B_{2(j-k)}^{-}\right) \times X_{(j-k)}^{\{q\}} x^{-1}y \\ z_{2j}^{+\{2l\}} = \sum_{k=0}^{j} w_{2j}^k \left(t_{2(j-k)}^{\{l\}} \odot B_{2(j-k)}^{+}\right) \times X_{(j-k)}^{\{q\}} \\ z_{2j+1}^{-\{2l+1\}} = \sum_{k=0}^{j} w_{2j+1}^k \left(t_{2(j-k)+1}^{\{l\}} \odot B_{2(j-k)+1}^{-}\right) \times X_{(j-k)}^{\{q\}} y \\ z_{2j+1}^{+\{2l+1\}} = \sum_{k=0}^{j} w_{2j+1}^k \left(t_{2(j-k)+1}^{\{l\}} \odot B_{2(j-k)+1}^{+}\right) \times X_{(j-k)}^{\{q\}} x \end{cases}, \tag{10}$$

where the terms of $B_{2j}^{-}, B_{2j}^{+}, B_{2j+1}^{-}$ and B_{2j+1}^{+} denote four coordinate transformation matrices from the polar coordinate system to the *Cartesian* coordinate system for corresponding cos/sin components in *Zernike* basis functions of even and odd order. These four transformation matrices form **the second block-wise recurrence** in *Zernike* computations, since the following recursion relation holds: [2]

$$\begin{cases} B_0^- = 0, B_0^+ = 1, B_1^- = 1, B_1^+ = 1 \\ B_{2j}^- = \begin{bmatrix} 0 & 0 \\ B_{2j-1}^+ + B_{2j-1}^- & 0 \end{bmatrix} \\ B_{2j}^+ = \begin{bmatrix} \begin{bmatrix} B_{2j-1}^+ & 0 \end{bmatrix} + \begin{bmatrix} 0 & B_{2j-1}^- \end{bmatrix} \\ \begin{bmatrix} B_{2j-1}^{+(j-1)} & 0 \end{bmatrix} - \begin{bmatrix} 0 & B_{2j-1}^{-(j-1)} \end{bmatrix} \end{bmatrix}, \\ B_{2j+1}^- = B_{2j}^+ + B_{2j}^- \\ B_{2j+1}^+ = B_{2j}^+ - \begin{bmatrix} 0 & B_{2j}^{-\{0\le k\le j-1\}} \end{bmatrix} \end{cases} \tag{11}$$

which is fully supported by the recursive nature of *Pascal*'s triangle, as are $w_n^{\{k\}}$ and $t_n^{\{l\}}$. By expanding the expressions $w_n^{\{k\}}, t_n^{\{l\}}$ and $B_n^{\mp}$ in equation (10) with their corresponding recursive forms, we obtain a block-wise recursive computation method for *Zernike* basis functions:

$$\begin{cases} z_0^- = 0, z_0^+ = 1, z_1^- = y, z_1^+ = x \\ z_{2j}^- = x \begin{bmatrix} 0 \\ \mathcal{O}_+(z_{2j-1}^-) \end{bmatrix} - y \begin{bmatrix} 0 \\ \mathcal{O}_-(z_{2j-1}^+) \end{bmatrix} - \begin{bmatrix} z_{2j-2}^- \\ 0 \end{bmatrix} \\ z_{2j}^+ = x \begin{bmatrix} 2z_{2j-1}^{+0} \\ \mathcal{O}_+(z_{2j-1}^+) \end{bmatrix} + y \begin{bmatrix} 2z_{2j-1}^{-0} \\ \mathcal{O}_-(z_{2j-1}^-) \end{bmatrix} - \begin{bmatrix} z_{2j-2}^+ \\ 0 \end{bmatrix} \\ z_{2j+1}^- = x\mathcal{O}_+(z_{2j}^-) - y\mathcal{O}_-(z_{2j}^+) - \begin{bmatrix} z_{2j-1}^- \\ 0 \end{bmatrix} \\ z_{2j+1}^+ = x\mathcal{O}_+(z_{2j}^+) + y\mathcal{O}_-(z_{2j}^-) - \begin{bmatrix} z_{2j-1}^+ \\ 0 \end{bmatrix} \end{cases}, \tag{12}$$

where $z_{2j}^{\mp} \equiv z_{2j}^{\mp\{0\le l\le j\}} \equiv z_{n=2j}^{\mp\{m=2l\}}$ denotes a column vector of the *Zernike* basis function of order $n = 2j$ with $j + 1$ elements at position (x, y), and analogously $z_{2j+1}^{\mp}$ for an odd order $2j + 1$. The two vector operators $\mathcal{O}_+$ and $\mathcal{O}_-$ are defined as follows:

$$\begin{cases} \mathcal{O}_+(z_n^{\mp}) = \begin{bmatrix} z_n^{\mp\{<\}} \\ 0 \end{bmatrix} + z_n^{\mp} \\ \mathcal{O}_-(z_n^{\mp}) = \begin{bmatrix} z_n^{\mp\{<\}} \\ 0 \end{bmatrix} - z_n^{\mp} \\ z_n^{\mp\{<\}} = \begin{bmatrix} \cancel{z_n^{\mp 0}} & z_n^{\mp 1} & z_n^{\mp 2} & \cdots & z_n^{\mp j} \end{bmatrix}^T \\ \qquad\quad = \begin{bmatrix} z_n^{\mp 1} & z_n^{\mp 2} & \cdots & z_n^{\mp j} \end{bmatrix}^T \end{cases}, \tag{13}$$

for performing offset addition and offset subtraction on a given vector.

3 New computation scheme

Zernike calculations in optical applications include, among others, the following topics: basis function calculation, surface (or wavefront) reconstruction, coefficients evaluation/transformation, derivative analysis, etc.

3.1 *Zernike* basis functions

Equation (10) can be rewritten as a

$$\begin{bmatrix} z_n^{\mp m} \end{bmatrix}_{1\times\frac{(n+1)(n+2)}{2}} = \begin{bmatrix} X_n^i \end{bmatrix}_{1\times\frac{(n+1)(n+2)}{2}} \mathcal{T}$$

$$\overset{\text{as}}{\rightarrow} \begin{bmatrix} z_0^0 \\ z_1^{-1} \\ z_1^1 \\ z_2^{-2} \\ z_2^0 \\ z_2^2 \\ \vdots \\ z_n^{-n} \\ z_n^{-n+2} \\ \vdots \\ z_n^n \end{bmatrix}^T = \begin{bmatrix} x^0y^0 \\ x^1y^0 \\ x^0y^1 \\ x^2y^0 \\ x^1y^1 \\ x^0y^2 \\ \vdots \\ x^ny^0 \\ x^{n-1}y^1 \\ \vdots \\ x^0y^n \end{bmatrix}^T \mathcal{T}, \tag{14}$$

where $\begin{bmatrix} X_n^i = x^{n-i}y^i \end{bmatrix}$ denotes the set of basis functions of homogeneous bivariate (x, y) polynomials with orders from 0 to n, $\begin{bmatrix} z_n^{\mp m} \end{bmatrix}$ denotes the set of basis functions of *Zernike* polynomials with the same orders, and $\mathcal{T}$ denotes the transformation matrix between these two sets of basis function. It is convenient to arrange xy basis functions in a naturally increasing order (equation (14)), and the same applies to *Zernike* basis functions, i.e., with the notation of $\mp m$[3].

There is the inverse transformation matrix $\mathcal{T}^{-1}$, with which we have

$$\begin{bmatrix} X_n^i \end{bmatrix} = \begin{bmatrix} z_n^{\mp m} \end{bmatrix} \mathcal{T}^{-1}. \tag{15}$$

Although $\mathcal{T}^{-1}$ can be obtained by matrix inversion of $\mathcal{T}$, it is recommended to construct it recursively on the basis of the two block-wise recurrences supported by *Pascal*'s triangle.[4]

The direct transformation method for calculating *Zernike* basis functions can be performed by first calculating the basis functions of homogeneous xy polynomials and then transforming them by multiplying the sparse matrix $\mathcal{T}$. Theoretically, the calculation of the xy-basis functions can be carried out efficiently and accurately according to

$$\begin{cases} \begin{bmatrix} x^ny^0 & x^{n-1}y^1 & \cdots & x^1y^{n-1} \end{bmatrix} \\ \quad = x\begin{bmatrix} x^{n-1}y^0 & x^{n-2}y^1 & \cdots & x^0y^{n-1} \end{bmatrix} \\ \begin{bmatrix} x^ny^0 & x^{n-1}y^1 & \cdots & x^1y^{n-1} & x^0y^n \end{bmatrix} \\ \quad = \begin{bmatrix} \begin{bmatrix} x^ny^0 & x^{n-1}y^1 & \cdots & x^1y^{n-1} \end{bmatrix} & y^n \end{bmatrix} \end{cases}. \tag{16}$$

[2] The mathematical derivation in this section briefly proves the two previously mentioned cases of block–wise recurrence in *Zernike* computations without order restriction. Further details can be found in Appendix A.

[3] This arrangement corresponds to the *OSA/ANSI* standard indices for *Zernike* polynomials [3].

[4] Table B1 in Appendix B gives $\mathcal{T}$, Table B2 its inverse $\mathcal{T}^{-1}$, each up to order *6*.

In addition to the direct block-wise transformation method, this work also presents a block-wise recursive method for calculating *Zernike* basis functions, which is described in equation (12). These two methods are not only mathematically equivalent but also practically interchangeable. The block-wise recursion can start either from any order n once the previous basis functions of order $n-1$ and $n-2$ are known, or from the 0th order.

3.2 Surface/wavefront reconstruction

Given a set of samples of a surface or wavefront, which can be either a set of 3D coordinate values, $\{(x_i, y_i, z_i)|\ i = 0, 1, 2, \cdots\}$ or a set of normal vectors $\{(x_i, y_i, (z_x = \partial z/\partial x, z_y = \partial z/\partial y)_i)|\ i = 0, 1, 2, \cdots\}$, a polynomial description of such a surface or wavefront can be reconstructed using a *Least-Squares* optimization:

$$C_{\{k\}} = \arg\min_{C_{\{k\}}} \sum_k w_{\{i\}}\left(Z_{\{i\}} - \left(X_{\{i\}\times\{k\}}\right) \times C_{\{k\}}\right)^2 \longrightarrow C_{\{k\}}$$

$$= \left(X^T\left(w_{\{i\}}I\right)X\right)^{-1}\left(X^T\left(w_{\{i\}}I\right)Z_{\{i\}}\right), \tag{17}$$

where $C_{\{k\}}$ denotes a set of coefficients, $X_{\{i\}\times\{k\}}$ basis function values at (x,y) with $\{i\}$ rows for individual samples and $\{k\}$ columns for individual basis functions, which can be either homogeneous xy basis functions, or *Zernike* basis functions. $Z_{\{i\}}$ denotes the set of measurement values of surface height $z_{\{i\}}$ or of surface normal vectors $(z_x, z_y)_{\{i\}}$, while $w_{\{i\}}$ assigns weights to the individual samples: if $w_{\{i\}} \equiv 1$, all samples are equally important for the reconstruction.

3.3. Polynomial coefficients

A surface/wavefront can be expressed based on *Zernike* polynomials with basis functions $[z_n^{\mp m}]$ or based on homogenous xy polynomials with basis functions $[x^{n-i}\ y^i]$, as

$$z(x, y/\rho, \varphi) = \sum_{n=0}^{N}\left(\sum_{\mp m=-n,\text{step2}}^{+n} d_n^{\mp m} z_n^{\mp m}(\rho, \varphi/x, y)\right)$$

$$\equiv \sum_{n=0}^{N}\left(\sum_{i=0}^{n} a_{ni} x^{n-i} y^i\right). \tag{18}$$

According to equation (17), $C_{\{k\}} = [d_n^{\mp m}]$ holds for the coefficient vector of *Zernike* polynomials, and $C_{\{k\}} = [a_{ni}]$ the coefficient vector for homogeneous xy polynomials. This leads to another group of transformations

$$\left[d_n^{\mp m}\right]_{\frac{(n+1)(n+2)}{2}\times 1} = \mathcal{T}^{-1}[a_{ni}]_{\frac{(n+1)(n+2)}{2}\times 1} \tag{19}$$

$$[a_{ni}]_{\frac{(n+1)(n+2)}{2}\times 1} = \mathcal{T}\left[d_n^{\mp m}\right]_{\frac{(n+1)(n+2)}{2}\times 1} \tag{20}$$

which allow us to freely transform not only the basis functions but also the corresponding coefficients between *Zernike* polynomials and homogeneous xy polynomials describing a surface or a wavefront.

3.4 Derivatives

Derivative analysis plays an important role in optical applications: first-order derivatives for vertex detection and ray tracing, second-order derivatives for surface curvature analysis, third-order derivatives for local apex detection (with local maximum/minimum curvature) and surface tilt evaluation, etc. Using the two sparse transformation matrices mentioned above, derivative calculations, as well as the evaluation of the optical properties of a surface/wavefront, can be flexibly performed based on its polynomial description, either *Zernike* or homogeneous xy polynomials. In particular, we have the following equation for homogeneous xy polynomials:

$$\begin{cases} \frac{\partial z(x,y/\rho,\varphi)}{\partial x} = \sum_{n=0}^{N}\sum_{i=0}^{n} a_{ni}(n-i)x^{n-i-1}y^i \\ \frac{\partial z(x,y/\rho,\varphi)}{\partial y} = \sum_{n=0}^{N}\sum_{i=0}^{n} a_{ni} i x^{n-i}y^{i-1} \end{cases}$$

$$n \geq i+1; i-1 \geq 0 \longrightarrow \begin{cases} \frac{\partial z}{\partial x} = \sum_{n=0}^{N-1}\sum_{i=0}^{n} a_{(n+1)i}(n-i+1)x^{n-i}y^i \\ \frac{\partial z}{\partial y} = \sum_{n=0}^{N-1}\sum_{i=0}^{n} a_{(n+1)(i+1)}(i+1)x^{n-i}y^i \end{cases}, \tag{21}$$

from which two first-order derivatives of a surface/wavefront described by homogeneous xy polynomials with order $0 \leq n \leq N$ correspond to two surfaces described by two homogeneous xy polynomials with order $0 \leq n \leq N-1$, according to the following equation

$$\begin{cases} \frac{\partial z(x,y/\rho,\varphi)}{\partial x} = \sum_{n=0}^{N-1}\sum_{i=0}^{n} a_{ni}^{(x)} x^{n-i}y^i \\ \frac{\partial z(x,y/\rho,\varphi)}{\partial y} = \sum_{n=0}^{N-1}\sum_{i=0}^{n} a_{ni}^{(y)} x^{n-i}y^i \end{cases}, \tag{22}$$

in which we newly introduce two sets of polynomial coefficients have: $a_{ni}^{(x)} = a_{(n+1)i}(n-i+1)$ for $\partial z/\partial x$ and $a_{ni}^{(y)} = a_{(n+1)(i+1)}(i+1)$ for $\partial z/\partial y$, which can be calculated directly from the coefficients of the original surface. Furthermore, second-order derivatives can be calculated from first-order derivatives, third-order derivatives from second-order derivatives, etc. Using equation (19), these also apply to *Zernike* polynomials in polar and *Cartesian* coordinate systems.

4 Discussion

The two block-wise recurrences discovered in Section 2 can be interpreted as a two-dimensional extension of the chronically observed **three-term-recurrence-relation** (*TTRR*) in orthogonal polynomials. The *Gram–Schmidt* orthogonalization process, which naturally introduces the *TTRR* (i.e., $P_n(x) = (A_n x + B_n)P_{n-1}x + C_n P_{n-2}(x)$), is widely used to construct orthogonal bases from homogeneous xy-polynomials. In particular, for *Zernike* polynomials,

two coordinate-value-independent processes: the orthogonalization process and the coordinate system transformation, respectively lead to these two cases of block-wise recurrence.

Building upon the block-wise recursions revealed above and the recursive nature of *Pascal*'s triangle, this work presents both a block-wise direct transformation method (equation (10)) and a block-wise recursive computation method (equation (12)) for *Zernike* basis functions. Mathematically, these two methods are equivalent. In contrast to the available *Zernike* computation techniques: the component-wise computation of *Zernike* polynomials using a list of functions in the *Cartesian* coordinate system [8], the brute-force conversion of *Zernike* polynomials to *Cartesian* polynomials (homogeneous *xy*-polynomials) [9], and the component-wise recursive computation method [6, 7], this work contributes to a block-wise understanding of *Zernike* polynomials.

Our special contributions include: 1. In addition to avoiding cos/sin calculations (as in [8]), we also avoid all calculations of repetitive factorials, divisions of large integers, and matrix inversions; 2. In contrast to the brute-force implementation of bidirectional conversion up to order 5 (with partial extension up to order 8) between *Zernike* basis functions and homogeneous *xy* basis functions [9], we separate the calculations irrelevant to the coordinate values from those relevant to the coordinate values, and offer either the recursive computation of the coordinate-irrelevant factors, or the recursive computation of the complete *Zernike* basis functions. Such recursions are based on block-wise recurrences for coordinate-irrelevant calculations and their close relationship to a *Pascal* triangle. We exploit these new insights for *Zernike*-related computations, thereby eliminating the tediousness/complexity of high-order *Zernike* components and thus achieving greater flexibility for optical applications without restriction of order. Furthermore, this work discovers a direct, but previously unnoticed, block-wise connection between *Zernike* polynomials in *Cartesian* form and their recursive computations (equations (10)–(12)).

4.1 Computational performance: a preliminary evaluation

The direct block-wise transformation method described in equation (10) and equation (14) for calculating *Zernike* basis functions can be interpreted as an upgraded version of the *xy*-form-LUT method mentioned in Section 1. Instead of explicitly defining a LUT of independent functions or equations for individual *Zernike* components in their *xy*-polynomial form, one or four recursively extendable transformation matrixes and a recursively extendable list of basis functions of *Cartesian xy*-polynomials (e.g., based on equation (16)) are combined by matrix multiplications, resulting in the required *Zernike* basis functions. Theoretically, this upgrade does not change the essential computational complexity for a single *Zernike* component, which can be measured as the number of two arithmetic CPU instructions: multiplication and addition. For a required *Zernike* component with polynomial order $n = 2j$ or $n = 2j + 1$ and azimuth order $m = \pm 2l$ or $m = \pm(2l + 1)$, the numbers of essential arithmetic instructions are listed in Table 3, where the construction of $\mathcal{T}$ is not counted because it is coordinate-independent and can therefore be calculated in advance as a factor LUT for the *Zernike* calculation. In short, this work improves the traditional *Zernike* computation scheme based on the *xy*-polynomial form with two significant advantages: 1) It is no longer necessary to repeatedly calculate the *xy*-basis functions in each calculation function/equation of individual *Zernike* components in *Cartesian* form; 2) There is no longer order restriction caused by the limited list of *Zernike* components with their *Cartesian* form given in the literature.

It is noticeable that the computational complexity of this direct transformation method is not the same for individual *Zernike* components with the same polynomial order n, but depends on their azimuth order $\pm m$, where z_n^{-n} requires minimum and z_n^0 (or z_n^1) maximum calculations. In Figure 1a, two solid curves represent the minimum and maximum number of all multiplications, including the computation of the *xy* basis functions and the matrix multiplication $[X] \times \mathcal{T}$ (equation (14)), for polynomial orders from 0 to 99. These two curves are compared with two additional dashed-dotted curves representing the corresponding performance measurements of the *Zernike* computation method based on component-wise recurrence, where the application scenario is set to compute a single *Zernike* component with given n and m for a coordinate position (x, y). The computational performance of component-wise recursion is achieved by manually programming the formulas (32)–(38) in [7] (*Andersen*, 2018), extending the three-term-recurrence to a five-term-recurrence for full *Zernike* polynomials. Similarly, Figure 1b shows the comparison of computational complexity based on the number of addition operations. These two comparisons demonstrate that the *Zernike* computation method based on direct block-wise transformation has high computational performance.

The above performance evaluation makes sense in theory, but calculating a single component is not so common in practice. In contrast, preparing all components from order 0 to order n is necessary for surface/wavefront reconstruction and analysis. The implementation of the corresponding methods significantly influences the final performance assessment. Two further aspects must be considered: 1) the reuse of lower-order results; 2) the use of modern computer hardware such as *SIMD* (Single-Instruction-Multiple-Data) technique and *Cache*-hierarchy. Block-wise recursion offers a direct route to modern, hardware-compatible implementations. Another computational performance comparison was performed between the component-wise recursive calculation method mentioned above and the block-wise recursive calculation method. Figure 2 shows the advantage of the latter method. Both methods were implemented and tested in the same MATLAB environment (*DELL* laptop, Intel(R) Core(TM) i5-1245U, 16GB RAM, 128MB Intel(R) UHD Graphics, Windows-11 OS, MATLAB-R2020b 64-bit) without any special coding optimization. Although this is only a preliminary assessment, the block-wise recursive method shows better computational performance, scaling linearly with the

Table 3. Counts of arithmetic instructions for a *Zernike* component $z_n^{\pm m}$, where sin and cos components of even and odd orders are considered separately.

Computations	Count of Multiplication (×)	Count of addition (+)
$\left[x^{2j-2l-1}y^{2l+1}\right]_{j\geq 1,0\leq l\leq j}$	$\frac{j(j+1)}{2}$	0
$\left[x^{2j-2l}y^{2l}\right]_{j\geq 0,0\leq l\leq j}$	$\frac{(j+1)(j+2)}{2}-2$	0
$\left[x^{2j-2l}y^{2l+1}\right]_{j\geq 0,0\leq l\leq j}$	$\frac{(j+1)(j+2)}{2}-1$	0
$\left[x^{2j-2l+1}y^{2l}\right]_{j\geq 0,0\leq l\leq j}$	$\frac{(j+1)(j+2)}{2}-1$	0
z_{2j}^{-2l}	$\frac{(j+l)(j-l+1)}{2}$	$\frac{(j+l)(j-l+1)}{2}-1$
z_{2j}^{+2l}	$\frac{(j+l+1)(j-l)+2(j+1)}{2}$	$\frac{(j+l+1)(j-l)+2(j+1)}{2}-1$
$z_{2j+1}^{-(2l+1)}$	$\frac{(j+l+1)(j-l)+2(j+1)}{2}$	$\frac{(j+l+1)(j-l)+2(j+1)}{2}-1$
$z_{2j+1}^{+(2l+1)}$	$\frac{(j+l+1)(j-l)+2(j+1)}{2}$	$\frac{(j+l+1)(j-l)+2(j+1)}{2}-1$

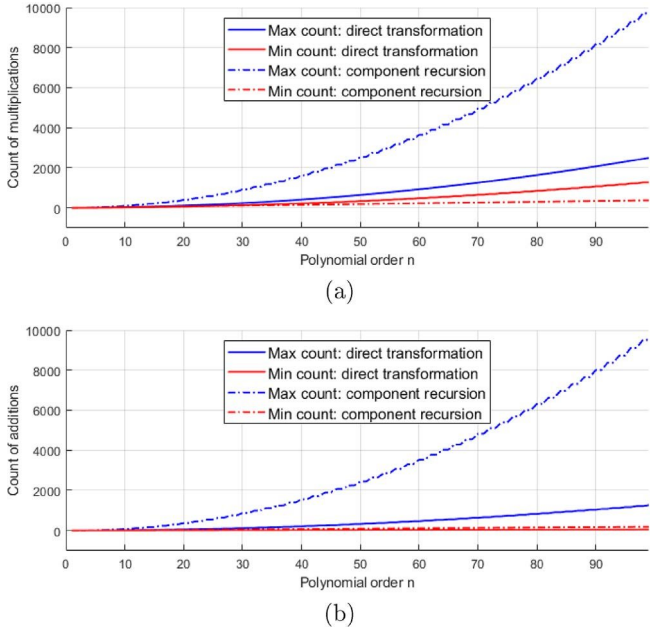


Figure 1. Comparison of computational complexity (I): (a) number of multiplications; (b) number of additions. Four curves in (a) show the maximum and minimum multiplication requirements for the component-wise recursion method and the direct transformation method for a required Zernike component at one (x, y) position; the same applies to four curves in (b) for addition operations.

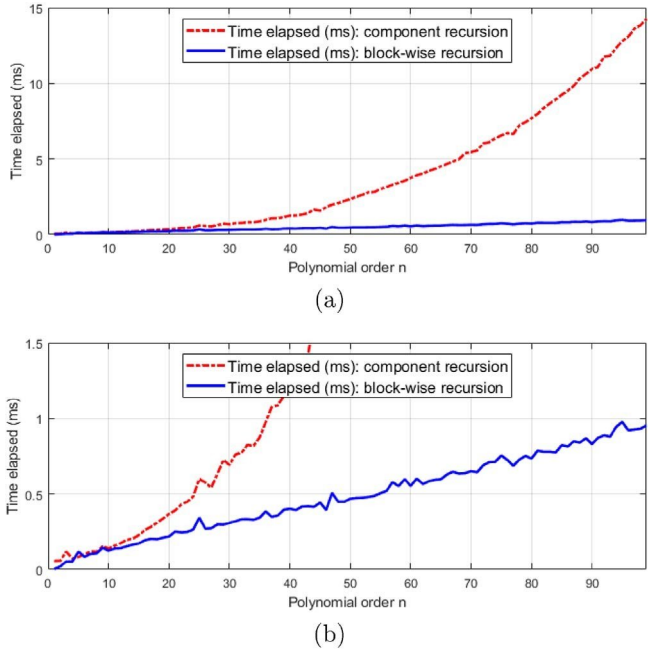


Figure 2. Comparison of computational complexity (II): Time elapsed for generating all *Zernike* components up to order 99 using component-wise recursion (red) compared to block-wise recursion (blue): (a) the comparison; (b) further details on the elapsed time of block-wise recursion (with y-axis ≤1.5 ms). The comparison was performed for one (x, y) position, and the elapsed time is measured in millisecond (10^{-3} s).

polynomial order while the component-wise recursive method appears to be exponentially relevant. This advantage stems from the inherent compatibility between the block-wise recursive method and modern hardware technology in civilian commercial computers.

4.2 Computational accuracy: a preliminary evaluation

In addition to the computational performance test described above, the computational accuracy was also compared using 120 spatial positions taken from 5 radius positions times 24 azimuth positions (see Fig. 3a). For each spatial position, all *Zernike* components from order 0 to order 99 were calculated using both the component-wise recursive method mentioned above ([7]) and the block-wise recursive method based on *Pascal*'s triangle (equation (12)). For each order, the maximum absolute difference between the results of both methods was recorded; thus, a difference curve along polynomial orders was created for each position. All 120 curves were recorded during the test. They are shown in Figure 3 in two versions: the raw difference curves (b) and their logarithmic (log10) version (c).

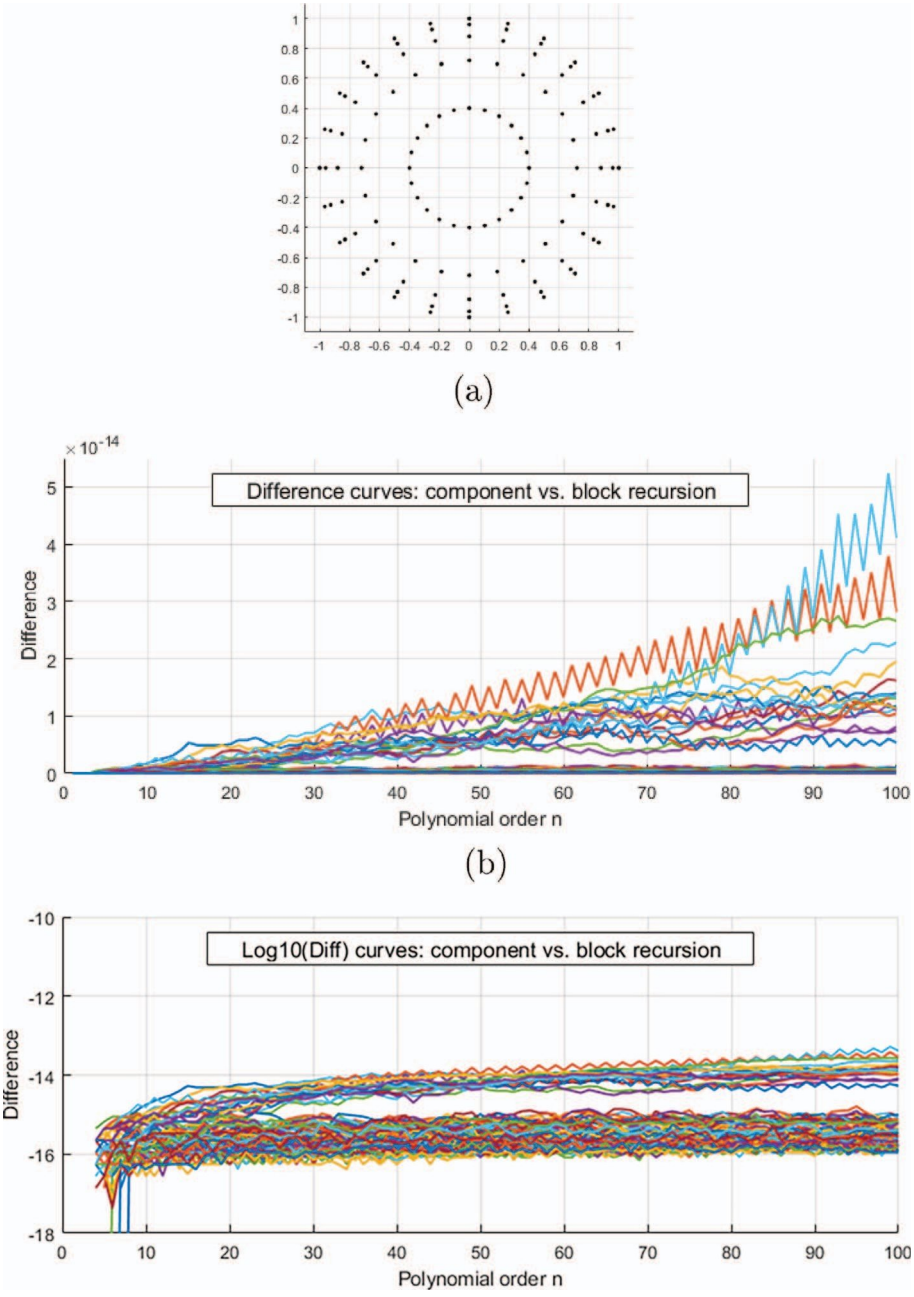


Figure 3. Accuracy assessment of the block–wise recursion compared to the component-wise recursion (*Andersen's* method [7]) for calculating Zernike basis functions: (a) 120 test positions from a circular region with five radii (1.00, 0.96, 0.88, 0.72, 0.40) and 24 uniformly distributed azimuth angles; (b) 120 difference curves (color–coded) with respect to the polynomial orders, each curve representing the maximum absolute differences between two sets of *Zernike* components along the orders from 0 to 99 for a position in (a); (c) the logarithmic representation (log10) of the corresponding curves in (b), of which 17 curves are distinguished from the others by their difference value of over 3×10^{-15} and whose corresponding positions all lie on the unit circle.

The comparison results show an absolute difference of less than 6×10^{-14}, with the maximum absolute difference exceeding 3×10^{-15} at 17 positions on the unit circle from order 50 onwards. This preliminary result confirmed our expectation that the block-wise recursive method can be interpreted as equivalent to the component-wise recursive method, although the former is proposed on the basis of one *Pascal* triangle and the latter was developed from the *TTRR* relationship in orthogonal polynomials.

Naturally, the question arises as to the above 17 positions on the unit circle: which method is closer to the truth? Recursive methods (component-wise and block-wise

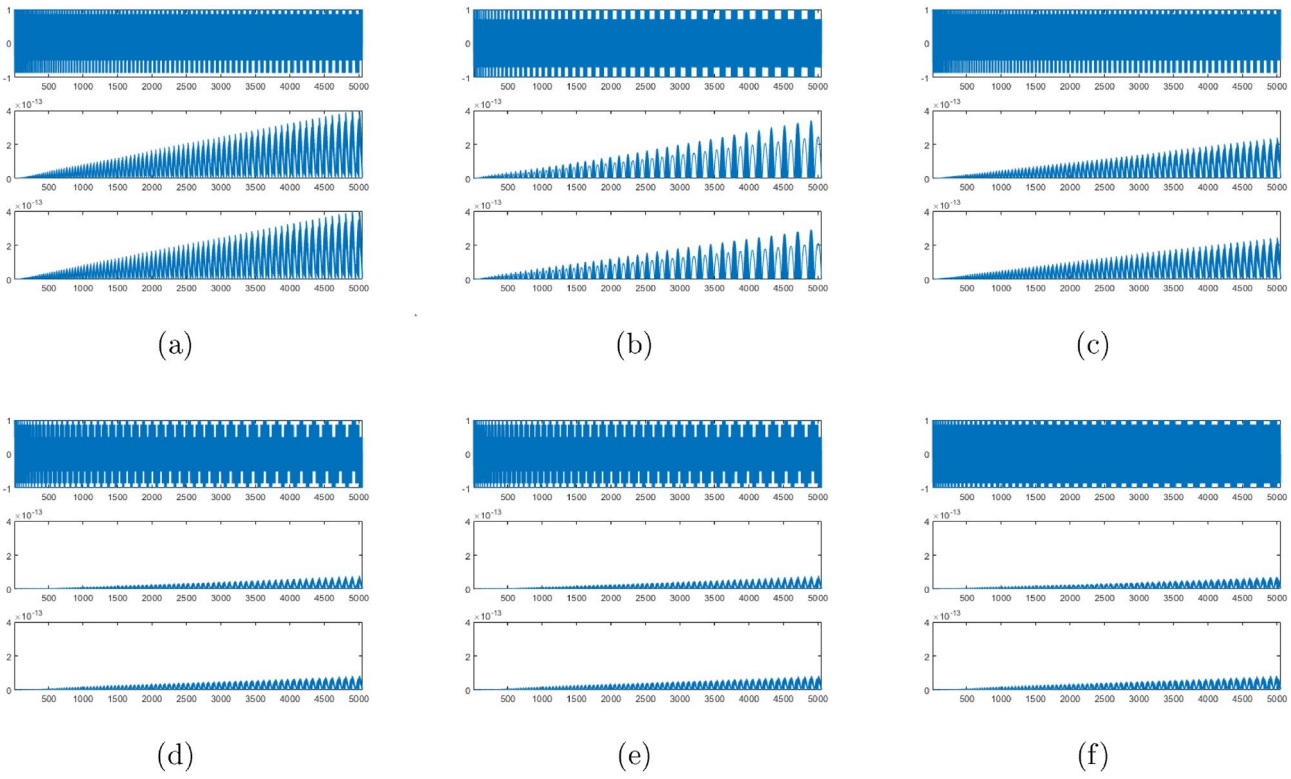


Figure 4. Accuracy comparisons (a)–(f) at six positions on the unit circle: $(x, y) \approx (-0.5000000, -0.8660254)$, $(-0.7071068, -0.7071068)$, $(-0.5000000, 0.8660254)$, $(0.5000000, 0.8660254)$. $(0.5000000, -0.8660254)$, and $(-0.8660254, -0.5000000)$. Each comparison includes three curves: the one above for 5050 *Zernike* components of orders 0 to 99, calculated using the direct transformation method, as the comparison reference; the middle one for the absolute difference of the component-wise recursive method to the reference; and the curve below for the block-wise recursive method. All difference curves are shown with the same value scale: $\leq 4 \times 10^{-13}$.

recursion) were further compared with the direct transformation method (equation (10)). Unlike that the two recursive methods, which were fully implemented in MATLAB with double precision, the direct transformation method requires additional support from an external library for "arbitrary-precision-arithmetic" due to the rapidly increasing integer values (i.e., binomial factors) for polynomials of order higher than 50. The comparison results for the six positions with the largest deviations, all located on the unit circle, are shown in Figure 4. It is evident that for both recursive methods, all deviations are smaller than approximately 4×10^{-13} compared to the direct transformation method.

It should be noted that this work does not constitute a definitive assessment of the accuracy of one method compared to others, but rather a preliminary evaluation to support the conceptual investigation of the relationship between *Zernike* polynomials and *Pascal*'s triangle. Further evaluations with more detailed application scenarios are to be expected, as the difference curves reveal some order-dependent structural information which, although very small, is not random noise.

4.3 Memory requirement

Memory requirements are often a critical factor in development due to detailed application scenarios. This work offers high flexibility for practical system/algorithm development: Block-wise recursive computation for generating basis functions requires no additional memory; similarly, virtually no additional memory is needed for surface/wavefront analysis, which uses transformation matrices for coefficient operations and derivative computations, since all transformation matrices can be rapidly computed by block-wise recursive algorithms, e.g., equation (11) for the forward coordinate system transformation in *Zernike* computations.

5 Conclusion

In this work, the computation of *Zernike* basis functions was divided into two parts: the part irrelevant to the coordinate values and the part relevant to the coordinate values. The first part comprises the computation of the coordinate transformation along with an orthogonalization process (based on homogeneous xy polynomials). This leads to two cases of block-wise recurrence in the coordinate-independent *Zernike* computation. The use of these two block-wise recurrences, supported by *Pascal* triangle, enables a block-wise direct transformation method and a block-wise recursive computation method for *Zernike* polynomials. The latter is more suitable for computing basis functions, while the former is better suited for operations and analyses

of surfaces using their polynomial coefficients. Theoretically, this significantly improves computational speed, reduces memory requirements, and increases application flexibility without compromising accuracy. Optical applications can benefit from these new insights.

This work provides a solid starting point for further theoretical and application-specific studies, such as a comprehensive evaluation of computational complexity, a dedicated algorithmic accuracy/stability investigation, and application-oriented algorithm design, *Zernike* analysis with non-circular pupil, fast higher-order derivatives of a surface described by *Zernike* polynomials, etc. It is also valuable for discovering similar cases of block-wise recurrence in other orthogonal polynomials constructed using the *Gram-Schmidt* process.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflicts of interest in regards to this article.

Data availability statement

The underlying data can be provided by the authors upon reasonable request.

References

1 Lakshminarayanan V, Fleck A, Zernike polynomials: a guide, *J. Modern Opt.* **58**(7), 545–561 (2011). https://doi.org/10.1080/09500340.2011.633763.
2 Schwiegerling J, *Description of Zernike Polynomials*, 2016. https://wp.optics.arizona.edu/visualopticslab/wp-content/uploads/sites/52/2016/08/Zernike-Notes-15Jan2016.pdf.
3 Niu K., Tian C, Zernike polynomials and their applications, *J. Opt.* **24**(12), 1–54 (2022). http://doi.org/10.1088/2040-8986/ac9e08.
4 Born M, Wolf E, Bhatia AB, Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light, 7th ed., Vol. 523 (Cambridge University Press, Cambridge, UK, 1999). ISBN 978-0-521-64222-4. https://books.google.de/books?id=nUHGp NsGyUCq=Zernikeredir_esc=y#v=snippetq=Zernikef=false.
5 Zernike F, Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode, *Physica* **1**(7–12), 689–704 (1934), http://doi.org/10.1016/S0031-8914(34)80259-5.
6 Shakibaei BH, Paramesran R, Recursive formula to compute Zernike radial polynomials, *Opt. Lett.* **38**(14), 2487–2489 (2013). http://doi.org/10.1364/OL.38.002487.
7 Andersen TB, Efficient and robust recurrence relations for the Zernike circle polynomials and their derivatives in Cartesian coordinates, *Opt. Exp.* **26**(15), 18878–18896 (2018). http://doi.org/10.1364/OE.26.018878.
8 van Brug H.H. (1997) Efficient Cartesian representation of Zernike polynomials in computer memory, *Proc. SPIE: Fifth Int. Topical Meet. Educ. Train. Opt.* 382–392, http://doi.org/10.1117/12.294412.
9 Mathar RJ, Zernike basis to Cartesian Transformations, *Serbian Astron. J.* **179**, 107–120 (2009). http://doi.org/10.2298/SAJ0979107M.

Appendix A: Mathematical proofs

A.1 Pascal triangle supported block-wise recurrence in Zernike polynomials

The *Zernike* radial polynomial has its binomial form (equation (8)). Defining $n = 2j$, $m = 2l$ for even orders and $n = 2j + 1$, $m = 2l + 1$ for odd orders, and defining two intermediate variables $J_k = j - k$ and $J_{2k} = 2J_k$, we obtain

$$\begin{cases} R_{2j}^{2l} = \sum_{k=0}^{j-l} (-1)^k \binom{J_{2k}+k}{J_{2k}} \binom{J_{2k}}{J_k+l} (\rho^2)^{J_k} \\ R_{2j+1}^{2l+1}; = \rho \sum_{k=0}^{j-l} (-1)^k \left(\binom{J_{2k}+k}{J_{2k}} + \binom{J_{2k}+k}{J_{2k}+1} \right) \left(\binom{J_{2k}}{J_k+l} + \binom{J_{2k}}{J_k+l+1} \right) (\rho^2)^{J_k} \end{cases}. \tag{A.1}$$

Further defining $w_k^e = \binom{J_{2k}+k}{J_{2k}}$, $w_k^o = w_k^e + \binom{J_{2k}+k}{J_{2k}+1}$, $t_{lk}^e = \binom{J_{2k}}{J_k+l}$, and $t_{lk}^o = t_{lk}^e + \binom{J_{2k}}{J_k+l+1}$, we obtain

$$\begin{cases} R_{2j}^{2l} = \sum_{k=0}^{j-l} (-1)^k w_k^e t_{lk}^e (\rho^2)^{J_k} \\ R_{2j+1}^{2l+1} = \rho \sum_{k=0}^{j-l} (-1)^k w_k^o t_{lk}^o (\rho^2)^{J_k} \end{cases}. \tag{A.2}$$

In particular, we have $w_k^e = \binom{J_{2k}+k}{J_{2k}} = \binom{2j-k}{k}$, for which we obtain a vector by continuously increasing k by 1 from 0 to $k = j - l$, provided $w_{\{k\}}^e = \left[\binom{2j}{0} \ \binom{2j-1}{1} \cdots \binom{j+l}{j-l} \right]$ as a row vector, where $2j = n$ for even orders and $l \le j$; we also have

$$w_k^o = w_k^e + \binom{J_{2k}+k}{J_{2k}+1} = \binom{2j-k}{k} + \binom{2j-k}{k-1} = \binom{2j+1-k}{k},$$

provided $w_{\{k\}}^o = \left[\binom{2j+1}{0} \ \binom{2j}{1} \ \cdots \ \binom{j+1+l}{j-l} \right]$, where $2j + 1 = n$ for odd orders and $l \le j$. Either $w_{\{k\}}^e$ or $w_{\{k\}}^o$ denotes a corresponding anti-diagonal vector in a left-aligned n-row *Pascal* triangle (e.g., Table 2 for $n \le 6$). For one even (or odd) order, $w_{\{k\}}^e$ (or $w_{\{k\}}^o$) depends on a single variable k and is independent of l or m.

Meanwhile, for an even order $n = 2j$, we have $t_{lk}^e = \binom{J_{2k}}{J_k+l} = \binom{2(j-k)}{(j-k)+l}$, where $t_{lk}^e|_{k=0} = \binom{2j}{j+l}$ subject to $k = 0$; continuously increasing k by 1 till $k = j - l$, we obtain a row vector $t_{l\{k\}}^e = \left[\binom{2j}{j+l} \ \binom{2j-2}{j-1+l} \ \binom{2j-4}{j-2+l} \cdots \binom{2l}{2l} \right]$ with respect to a particular l, and obtain an anti-upper-triangular matrix for all the $0 \le l \le j$, as

$$T_{n=2j}\left(t^{e}_{lk}\right)=t^{e}_{\{l\}\{k\}}$$

$$=\begin{bmatrix} \binom{2j}{j} & \binom{2j-2}{j-1} & \binom{2j-4}{j-2} & \cdots & \binom{2}{1} & \binom{0}{0} \\ \binom{2j}{j+1} & \binom{2j-2}{j} & \binom{2j-4}{j-1} & \cdots & \binom{2}{2} & \\ \binom{2j}{j+2} & \binom{2j-2}{j+1} & \binom{2j-4}{j} & \cdots & & \\ \cdots & \cdots & \ddots & & & \\ \binom{2j}{2j-2} & \binom{2j-2}{2j-2} & & & & \\ \binom{2j}{2j} & & & & & \end{bmatrix}_{(j+1)\times(j+1)} . \quad (A.3)$$

Also for an odd order $n=2j+1$, we have $t^{o}_{lk}=t^{e}_{lk}+\binom{J_{2k}}{J_k+l+1}=\binom{2(j-k)}{(j-k)+l}+\binom{2(j-k)}{(j-k)+l+1}=\binom{2(j-k)+1}{((j-k)+1)+l}$, and an anti-upper-triangular matrix as follows:

$$T_{n=2j+1}\left(t^{o}_{lk}\right)=t^{o}_{\{l\}\{k\}}$$

$$\begin{bmatrix} \binom{2j+1}{j+1} & \binom{2j-1}{j} & \binom{2j-3}{j-1} & \cdots & \binom{3}{2} & \binom{1}{1} \\ \binom{2j+1}{j+2} & \binom{2j-1}{j+1} & \binom{2j-3}{j} & \cdots & \binom{3}{3} & \\ \binom{2j+1}{j+3} & \binom{2j-1}{j+2} & \binom{2j-3}{j+1} & \cdots & & \\ \cdots & \cdots & \ddots & & & \\ \binom{2j+1}{2j} & \binom{2j-1}{2j-1} & & & & \\ \binom{2j+1}{2j+1} & & & & & \end{bmatrix}_{(j+1)\times(j+1)} . \quad (A.4)$$

The results of mathematical derivation starting from binomial equivalence agree with the intuitive visual observation in the previous Section 2.1:

$$\begin{cases} (-1)^{k}w^{e}_{k}\equiv w^{k}_{n=2j} \\ (-1)^{k}w^{o}_{k}\equiv w^{k}_{n=2j+1} \\ t^{e}_{lk}|_{k=0}\equiv t^{m=2l}_{n=2j} \\ t^{o}_{lk}|_{k=0}\equiv t^{m=2l+1}_{n=2j+1} \\ T_n\left(t^{o/e}_{lk}/0\right)\equiv T_n=\begin{bmatrix} t^{\{m\}}_{n} & \begin{bmatrix} T_{n-2} \\ 0 \end{bmatrix} \end{bmatrix} \end{cases}, \quad (A.5)$$

where $T_n\left(t^{o/e}_{lk}/0\right)$ denotes a variation of $T_n\left(t^{o/e}_{lk}\right)$ by filling the elements below the anti-diagonal elements with zeros. Furthermore, by adopting

$$\begin{cases} \mathrm{diag}\left(w^{\{k\}}_{2j}\right)=\begin{bmatrix} (-1)^{0}w^{e}_{k=0} & & & \\ & (-1)^{1}w^{e}_{k=1} & & \\ & & \ddots & \\ & & & (-1)^{j-l}w^{e}_{k=j} \end{bmatrix}_{(j+1)\times(j+1)} \\ \mathrm{diag}\left(w^{\{k\}}_{2j+1}\right)=;\begin{bmatrix} (-1)^{0}w^{o}_{k=0} & & & \\ & (-1)^{1}w^{o}_{k=1} & & \\ & & \ddots & \\ & & & (-1)^{j-l}w^{o}_{k=j} \end{bmatrix}_{(j+1)\times(j+1)} \end{cases}, \quad (A.6)$$

Equation (A.2) can be re-written as

$$\begin{cases} R^{\{2l\}}_{2j}=T_{2j}\times\mathrm{diag}\left(w^{\{k\}}_{2j}\right)\times\left(\begin{bmatrix}(\rho^2)^{j} & (\rho^2)^{j-1} & \cdots & (\rho^2) & 1\end{bmatrix}^{T}\right)_{(j+1)\times1} \\ R^{\{2l+1\}}_{2j+1}=T_{2j+1}\times\mathrm{diag}\left(w^{\{k\}}_{2j+1}\right)\times\left(\begin{bmatrix}(\rho^2)^{j} & (\rho^2)^{j-1} & \cdots & (\rho^2) & 1\end{bmatrix}^{T}\right)_{(j+1)\times1}\times\rho \end{cases}, \quad (A.7)$$

where the radial coordinate-dependent vector $\begin{bmatrix}(\rho^2)^{j} & (\rho^2)^{j-1} & \cdots & (\rho^2) & 1\end{bmatrix}^{T}$ contains all possible $(\rho^2)^{J_k}$ in equation (A.7), $\forall k\in[0,\ j-1]$, $\forall l\in[0,\ j]$. Essentially equation (A.7) is equivalent to equation (5), and equations (A.3) and ((A.4)) are equivalent to equation (7), without an order limit ($\forall n\geq 0$).

It is noticed that the close relationship between the radial polynomials and *Pascal* triangle can be straightforwardly obtained even without the intermediate variables J_k and J_{2k}: The ordered set of binomial factors $\left\{\binom{n-k}{k}\middle| k=0,1,...(n-m)/2\right\}$ corresponds to an anti-diagonal vector of a left-aligned *Pascal* triangle, where k increases while $n-k$ decreases, each in steps of 1, as shown in Table 2 for $n\leq 6$; **the binomial factor $\binom{\boldsymbol{n-2k}}{\boldsymbol{(n-m)/2-k}}$ can be rewritten as $\binom{\boldsymbol{n-2k}}{\boldsymbol{(n-2k)/2-m/2}}$, which is equivalent to $\binom{\boldsymbol{n-2k}}{\boldsymbol{(n-2k)/2+m/2}}$.** If k varies as the column index and $l=m/2$ for even orders or $l=(m-1)/2$ for odd orders varies as the row index, the ordered factor set $\left\{\binom{n-2k}{(n-2k)/2+l}\middle| l=0,1,j;k=0,1,...j-l\right\}$ forms a 2D matrix. For a certain k, the column vector $\binom{n-2k}{(n-2k)/2+\{l\}}$ corresponds to the right half of the $(n-2k)$th row in *Pascal* triangle (Table 1 for $n\leq 6$). Nevertheless, using the intermediate variables J_k and J_{2k}, additional block-wise relationships such as $w^{o}_{k}=w^{e}_{k}+\binom{J_{2k}+k}{J_{2k}+1}$ and $t^{o}_{lk}=t^{e}_{lk}+\binom{J_{2k}}{J_k+l+1}$ can be obtained, which, although not as intuitive, are useful for further algorithm development.

A.2 Zernike radial polynomials in Cartesian: the second block-wise recurrence

Considering that $\rho^2=x^2+y^2$, its power value $(\rho^2)^{j}$ can be expressed as the multiplication of two vectors:

$$(\rho^2)^{j}=\sum_{q=0}^{j}b^{q}_{j}x^{2(j-q)}y^{2q}=b^{\{q\}}_{j}\times X^{\{q\}}_{j} \quad (A.8)$$

where $b^{\{q\}}_{j}$ denotes a row vector containing all binomial factors of $\{b^{q}_{j}=\binom{j}{q}|q=0,1,2,...j\}$, corresponding to the jth row in an n-row *Pascal* triangle, and $X^{\{q\}}_{j}$ denotes a column vector of the sequentially listed jth basis functions of homogeneous bivariate polynomials for $(x^2,\ y^2)$, as $\begin{bmatrix}x^{2j}y^{0} & x^{2(j-1)}y^{2} & \cdots x^{0}y^{2j}\end{bmatrix}^{T}$.

Further, the radial coordinate-dependent vector in equation (A.7) can be represented in matrix form as follows:

$$\begin{bmatrix} (\rho^2)^j \\ (\rho^2)^{j-1} \\ \vdots \\ (\rho^2) \\ 1 \end{bmatrix}_{(j+1)\times 1} = \begin{bmatrix} b_j^{\{q\}} & 0 & \cdots & 0 & 0 \\ 0 & b_{j-1}^{\{q\}} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & b_1^{\{q\}} & 0 \\ 0 & 0 & \cdots & 0 & b_0^{\{q\}} \end{bmatrix}_{(j+1)\times\frac{(j+1)(j+2)}{2}} \times \begin{bmatrix} X_j^{\{q\}} \\ X_{j-1}^{\{q\}} \\ \vdots \\ X_1^{\{q\}} \\ X_0^{\{q\}} \end{bmatrix}_{\frac{(j+1)(j+2)}{2}\times 1}, \tag{A.9}$$

which can be iteratively represented in submatrix form as follows:

$$\begin{bmatrix} (\rho^2)^j \\ (\rho^2)^{j-1} \\ \vdots \\ (\rho^2) \\ 1 \end{bmatrix}_{(j+1)\times 1} = \mathcal{B}_j \times \mathcal{X}_j = \begin{bmatrix} b_j^{\{q\}} & 0 \\ 0 & \mathcal{B}_{j-1} \end{bmatrix}_{(j+1)\times\frac{(j+1)(j+2)}{2}} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix}_{\frac{(j+1)(j+2)}{2}\times 1} \tag{A.10}$$

where $\mathcal{B}_j$ denotes the complete matrix with $(j+1)$ rows and $\frac{(j+1)(j+2)}{2}$ columns in equation (A.9), $\mathcal{B}_{j-1}$ is its right/below submatrix with j rows and $\frac{(j)(j+1)}{2}$ columns, and the same applies to $\mathcal{X}_j$ and its sub–vector $\mathcal{X}_{j-1}$ for the right column vector containing all $X_j^{\{q\}}$s. Subsequentially equation (A.7) can be written as

$$\begin{cases} R_{2j}^{\{2l\}} = T_{2j} \times \mathrm{diag}\left(w_{2j}^{\{k\}}\right) \times \mathcal{B}_j \times \mathcal{X}_j \\ R_{2j+1}^{\{2l+1\}} = T_{2j+1} \times \mathrm{diag}\left(w_{2j+1}^{\{k\}}\right) \times \mathcal{B}_j \times \mathcal{X}_j \times \rho \end{cases} \tag{A.11}$$

inside which we have two cases of coordinate-irrelevant block-wise recurrence, as follows:

$$\begin{cases} R_{2j}^{\{2l\}} = \begin{bmatrix} t_{2j}^{\{2l\}} & \begin{bmatrix} T_{2j-2} \\ 0 \end{bmatrix} \end{bmatrix} \times \mathrm{diag}\left(w_{2j}^{\{k\}}\right) \times \begin{bmatrix} b_j^{\{q\}} & 0 \\ 0 & \mathcal{B}_{j-1} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} \\ R_{2j+1}^{\{2l+1\}} = \begin{bmatrix} t_{2j+1}^{\{2l+1\}} & \begin{bmatrix} T_{2j-1} \\ 0 \end{bmatrix} \end{bmatrix} \times \mathrm{diag}\left(w_{2j+1}^{\{k\}}\right) \times \begin{bmatrix} b_j^{\{q\}} & 0 \\ 0 & \mathcal{B}_{j-1} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} \times \rho \end{cases}. \tag{A.12}$$

For weight factors, $w_n^0 = 1$ always applies, and subsequentially $w_n^{\{k\}} = \begin{bmatrix} 1 & w_n^{\{k>0\}} \end{bmatrix}$, where $w_n^{\{k>0\}}$ denotes the sub-vector of $w_n^{\{k\}}$ without its first element. Equation (A.12) can be written as follows:

$$\begin{cases} R_{2j}^{\{2l\}} = \begin{bmatrix} t_{2j}^{\{2l\}} & \begin{bmatrix} T_{2j-2} \\ 0 \end{bmatrix} \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}\left(w_{2j}^{\{k>0\}}\right) \end{bmatrix} \times \begin{bmatrix} b_j^{\{q\}} & 0 \\ 0 & \mathcal{B}_{j-1} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} \\ R_{2j+1}^{\{2l+1\}} = \begin{bmatrix} t_{2j+1}^{\{2l+1\}} & \begin{bmatrix} T_{2j-1} \\ 0 \end{bmatrix} \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}\left(w_{2j+1}^{\{k>0\}}\right) \end{bmatrix} \times \begin{bmatrix} b_j^{\{q\}} & 0 \\ 0 & \mathcal{B}_{j-1} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} \times \rho \end{cases} \tag{A.13}$$

where all coordinate-irrelevant calculations, i.e., calculations of and between t, T, w, b and $\mathcal{B}$, are fully supported by one n-row *Pascal* triangle.

A.3 Complete Zernike by one Pascal triangle

If we further define $J_{kl} = j - k - l$ and combine *de Moivre*'s formula with equation (A.2), we obtain

See the Equation (A.14) bottom of the page

which is equivalent to the following equation in matrix form:

See the Equation (A.15) top of the next page

This is an extension of equation (A.13), where we have distinguished four submatrices $\mathcal{B}_{j-1}^{e-}, \mathcal{B}_{j-1}^{e+}, \mathcal{B}_{j-1}^{o-}$ and $\mathcal{B}_{j-1}^{o+}$ for the sin and cos components of even and odd orders, respectively, as well as the submatrix with index (l,k) introduced by the intermediate variable $J_{kl} = j - k - l$, as follows:

$$B_{2j}^{+} = \left[b_{2l}^{\{2p\}} \otimes b_{j-l}^{\{q\}}\right]^{\{l\}\{k\}} = \begin{bmatrix} b_0^{\{2p\}} \otimes b_j^{\{q\}} \\ b_2^{\{2p\}} \otimes b_{j-1}^{\{q\}} \\ \vdots \\ b_{2j}^{\{2p\}} \otimes b_0^{\{q\}} \end{bmatrix}_{(j+1)\times(j+1)}. \tag{A.16}$$

Each submatrix in the matrix above corresponds to a convolution of two vectors as follows:

$$b_{2l}^{\{2p\}} \otimes b_{j-l}^{\{q\}} = \left(\begin{bmatrix} (-1)^0 b_{2l}^0 \\ (-1)^1 b_{2l}^2 \\ \vdots \\ (-1)^l b_{2l}^{2l} \end{bmatrix}^T \right)_{1\times(l+1)}$$

$$\begin{cases} z_{2j}^{+2l} = \sum_{k=0}^{j-l} (-1)^k w_k^e t_{lk}^e & \left(\left(\sum_{q=0}^{J_{kl}} b_{J_{kl}}^q (x^2)^q (y^2)^{J_{kl}-q} \right) \times \left(\sum_{p=0}^{l} (-1)^p b_{2l}^{2p} (x^2)^{l-p} (y^2)^p \right) \right) \\ z_{2j}^{-2l} = x^{-1} y \sum_{k=0}^{j-l} (-1)^k w_k^e t_{lk}^e & \left(\left(\sum_{q=0}^{J_{kl}} b_{J_{kl}}^q (x^2)^q (y^2)^{J_{kl}-q} \right) \times \left(\sum_{p=0}^{l-1} (-1)^p b_{2l}^{2p+1} (x^2)^{l-p} (y^2)^p \right) \right) \\ z_{2j+1}^{+(2l+1)} = x \sum_{k=0}^{j-l} (-1)^k w_k^o t_{lk}^o & \left(\left(\sum_{q=0}^{J_{kl}} b_{J_{kl}}^q (x^2)^q (y^2)^{J_{kl}-q} \right) \times \left(\sum_{p=0}^{l} (-1)^p b_{2l+1}^{2p} (x^2)^{l-p} (y^2)^p \right) \right) \\ z_{2j+1}^{-(2l+1)} = y \sum_{k=0}^{j-l} (-1)^k w_k^o t_{lk}^o & \left(\left(\sum_{q=0}^{J_{kl}} b_{J_{kl}}^q (x^2)^q (y^2)^{J_{kl}-q} \right) \times \left(\sum_{p=0}^{l} (-1)^p b_{2l+1}^{2p+1} (x^2)^{l-p} (y^2)^p \right) \right) \end{cases}, \tag{A.14}$$

$$\begin{cases} z_{2j}^{-\{2l\}} = \left[t_{2j}^{\{2l\}} \quad \begin{bmatrix} T_{2j-2} \\ 0 \end{bmatrix} \right] \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}(w_{2j}^{\{k>0\}}) \end{bmatrix} \times \begin{bmatrix} [b_{2l}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}}]^{\{l\}\{k\}} & 0 \\ 0 & \mathcal{B}_{j-1}^{e-} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} x^{-1} y \\ z_{2j}^{+\{2l\}} = \left[t_{2j}^{\{2l\}} \quad \begin{bmatrix} T_{2j-2} \\ 0 \end{bmatrix} \right] \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}(w_{2j}^{\{k>0\}}) \end{bmatrix} \times \begin{bmatrix} [b_{2l}^{\{2p\}} \otimes b_{j-l}^{\{q\}}]^{\{l\}\{k\}} & 0 \\ 0 & \mathcal{B}_{j-1}^{e+} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} \\ z_{2j+1}^{-\{2l+1\}} = \left[t_{2j+1}^{\{2l+1\}} \quad \begin{bmatrix} T_{2j-1} \\ 0 \end{bmatrix} \right] \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}(w_{2j+1}^{\{k>0\}}) \end{bmatrix} \times \begin{bmatrix} [b_{2l+1}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}}]^{\{l\}\{k\}} & 0 \\ 0 & \mathcal{B}_{j-1}^{o-} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} y \\ z_{2j+1}^{+\{2l+1\}} = \left[t_{2j+1}^{\{2l+1\}} \quad \begin{bmatrix} T_{2j-1} \\ 0 \end{bmatrix} \right] \times \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{diag}(w_{2j+1}^{\{k>0\}}) \end{bmatrix} \times \begin{bmatrix} [b_{2l+1}^{\{2p\}} \otimes b_{j-l}^{\{q\}}]^{\{l\}\{k\}} & 0 \\ 0 & \mathcal{B}_{j-1}^{o+} \end{bmatrix} \times \begin{bmatrix} X_j^{\{q\}} \\ \mathcal{X}_{j-1} \end{bmatrix} x \end{cases}. \tag{A.15}$$

$$\times \begin{bmatrix} b_{j-l}^0 & b_{j-l}^1 & \cdots & b_{j-l}^{j-l} & 0 & \cdots & 0 \\ 0 & b_{j-l}^0 & \cdots & b_{j-l}^{j-l-1} & b_{j-l}^{j-l} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & b_{j-l}^0 & b_{j-l}^1 & \cdots & b_{j-l}^{j-l} \end{bmatrix}_{(l+1)\times(j+1)}, \tag{A.17}$$

which performs the multiplication between the azimuth polynomials and the radial polynomials for $n = 2j$ of even order, $+m = +(2l)$ part, just as $b_{2l}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}}$ for $-m = -(2l)$ part, while $b_{2l+1}^{\{2p\}} \otimes b_{j-l}^{\{q\}}$ and $b_{2l+1}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}}$ stand for $n = 2j+1$ of odd order, $+m = (2l+1)$ and $-m = -(2l+1)$ parts, respectively.

Overall, equation (A.14) and its matrix form (equation (A.15)) show two cases of block-wise recurrence in complete *Zernike* calculations, which are irrelevant to the coordinates, since equation (A.15) can be represented as follows:

$$\begin{cases} z_{2j}^{-\{2l\}} = T_{2j} \times \mathrm{diag}(w_{2j}^{\{k\}}) \times \mathcal{B}_j^{e-} \times \mathcal{X}_j x^{-1} y \\ z_{2j}^{+\{2l\}} = T_{2j} \times \mathrm{diag}(w_{2j}^{\{k\}}) \times \mathcal{B}_j^{e+} \times \mathcal{X}_j \\ z_{2j+1}^{-\{2l+1\}} = T_{2j+1} \times \mathrm{diag}(w_{2j+1}^{\{k\}}) \times \mathcal{B}_j^{o-} \times \mathcal{X}_j y \\ z_{2j+1}^{+\{2l+1\}} = T_{2j+1} \times \mathrm{diag}(w_{2j+1}^{\{k\}}) \times \mathcal{B}_j^{o+} \times \mathcal{X}_j x \end{cases} \tag{A.18}$$

where we have

$$\begin{cases} \mathcal{B}_0^{e-} = 0, \mathcal{B}_0^{e+} = 1 \\ \mathcal{B}_0^{o-} = 1, \mathcal{B}_0^{o+} = 1 \\ \mathcal{B}_j^{e-} = \begin{bmatrix} B_{2j}^- & 0 \\ 0 & \mathcal{B}_{j-1}^{e-} \end{bmatrix}, B_{2j}^- = \left[b_{2l}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}} \right]^{\{l\}\{k\}} \\ \mathcal{B}_j^{e+} = \begin{bmatrix} B_{2j}^+ & 0 \\ 0 & \mathcal{B}_{j-1}^{e+} \end{bmatrix}, B_{2j}^+ = \left[b_{2l}^{\{2p\}} \otimes b_{j-l}^{\{q\}} \right]^{\{l\}\{k\}} \\ \mathcal{B}_j^{o-} = \begin{bmatrix} B_{2j+1}^- & 0 \\ 0 & \mathcal{B}_{j-1}^{o-} \end{bmatrix}, B_{2j+1}^- = \left[b_{2l+1}^{\{2p+1\}} \otimes b_{j-l}^{\{q\}} \right]^{\{l\}\{k\}} \\ \mathcal{B}_j^{o+} = \begin{bmatrix} B_{2j+1}^+ & 0 \\ 0 & \mathcal{B}_{j-1}^{o+} \end{bmatrix}, B_{2j+1}^+ = \left[b_{2l+1}^{\{2p\}} \otimes b_{j-l}^{\{q\}} \right]^{\{l\}\{k\}} \end{cases}, \tag{A.19}$$

such that both $T_{n=2j}$, $T_{n=2j+1}$, $\mathcal{B}_{j=n/2}^{e\pm}$, and $\mathcal{B}_{j=(n-1)/2}^{o\pm}$ have their recursive submatrix decomposition and are fully supported by an n-row *Pascal* triangle.

A.4 Block-wise Zernike calculation

Since $w_n^0 \equiv 1$ always holds, all formulas in equation (A.18) have an alternative pseudo-iterative form. For example, we have

$$\begin{aligned} z_{2j}^{+\{2l\}} = &\left(t_{2j}^{\{l\}} \odot B_{2j}^+ \right) \times X_j^{\{q\}} + T_{2(j-1)} \\ &\times \mathrm{diag}\left(w_{2j}^{\{k>0\}} \right) \times \mathcal{B}_{j-1}^{e+} \times \mathcal{X}_{j-1} \end{aligned} \tag{A.20}$$

where $\odot$ denotes the *Hadamard* product as element-wise multiplication between a vector and a matrix, e.g., we have

$$\begin{aligned} t^{\{l\}} \odot B^{\{l\}\{k\}} = &\begin{bmatrix} t^0 \\ t^1 \\ \vdots \\ t^j \end{bmatrix}_{(j+1)\times 1} \\ &\odot \begin{bmatrix} B^{0,0} & B^{0,1} & \cdots & B^{0,j} \\ B^{1,0} & B^{1,1} & \cdots & B^{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ B^{j,0} & B^{j,1} & \cdots & B^{j,j} \end{bmatrix}_{(j+1)\times(j+1)} \\ = &\begin{bmatrix} t^0 B^{0,0} & t^0 B^{0,1} & \cdots & t^0 B^{0,j} \\ t^1 B^{1,0} & t^1 B^{1,1} & \cdots & t^1 B^{1,j} \\ \vdots & \vdots & \ddots & \vdots \\ t^j B^{j,0} & t^j B^{j,1} & \cdots & t^j B^{j,j} \end{bmatrix}_{(j+1)\times(j+1)} \end{aligned}. \tag{A.21}$$

Equation (A.2) is equivalent to

$$z_{2j}^{+\{2l\}} = \sum_{k=0}^{j} w_{2j}^k \left(t_{2(j-k)}^{\{l\}} \odot B_{2(j-k)}^+ \right) \times X_{(j-k)}^{\{q\}}, \tag{A.22}$$

which offers us a direct block-wise transformation method for calculating the *cos*-relevant components of even-order ($n = 2j$) *Zernike* basis functions in the *Cartesian* coordinate system. Such a transformation applies to all *Zernike* basis functions, as shown in equation (10).

Appendix B: Forward and inverse transformation matrices, up to order 6

Table B1. The forward transformation matrix T, up to order 6.

n		0	1		2			3				4					5						6						
	i/m	0	−1	1	−2	0	2	−3	−1	1	3	−4	−2	0	2	4	−5	−3	−1	1	3	5	−6	−4	−2	0	2	4	6
0	0	1	0	0	0	−1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0
1	0	0	0	1	0	0	0	0	0	−2	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	−2	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	2	1	0	0	0	0	0	0	−6	−3	0	0	0	0	0	0	0	0	0	0	12	6	0	0
	1	0	0	0	2	0	0	0	0	0	0	0	−6	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0
	2	0	0	0	0	2	−1	0	0	0	0	0	0	−6	3	0	0	0	0	0	0	0	0	0	0	12	−6	0	0
3	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	−12	−4	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	3	3	0	0	0	0	0	0	0	0	−12	−12	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	3	−3	0	0	0	0	0	0	0	0	−12	12	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	−1	3	0	0	0	0	0	0	0	0	4	−12	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	6	4	1	0	0	0	0	0	0	0	0	0	−30	−20	−5	0
	1	0	0	0	0	0	0	0	0	0	0	4	8	0	0	0	0	0	0	0	0	0	0	−20	−40	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	12	0	−6	0	0	0	0	0	0	0	0	0	−60	0	30	0
	3	0	0	0	0	0	0	0	0	0	0	−4	8	0	0	0	0	0	0	0	0	0	0	20	−40	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	6	−4	1	0	0	0	0	0	0	0	0	0	−30	20	−5	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	5	1	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	15	10	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	−10	−10	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−10	10	20	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	−15	5	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−5	10	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	15	6	1
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	24	30	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	15	−30	−15
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−20	0	60	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	−15	−30	15
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	−24	30	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	−15	6	−1

Table B2. The inverse transformation matrix $\mathcal{T}^{-1}$, up to order 6, with 2 decimal digits for visualization.

	0	1	2		3			4				5					6											
m/i	0	0	1	0	1	2	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6
0	0	1.00	0	0	0.25	0	0.25	0	0	0	0	0.13	0	0.04	0	0.13	0	0	0	0	0	0	0.08	0	0.02	0	0.02	0	0.08
1	−1	0	0	1.00	0	0	0	0	0.17	0	0.50	0	0	0	0	0	0	0	0.06	0	0.06	0	0.31	0	0	0	0	0	0	0
	1	0	1.00	0	0	0	0	0.50	0	0.17	0	0	0	0	0	0	0	0.31	0	0.06	0	0.06	0	0	0	0	0	0	0	0
2	−2	0	0	0	0	0.50	0	0	0	0	0	0	0.19	0	0.19	0	0	0	0	0	0	0	0	0	0.09	0	0.06	0	0.09	0
	0	0	0	0	0.25	0	0.25	0	0	0	0	0.19	0	0.06	0	0.19	0	0	0	0	0	0	0.14	0	0.03	0	0.03	0	0.14	
	2	0	0	0	0.50	0	−0.50	0	0	0	0	0.38	0	0.00	0	−0.38	0	0	0	0	0	0	0.28	0	0.02	0	−0.02	0	−0.28	
3	−3	0	0	0	0	0	0	0	0.25	0	−0.25	0	0	0	0	0	0	0.15	0	0.05	0	−0.25	0	0	0	0	0	0	0	
	−1	0	0	0	0	0	0	0	0.08	0	0.25	0	0	0	0	0	0	0.05	0	0.05	0	0.25	0	0	0	0	0	0	0	
	1	0	0	0	0	0	0	0.25	0	0.08	0	0	0	0	0	0	0.25	0	0.05	0	0.05	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0.25	0	−0.25	0	0	0	0	0	0	0.25	0	−0.05	0	−0.15	0	0	0	0	0	0	0	0	
4	−4	0	0	0	0	0	0	0	0	0	0	0	0.13	0	−0.13	0	0	0	0	0	0	0	0	0	0.10	0	0.00	0	−0.10	0
	−2	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0.06	0	0	0	0	0	0	0	0	0	0.05	0	0.03	0	0.05	0
	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0.02	0	0.06	0	0	0	0	0	0	0.08	0	0.02	0	0.02	0	0.08	
	2	0	0	0	0	0	0	0	0	0	0	0.13	0	0.00	0	−0.13	0	0	0	0	0	0	0.16	0	0.01	0	−0.01	0	−0.16	
	4	0	0	0	0	0	0	0	0	0	0	0.13	0	−0.13	0	0.13	0	0	0	0	0	0	0.16	0	−0.05	0	−0.05	0	0.16	
5	−5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	−0.06	0	0.06	0	0	0	0	0	0	0	
	−3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0.01	0	−0.06	0	0	0	0	0	0	0	
	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.01	0	0.06	0	0	0	0	0	0	0	
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	0.01	0	0.01	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	−0.01	0	−0.04	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.06	0	−0.06	0	0.06	0	0	0	0	0	0	0	0	
6	−6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	−0.03	0	0.03	0
	−4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0.00	0	−0.02	0
	−2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.01	0	0.01	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0.00	0	0.00	0	0.02	
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0.00	0	0.00	0	−0.03	
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	−0.01	0	−0.01	0	0.03	
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	−0.03	0	0.03	0	−0.03	

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 50
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026034
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Visualization of wavefront aberrations by Zernike polynomials

Stephan Reichelt*

Institute of Applied Optics (ITO), University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Received 28 January 2026 / Accepted 31 March 2026

Abstract. Zernike aberration coefficients are typically presented in tabular form or as simple bar graphs, making it difficult to intuitively interpret their meaning and relate them to the specific aberration term. In this work, we present intuitive and rapidly interpretable visual representations of Zernike aberration terms that highlight the magnitude and orientation of their dominant contributions. Depending on the aberration order, we recommend different graphical formats – such as bubble plots and heatmaps – for visualizing low-order and mid-spatial frequency Zernike terms. These graphical representations are particularly valuable when quick visual feedback on individual aberration terms is needed, such as during aberration compensator adjustment, real-time wavefront visualization of dynamic processes, or optical alignment procedures. They can also enhance the clarity and interpretability of inspection reports or measurement certificates. Furthermore, an alternative definition of the azimuthal orientation of Zernike terms with $m \neq 0$ is proposed, enabling a more effective analysis of the shape and bending of spoke-shaped aberrations.

Keywords: Zernike polynomials, Visualization, Wavefront analysis, Wavefront aberrations, Optical testing.

1 Introduction

Polynomial sets such as Zernike polynomials [1–4] are widely used in various applications to approximate optical aberrations in a compact and sufficiently accurate manner. For example, in precision optical manufacturing, Zernike polynomials are used to characterize and approximate surface figure errors of optical components (e.g., lenses, flats) or wavefront aberrations of optical systems measured by interferometers or wavefront sensors. Freeform surfaces, which can also be described by Zernike polynomials, play an important role in optical design. In addition, Zernike polynomials are typically used in adaptive optics to compensate for dynamic aberrations, such as those caused by atmospheric turbulence.

The present work is motivated by the aim of providing a representation of Zernike approximation results that is as simple and intuitive as possible. Conventional formats such as tabular listings or bar plot visualizations, while widely used, often fail to adequately convey structural relationships and limit the direct interpretability of the underlying wavefront or surface characteristics. For users working with Zernike polynomial representations in optical design, metrology, or ophthalmology, the interpretation of coefficient sets can be non-trivial, particularly in the presence of multiple interacting modes or when spatial patterns must be inferred indirectly. This can make it difficult to extract relevant insights efficiently, even for experienced practitioners. To address this, the present work introduces a novel representation that is specifically designed to make structurally relevant features more directly accessible, thereby facilitating interpretation and supporting more effective analysis of Zernike-based wavefront or surface descriptions.

2 Definition of Zernike polynomials

Figure 1 illustrates the first 15 Zernike polynomials, arranged in a pyramid up to the 4th order.

Zernike polynomials are expressed either in Cartesian (x, y) or conveniently in polar coordinates (r, θ). They are defined within the unit circle, where r is the normalized radial distance from the origin, and θ is the angle from the x axis (common right-handed coordinate system). Apart from a preceding normalization factor N_n^m, Zernike polynomials are the product of a radial polynomial $R_n^m(r)$ and an azimuthal function $M_m(\theta)$

$$Z_n^m(r,\theta) = N_n^m \cdot R_n^m(r) \cdot M_m(\theta), \tag{1}$$

where n is the degree of the polynomial and m is the azimuthal dependence parameter [2, 5]. The azimuthal order m is an integer ($m \in \mathbb{Z}$), whereas the polynomial degree or radial order n is a non-negative integer ($n \in \mathbb{N}_0$). For a particular Zernike polynomial Z_n^m, the integral values n and m are either both even or both odd. The double-indexing scheme, in which n denotes the highest power of the

* Corresponding author: reichelt@ito.uni-stuttgart.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

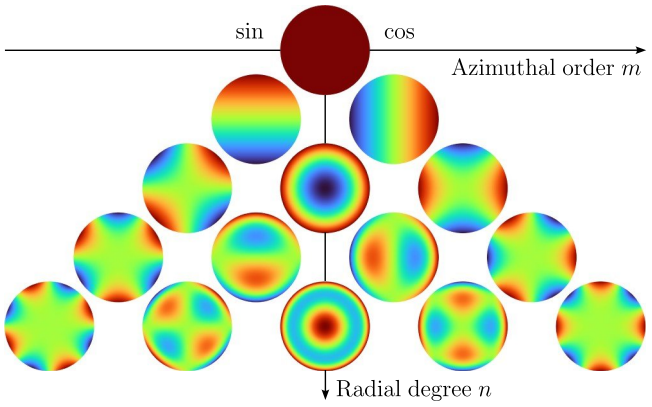


Fig. 1. Pyramid representation of the non-normalized Zernike polynomials Z_n^m up to the 4th order.

radial polynomial and m the azimuthal order, is essential for unambiguously describing the Zernike polynomials.

The radial polynomials $R_n^m(r)$ are explicitly defined by

$$R_n^m(r) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!} r^{n-2s}, \qquad (2)$$

whereas the azimuthal function $M_m(\theta)$ is given by[1]

$$M_m(\theta) = \begin{cases} -\sin(m\theta) & \text{for } m < 0 \\ \cos(m\theta) & \text{for } m \geq 0 \end{cases}. \qquad (3)$$

Optionally, a normalization factor N_n^m may be applied, defined as

$$N_n^m = \begin{cases} 1 & \text{for no normalization} \\ \sqrt{(2-\delta_{m0})(n+1)} & \text{for RMS normalization} \end{cases} \qquad (4)$$

where δ_{m0} is the Kronecker delta function.[2] If the normalization factor N_n^m is set to 1, the Zernike polynomials are referred to as non-normalized, implying that either their maximum is 1 or that their maximum and minimum are ± 1, respectively. In contrast, for RMS-normalized polynomials, the normalization factor is defined such that the root-mean-square (RMS) value of each polynomial is 1. This ensures that, when a polynomial fit is applied to a data set, the absolute value $|c_n^m|$ of the fitted Zernike coefficients directly represent the RMS values of the associated aberration terms. For numerical reasons, it is advantageous to apply the RMS normalization not to the Zernike polynomials themselves, but to the coefficients obtained from the fit.

Note that each Zernike polynomial pair, Z_n^m and Z_n^{-m} (for $m \neq 0$), has the same normalization factor $N_n^m = N_n^{-m}$ as well as the same radial polynomial $R_n^m(r) = R_n^{-m}(r)$.

[1] Note that the sine function is an odd function where $\sin(-x) = -\sin(x)$. For $m < 0$, the negative sign of the sine function is necessary for a correct notation.

[2] Kronecker delta is defined as $\delta_{m0} = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{for } m \neq 0 \end{cases}$.

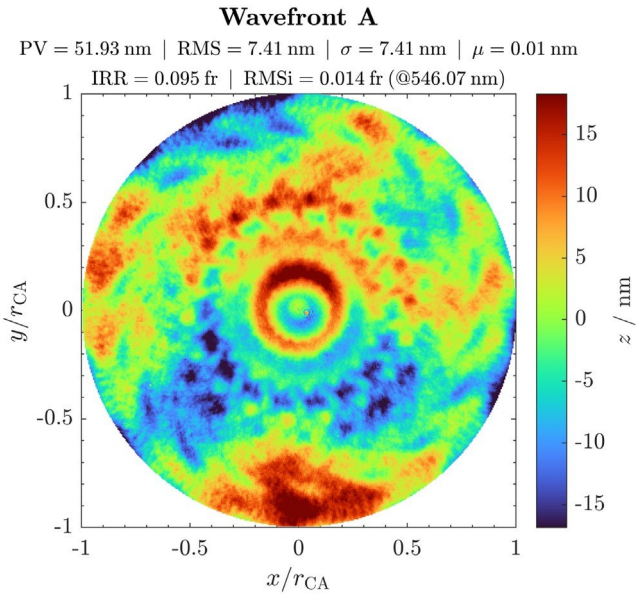


Fig. 2. Sample wavefront A: measurement data of a lens surface.

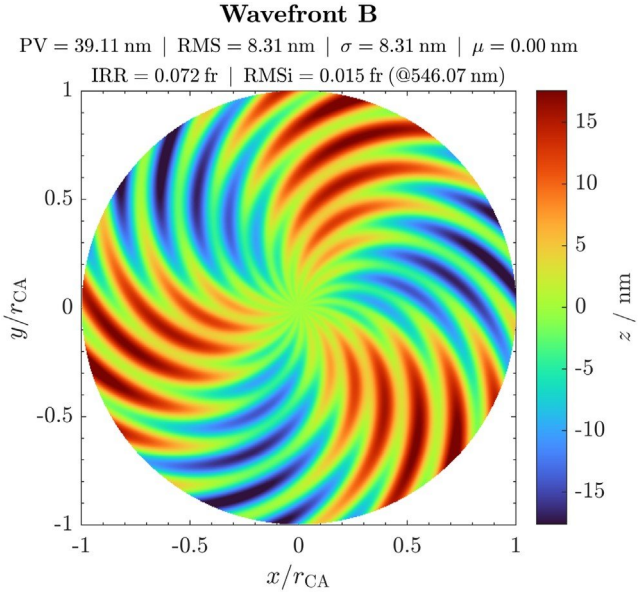


Fig. 3. Sample wavefront B: synthetic aberration data.

3 Sample wavefront data

Two representative wavefronts are employed to illustrate the visualization methods discussed in this paper. Wavefront A (Fig. 2) corresponds to the measured aberrations of a lens surface, containing both low- and mid-spatial frequency contributions, while wavefront B (Fig. 3) is synthetically generated to highlight the orientation of individual azimuthal orders.

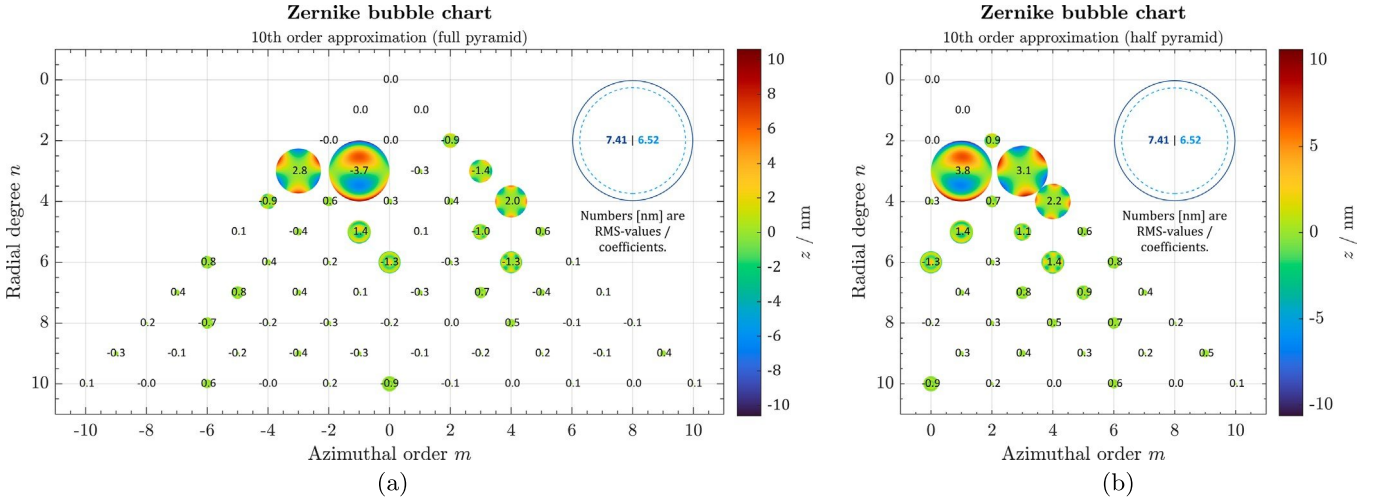


Fig. 4. Bubble chart plots of low-order Zernike aberration coefficients in full-pyramid representation (a) and in half-pyramid representation (b). The area of the depicted aberration terms (bubble size) is proportional to the magnitude of its RMS-normalized Zernike coefficient c_n^m. The blue circles on the upper right represent the areas of the total RMS-deviation of the wavefront error (dark blue) and its fitted RMS-deviation by the shown low-order aberrations fit of a 10th order polynomial (light blue). The numbers within the pyramid represent the RMS-normalized Zernike coefficients [nm], whereas the numbers in the blue reference circles (upper right corner) are the RMS-values [nm] of the original wavefront and its fitted 10th-order counterpart. Primary coma and trefoil are the main aberration terms.

4 Representing low-order Zernike aberrations

Often, the coefficients resulting from a least-squares Zernike fit are presented in tabular form or as simple bar graphs using a single-indexing scheme. However, Zernike polynomials depend on two integer parameters, n and m, which makes any single-indexing scheme arbitrary. This makes it difficult to interpret their meaning and relate them to the specific aberration term, especially since different single-index ordering schemes are in use, such as ANSI/ISO 24157, Born/Wolf, Standard, or Fringe.

Evans et al. [6] proposed a graphical representation, showing the magnitude of the various Zernike orders in dependence of the azimuthal order m and the degree n of the radial polynomial, a method that was adopted by the author years ago for plotting the aberration coefficients as two-dimensional bar plots, e.g. in [7].

Inspired by this, an even more intuitive way for representing low-order Zernike aberration coefficients is proposed here. It makes use of the Zernike pyramid and the fact, that the RMS-normalized Zernike coefficients are uncorrelated to each other. Hence, the root-sum-of-squares (rss) level of individual RMS-normalized Zernike coefficients c_n^m yields the total RMS-value of the aberration

$$\mathrm{RMS}_{\mathrm{total}} = \sqrt{\sum_{i=1}^{\infty} \mathrm{RMS}_i^2} = \sqrt{\sum_{i=1}^{\infty} \left(c_n^m\right)_i^2}. \qquad (5)$$

The basic idea is just use a bubble plot depiction of the Zernike pyramid and scale the size of each individual bubble (or individual Zernike polynomial) with the magnitude of the RMS-normalized Zernike coefficient c_n^m. There is a physical meaning in this scaling as the sum of all individual bubble areas corresponds to the total RMS-value. By way of example this is shown in Figure 4a for an X/Y representation of the full Zernike pyramid.

Here, the total RMS deviation of sample wavefront A having 7.41 nm is taken as the reference area (area $\propto \mathrm{RMS}_{\mathrm{total}}^2 \propto r_{\mathrm{total}}^2$, cf. dark blue circle in the upper right corner). The individual radii scale then with $\left|c_n^m\right|_i/\mathrm{RMS}_{\mathrm{total}}$, i.e. the total area of all shown bubbles (10th order fit with 66 polynomials) therefore corresponds to the light blue dotted area $\propto \mathrm{RMS}_{10\mathrm{th}}^2 = (6.52\,\mathrm{nm})^2$. It can be easily seen in Figure 4a that primary Y–coma and primary Y–trefoil are the main aberration contributors. An advantage of the full pyramid or X/Y representation is that the signs of the Zernike coefficients c_n^m are preserved, either through the positions of the maxima and minima of the individual Zernike terms or through additional numerical annotation.

Another, even more convenient depiction is a Mag/Angle representation utilizing only half of the Zernike pyramid, see Figure 4b. The magnitude of the paired X/Y coefficients is given by

$$|c_n^m| = \sqrt{\left(c_n^{-m}\right)^2 + \left(c_n^{+m}\right)^2}, \qquad (6)$$

whereas the angular orientation results just from adding up $\left(c_n^{-m} Z_n^{-m} + c_n^{+m} Z_n^{+m}\right)$ pairwise. Magnitude and orientation can be then interpreted intuitively.

Note that in Figure 4, the terms piston, tilt, and defocus have been set to zero intentionally, since they describe alignment parameters rather than intrinsic surface or wavefront aberrations.

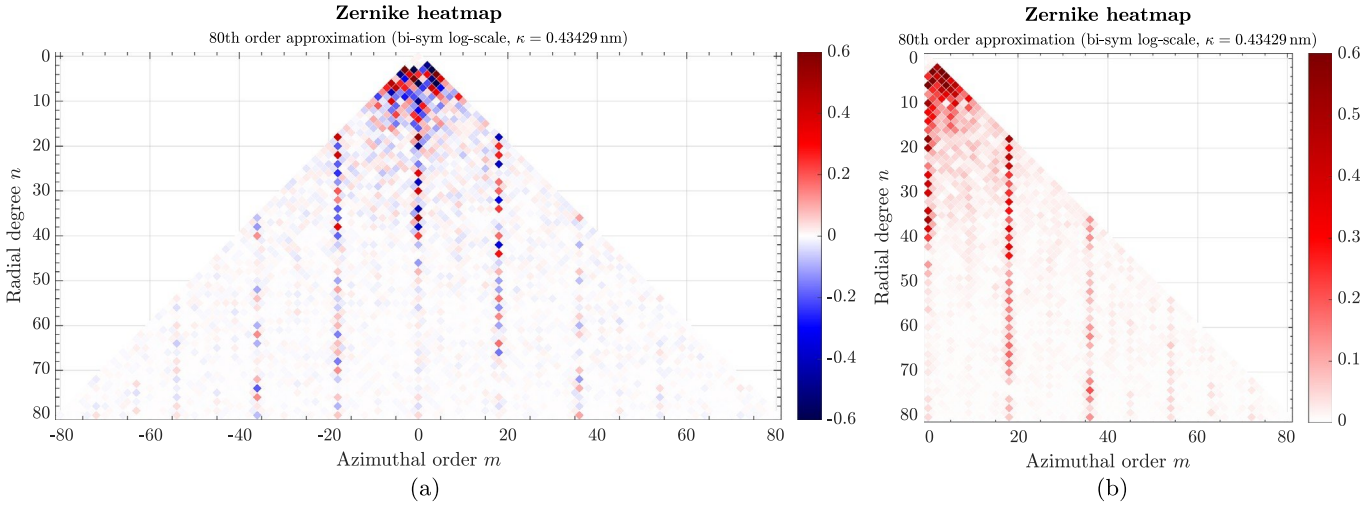


Fig. 5. Heatmaps of Zernike aberration coefficients for a 80th-order approximation in full-pyramid representation (a) and in half-pyramid representation (b). To emphasize coefficients with small magnitudes, the coefficients are displayed on a bi-symmetric logarithmic scale [8]. In the full-pyramid representation, the signs of the coefficients are preserved, whereas in the half-pyramid representation, the magnitudes of the respective coefficient pairs are shown.

5 Representing mid- and high-spatial frequency Zernike terms

Whereas the Zernike pyramid representation with scaled bubbles is adequate for visualizing polynomial degrees up to the 10th or 12th order, it becomes impractical for high-order fits. Zernike coefficient maps (also called Zernike spectra) are ideally suited for this purpose.

Figure 5 presents heatmaps of Zernike aberration coefficients obtained from a 80th-order fit of sample wavefront A. To emphasize coefficients with small magnitudes, the data are displayed using a bi-symmetric logarithmic scale [8]. The scaling constant of the bi-symmetric transfer function, which controls the slope near the origin, was set here to $\kappa = 1/\ln(10)$ nm.

Note that the heatmaps are not displayed as a conventional pixel grid, but instead as a square tiling rotated by 45° and rendered using patches. This representation avoids gaps in the spectrum and ensures that the pyramid is fully filled. Alternative spectral representations [9, 10] employ a modified radial index n_F for the same purpose, namely to circumvent such gaps. The representation presented here simultaneously provides a fully filled pyramid while preserving the conventional ordering of the radial polynomial degree n along the ordinate.

The log-based representation clearly reveals characteristic azimuthal orders with $m = 18$ and their higher harmonics ($m = 36$, 54, 72) in the wavefront. The heatmaps therefore reveal specific mid-spatial-frequency errors, such as ripple and spoke-like structures. Rotationally variant irregularities, or spoke-shaped aberrations, represent a specific form of surface undulations characterized by regular azimuthal orders extending toward the origin of the wavefront map. Such azimuthal waviness is typically introduced by manufacturing processes involving rotational motion about the workpiece center. Examples include precision contour grinding of lenses, where chatter marks may arise from ring tools, as well as diamond turning or direct laser beam writing using air-bearing spindles, where slight imbalances can cause run-out errors.

6 Azimuthal orientation

The angular or azimuthal orientation refers to the angle with respect to the positive x-axis at which the aberration occurs. For example, combining equal positive tilts in the x- and y-directions produces a resultant tilt oriented at +45°. For a tilt, the direction of the aberration is unambiguous when referenced to its maximum. However, once the radial polynomial contains more than one term, local maxima or minima arise along the radial profile. How, for example, should the orientation of primary coma (cf. Fig. 6, center) be specified? Two approaches can be considered. One may, for instance, refer to the maximum value at the edge of the aperture, i.e. at $r = 1$ (previous definition). Alternatively, one may refer to the first local maximum in radial direction, which for primary coma occurs at the radius $r = \sqrt{2}/3 \approx 0.471$ (proposed definition). These two variants are discussed in more detail in the following.

In Evans et al. [6] an equation and procedure to determine the first peak in *angular* direction at the edge of the aperture (i.e. at radius $r = 1$) is derived, which further simplifies with the atan2-function to

$$\theta_0^{\mathrm{lp}} = \frac{1}{m}\,\mathrm{atan2}\left(c_n^{-m}, c_n^{m}\right), \tag{7}$$

where $m \in \mathbb{N}$. This variant is referred to here as the previous definition. The angular offset angle is denoted hereafter by the superscript "lp" which stands for "last peak in *radial* direction". As can be seen from Figure 1, for all Zernike polynomials with a cosine azimuthal dependence ($m > 0$, right

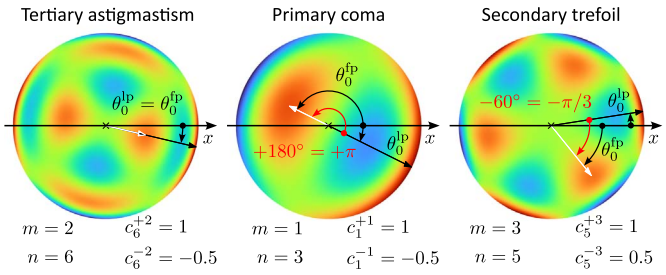


Fig. 6. Comparison of the definitions for azimuthal orientation of $\pm m$-paired Zernike polynomials. The angular orientation of the Zernike polynomials $Z_n^{\pm m}$ is illustrated for different choices of θ_0 (black arrow: common definition θ_0^{lp}; white arrow: proposed new definition θ_0^{fp}). Left: tertiary astigmatism $Z_6^{\pm 2}$ with peaks at same azimuthal orientation. Middle: primary coma $Z_3^{\pm 1}$, where $\pi = 180°$ has to be added. Right: secondary trefoil $Z_5^{\pm 3}$, where $\pi/3 = 60°$ must be subtracted.

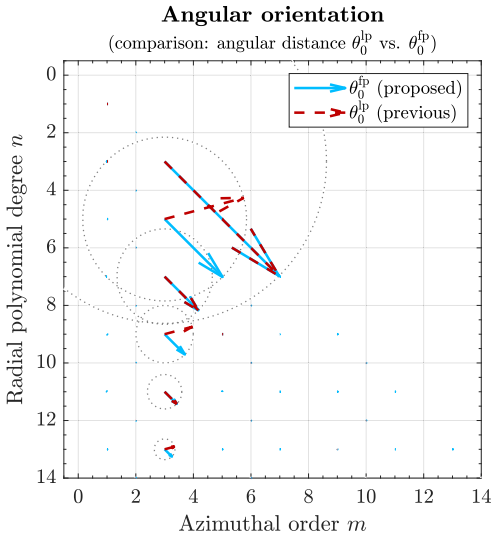


Fig. 7. Comparison of the two definitions for angular orientation. The vector length represents the magnitude of the Zernike coefficient $|c_n^m|$, and the angle represents its orientation.

side of pyramid) this last peak in radial direction at $r = 1$ is always located on the positive x-axis (at $x = 1$, $y = 0$).

Here, we define another angle θ_0^{fp} for the angular orientation, the one that points to the first local peak in the radial direction (proposed definition). The difference is illustrated in Figure 6.

If $(n \bmod 4) = (m \bmod 4)$, then there is no difference in the definition. However, in all other cases (i.e., for every second diagonal wing pair of the Zernike pyramid, beginning with the primary coma pair), a special treatment is required. π/m must either be added or subtracted, depending on the signs of the coefficient pair, i.e.

$$\theta_0^{\mathrm{fp}} = \begin{cases} \theta_0^{\mathrm{lp}} & \text{if } (n \bmod 4) = (m \bmod 4) \\ \theta_0^{\mathrm{lp}} + \dfrac{\pi}{m} & \text{else if } (c_n^{-m} < 0) \vee (c_n^{-m} = 0 \wedge c_n^m > 0) \\ \theta_0^{\mathrm{lp}} - \dfrac{\pi}{m} & \text{else if } (c_n^{-m} > 0) \vee (c_n^{-m} = 0 \wedge c_n^m < 0) \end{cases} . \tag{8}$$

The three conditions can be combined into a compact formula suitable for implementation

$$\theta_0^{\mathrm{fp}} = \frac{1}{m}\mathrm{atan2}\left(c_n^{-m}, c_n^m\right) - \frac{\pi}{m}\{[(n-m) \bmod 4] \neq 0\}\mathrm{sign}\left[\mathrm{atan2}\left(c_n^{-m}, c_n^m\right)\right], \tag{9}$$

where the first term corresponds to equation (7), the bracketed factor {.} represents a logical flag [0, 1] to distinguish the cases, and the sign-function determines whether π/m is added or subtracted.

The alternative definition of the azimuthal orientation of Zernike polynomials with $m \neq 0$ is motivated by the aim of enabling a more effective analysis of spoke-shaped aberrations. Consider the synthetically generated sample wavefront B (Fig. 3), which contains rotation-invariant aberration components only and is composed of spoke-shaped aberrations with azimuthal orders $m = 3$ (straight spokes) and $m = 18$ (bended spokes). The angular orientation of the straight, spoke-shaped trefoil in sample wavefront B is compared for both definitions in Figure 7. According to the previous definition (θ_0^{lp} as defined in equation (7), shown as red dashed arrows), the angle is rotated by π/m for every second occurrence of the polynomial degree n (i.e. $n = 5$, 9, 13, ...). In contrast, under the proposed definition (θ_0^{fp} according to equation (9), shown as light blue arrows), the angular orientation always points in the same direction (here $-45°$). This makes it possible to trace the radial progression of the spokes.

To analyze the bending or the shape of the spokes, representations other than the phasor plot in Figure 7 are more suitable. First, dominant azimuthal orders are identified by applying a threshold to the root-sum-square combination of the coefficients of a given order m. To plot the determined angles θ_0^{fp} as a function of radius, or to directly visualize their variation within the circular aperture, the radial coordinate of the first local maximum is required (cf. the endpoints of the white arrows in Fig. 6). This radius is determined numerically by searching for the first extremum of the corresponding radial polynomial R_n^m. The values are computed once and stored in a look-up table.

Note that the angle θ_0 in both equations (7) and (8) is confined to a circular sector spanning the arc $0 \pm \pi/m$. With increasing m, this minor sector becomes progressively smaller. As a consequence, owing to the arctan function, the evaluated angle θ_0 is wrapped modulo $2\pi/m$. To consistently trace the trajectory of a spoke from the aperture rim toward the coordinate origin, it can therefore be advantageous to unwrap the initial angle θ_0.

Figure 8 illustrates the proposed definition of the angle θ_0^{fp}, shown in (a) as a function of the radius and in (b) as a polar plot over the circular aperture.

The dominant azimuthal orders ($m = 3$ and $m = 18$) of sample wavefront B are presented, including both the wrapped angle (solid line with markers) and the unwrapped angle (dashed line without markers). Since, for $m = 3$, the

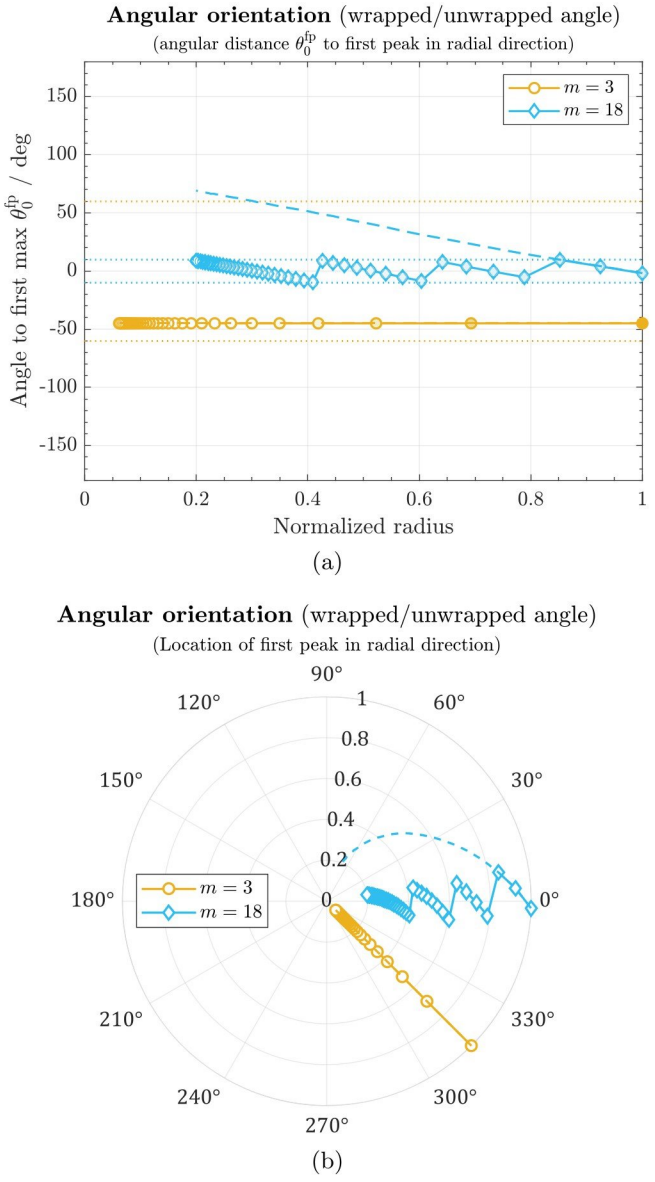


Fig. 8. Visualization of the proposed definition for azimuthal orientation. (a) Angle $\theta_0^{\mathrm{fp}}(r(n), m)$ over the radius where the first peak in radial direction occurs. (b) Polar plot of the same data.

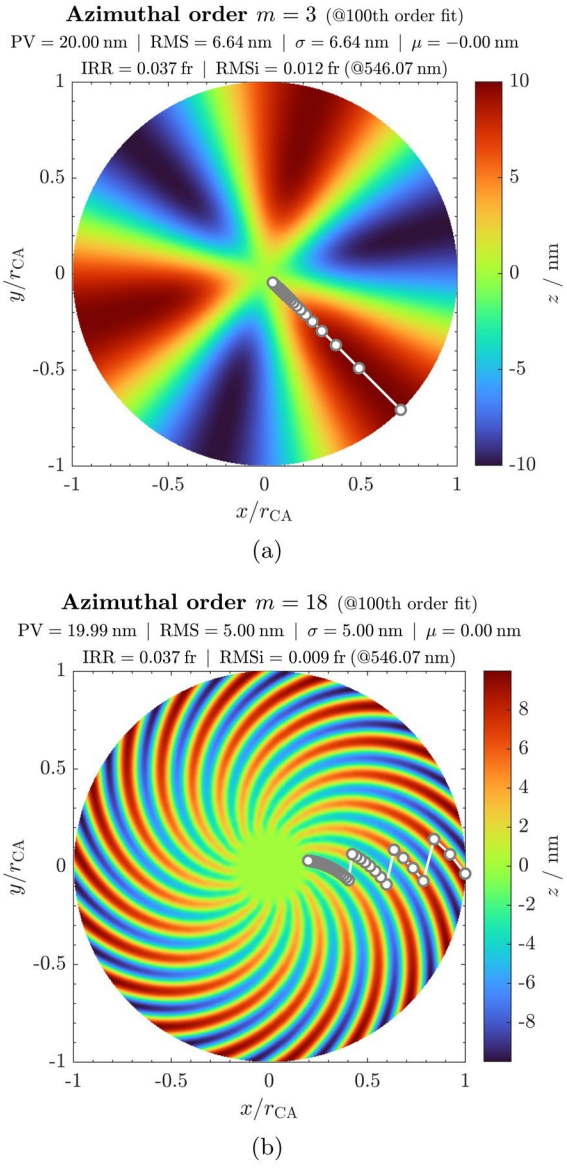


Fig. 9. Zernike-fit decomposition of the azimuthal orders of sample wavefront B, overlaid with the trajectory of the unwrapped angle θ_0^{fp}. A 100th-order Zernike fit was applied, with panel (a) showing the azimuthal order $m = 3$ and panel (b) showing $m = 18$.

wrapped angle never reaches the limit of uniqueness (indicated by the dotted line), the wrapped and unwrapped angles are here identical. The representation in Figure 8a reveals the radial dependence of the spoke bending. For $m = 3$, the bending is constant (straight spokes), whereas for $m = 18$ it exhibits a linear dependence on the radius. The polar plot in Figure 8b provides a convenient representation for tracing the trajectory of a single spoke of a given azimuthal order.

Finally, Figure 9 illustrates the fitted dominant orders (i.e., the sum of all Zernike terms with the same m), overlaid with the trajectory of the unwrapped angle θ_0^{fp}. Analyzing the shape and curvature of these spoke-like aberrations can therefore provide valuable insights into the underlying machine kinematics and support process optimization.

7 Summary

This paper presents a set of graphical visualization methods for results obtained from a Zernike approximation of wavefront aberrations or surface figure errors. These representations are intended to provide an intuitive visualization of the extensive information contained in the RMS-normalized Zernike coefficient vector. In particular, dominant characteristics are intended to be rapidly recognized and extracted.

Zernike bubble charts are well suited for representing low-order aberrations up to the 10th or 12th order. The area of each bubble is scaled according to the RMS value

of the corresponding Zernike aberration, such that the relative scaling has a clear physical meaning. These bubble charts are particularly valuable when rapid visual feedback on individual aberration terms is required, for example during aberration compensator adjustment, real-time wavefront visualization of dynamic processes, or optical alignment procedures. In addition, they can improve the clarity and interpretability of inspection reports or measurement certificates for lenses and optical systems.

Zernike heatmaps, by contrast, are well suited for representing mid- and high-spatial-frequency errors. For this purpose, the magnitude of the RMS-normalized coefficients is displayed on a logarithmic scale in order to enhance the visibility of contributions with small amplitudes. Such heatmaps reveal characteristic manufacturing artifacts, such as spoke-like structures or ripple patterns.

Furthermore, this work introduces an alternative definition of the azimuthal orientation of Zernike terms for $m \neq 0$. This approach enables the identification of the shape and bending of spoke-like aberrations. For such specific analyses, other numerical methods, such as Fourier-based filtering or multi-angle averaging, may also be employed [11]. However, given the availability of convenient numerical methods for fitting wavefronts or aberrations to very high Zernike orders [12], these features can now be directly extracted from the Zernike coefficients themselves.

Funding

This research did not receive any specific funding.

Conflicts of interest

The author declares no conflicts of interest.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. The MATLAB® code for visualizing Zernike coefficients (bubble charts, heatmaps) are publicly available on the MATLAB® File Exchange at the following URL: https://de.mathworks.com/matlabcentral/fileexchange/183649-zernike-visualization-heatmaps-and-bubble-plots.

Author contribution statement

This work was conducted and written solely by the author, who is fully responsible for all contributions and content.

References

1 Zernike F, Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode, *Physica* **1**(7), 689–704 (1934). https://doi.org/10.1016/S0031-8914(34)80259-5.
2 Malacara D (ed.), *Optical Shop Testing*, 2nd edn. (John Wiley & Sons, New York, 1992) https://doi.org/10.1002/9780470135976
3 Gross H, Dörband B, Müller H (eds.), *Handbook of Optical Systems, Volume 5: Metrology of Optical Components and Systems*, 1st edn. (Wiley-VCH, Weinheim, 2012) https://doi.org/10.1002/9783527699230.
4 Niu K, Tian C, Zernike polynomials and their applications, *J. Opt.* **24**(12), 123001 (2022). https://doi.org/10.1088/2040-8986/ac9c88.
5 Born M, Wolf E, *Principles of Optics*, 7th (expanded) edn. (Cambridge University Press, Cambridge, 1999).
6 Evans CJ, Parks RE, Sullivan PJ, Taylor JS, Visualization of surface figure by the use of Zernike polynomials, *Appl. Opt.* **34**(34), 7815–7819 (1995). https://doi.org/10.1364/AO.34.007815.
7 Reichelt S, Pruss C, Tiziani HJ, Absolute interferometric test of aspheres by use of twin computer-generated holograms, *Appl. Opt.* **42**(22), 4468–4479 (2003). https://doi.org/10.1364/AO.42.004468.
8 Webber JBW, A bi-symmetric log transformation for wide-range data, *Meas. Sci. Technol.* **24**(2), 027001 (2013). https://doi.org/10.1088/0957-0233/24/2/027001.
9 Forbes GW, Fitting freeform shapes with orthogonal bases, *Opt. Express* **21**(16), 19061–19081 (2013). https://doi.org/10.1364/OE.21.019061.
10 Hosseinimakarem Z, Davies AD, Evans CJ, Zernike polynomials for mid-spatial frequency representation on optical surfaces, *Proc. SPIE.* **9961**, 9961–18 (2016). https://doi.org/10.1117/12.2238514.
11 Reichelt S, Decomposition of non-rotationally symmetric wavefront aberrations into their azimuthal orders, in *Applied Optical Metrology III*, edited by E. Novak, J.D. Trolinger, (Proc. SPIE 11102, 2019), pp. 63–75. https://doi.org/10.1117/12.2526938.
12 Fan Y, Forbes GW, Rolland JP, Fast Zernike fitting of freeform surfaces using the Gauss-Legendre quadrature, *Opt. Express* **32**(11), 20011–20023 (2024). https://doi.org/10.1364/OE.498096.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 35
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026035
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE

OPEN ACCESS

Large resolution-weighted étendue space telescope designs categorized by tertiary mirror location diversity

Hyejeon Cho[1,2], Hyosun Park[1], M. James Jee[1,3], and Daewook Kim[4,5,6,*]

[1] Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[2] Center for Galaxy Evolution Research, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[3] Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
[4] Wyant College of Optical Sciences, University of Arizona, 1630 E. University Blvd., Tucson, AZ 85721, USA
[5] Department of Astronomy, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA
[6] Large Binocular Telescope Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA

Received 21 January 2026 / Accepted 31 March 2026

Abstract. Modern observational astronomy makes extensive use of deep, wide-field imaging and large spectroscopic surveys, defining telescope design spaces in which angular resolution, achievable depth, and survey efficiency are considered jointly rather than independently. Conventional étendue alone does not fully capture the efficiency with which resolved and deblended information can be acquired. We introduce a resolution-weighted étendue metric that provides a quantitative basis for comparing information throughput across different facilities. Applying this metric, we illustrate how accounting for angular resolution reshapes the relative placement of existing telescopes in design parameter space, with a representative space telescope design serving as an example of this shift. We consider two distinct reflective space telescope designs utilizing a variation of the three-mirror anastigmat configuration. The strategic placement of the tertiary mirror significantly influences the telescope's form factor, enabling compact architectures for very large aperture systems. These designs achieve large resolution-weighted étendue and imaging throughput, making them well-suited for precision cosmological and astrophysical measurements.

Keywords: Resolution-weighted étendue, Space telescope design, Tertiary mirror location, Astronomical image processing.

1 Introduction

Recent decadal surveys and community planning documents in astronomy and astrophysics show that current and future observational studies make extensive use of deep and wide-field imaging and large spectroscopic surveys across a broad range of wavelengths (e.g., [1, 2]). These approaches underpin precision cosmology, statistically robust studies of star and galaxy formation and evolution, time-domain astronomy, exoplanet detection and characterization, and related topics in astrophysics and fundamental physics. Achieving these science goals requires observations that deliver statistically powerful source samples together with precise, stable, and well-calibrated measurements, thereby constraining the parameter space for telescope design and optimization.

Ground-based wide-field facilities, such as the Subaru Telescope Hyper Suprime-Cam (HSC) [3] and the Vera C. Rubin Observatory Legacy Survey of Space and Time Camera (Rubin/LSST) [4], are optimized for large étendue—a combination of wide Field of View (FoV) and large photon collecting area—to enable efficient surveys over very large angular scales. Using a uniform, simplified definition of effective collecting area, their étendue values are of order tens to hundreds of m^2 deg^2. Space-based facilities occupy complementary regions of observational parameter space by providing stable, diffraction-limited performance. The James Webb Space Telescope (JWST) [5, 6], for example, prioritizes sensitivity and angular resolution over a relatively small FoV, yielding étendue values $\ll 1$ m^2 deg^2, whereas survey-oriented missions such as Euclid [7] and the Roman Space Telescope [8] target wide-area coverage at intermediate étendue.

Advances in launch-vehicle and spacecraft technologies have reduced space access costs and development barriers for small satellite missions [9]. Science cases at far-infrared wavelengths further illustrate how space telescope architectures can be driven toward large apertures, deployable structures, and active control [10]. These developments

* Corresponding author: dkim@optics.arizona.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

motivate a class of space telescope designs that extend established architectures toward larger photon collecting area and instantaneous FoV, while delivering high angular resolution, high photon throughput, and stable, spatially uniform Point Spread Functions (PSFs). Within this design space, optical performance, survey efficiency, and information yield are more tightly coupled. In this work, we introduce a practical metric for quantitative comparisons of information collection rates across facilities, and examine two illustrative optical system architectures and the science opportunities they enable.

2 Resolution-weighted étendue for space-based astronomy

Conventional étendue, $A\boldsymbol{\Omega}$, has been adopted as a first-order figure of merit for survey efficiency (often quantified as survey speed), particularly in the sky-noise-limited regime, where the sky area surveyed to a given depth per unit time scales with the photon collecting area A and the FoV $\boldsymbol{\Omega}$. However, $A\boldsymbol{\Omega}$ alone does not reflect the rate at which independent information is acquired at finite angular resolution. In particular, source blending and crowding impose a resolution-dependent limit on source detection and measurement such that increases in exposure time or étendue do not translate linearly into gains in effective depth or information acquisition efficiency.

Previous studies in survey telescope design and optimization have shown that survey efficiency depends not only on photon throughput but also on the PSF footprint (e.g., [11]). Building on this physical insight, we introduce a Resolution-Weighted Étendue (RWE) as a practical comparative metric. The RWE quantifies differences in the rate at which independent information is collected across telescopes by capturing both the light-gathering power per pointing and the delivered angular resolution.

In its most general form, the RWE may be written as

$$\mathcal{E}_{RW,\lambda} = \frac{A\Omega_{\mathrm{FoV}}}{\Omega_{\mathrm{PSF},\lambda}}, \qquad (1)$$

where Ω_{FoV} denotes the instantaneous FoV. To ensure RWE serves as a practical diagnostic tool of effective information throughput on an equivalent basis, we define $\Omega_{\mathrm{PSF},\lambda}$ as the solid angle of the delivered system-level PSF measured at a chosen wavelength λ. Since the PSF size is intrinsically tied to the observing wavelength, the resulting RWE is a band-specific metric. Consequently, a meaningful cross-comparison of different facilities requires evaluating them at consistent or closely overlapping spectral bands. For Gaussian, near-circular PSFs, $\Omega_{\mathrm{PSF},\lambda} \propto \vartheta^2_{\mathrm{PSF},\lambda}$, allowing the metric to be equivalently expressed in terms of the PSF Full Width at Half Maximum (FWHM) at the specified wavelength λ, $\vartheta_{\mathrm{PSF},\lambda}$ (typically in arcseconds).

In contrast to an idealized diffraction limited PSF such as Airy pattern for a circular aperture, $\boldsymbol{\Omega}_{\mathrm{PSF},\lambda}$ accounts for the end-to-end performance of the actual optical train and its operational environment[1]. This allows RWE to quantify the realistic information yield of a system and to penalize designs where theoretical resolution is degraded by practical constraints such as wavefront errors, pointing jitter, and detector-level sampling effects.

For consistent numerical comparisons across facilities, we further define a dimensionless RWE using a per-unit normalization as

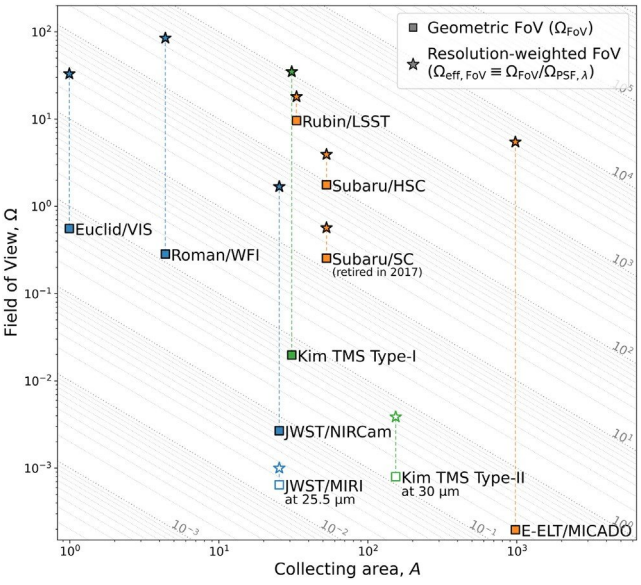


Figure 1. Étendue comparison for two distinct groups: optical/near-IR (filled symbols) and far-IR (open symbols) telescopes. The sample includes ground-based (orange) and space-based (blue and green) telescopes, selected to span a wide range of aperture sizes and FoVs, reflecting the diverse capabilities of current and next-generation facilities. Collecting area is evaluated assuming no field vignetting and excluding wavelength-dependent transmission, enabling a uniform, geometry-based comparison. The dotted gray contour lines show loci of constant étendue and serve as a visual reference for comparison. Standard étendue (squares) and RWE (stars) are projected onto a common FoV-based representation, enabling direct comparison on the same A–$\boldsymbol{\Omega}$ plane. Specifically, the square symbols are plotted at $(A, \boldsymbol{\Omega}_{\mathrm{FoV}})$, while the star symbols are plotted using an effective, resolution-weighted FoV, $\boldsymbol{\Omega}_{\mathrm{eff,FoV}} \equiv \boldsymbol{\Omega}_{\mathrm{FoV}}/\boldsymbol{\Omega}_{\mathrm{PSF},\lambda}$. The Kim TMS Type-I space telescope and the gigantic ground-based telescope E-ELT [12–14] exhibit substantially larger RWE, corresponding to greater depth and more resolved detail per pointing despite its smaller geometric FoV.

[1] For operational facilities (e.g., JWST), $\boldsymbol{\Omega}_{\mathrm{PSF},\lambda}$ refers to the measured on-sky PSF. For facilities that are in the design, development, or construction phase, it represents the predicted "delivered" PSF, which incorporates the nominal optical performance combined with field-dependent aberration balancing tradeoff and expected system-level error budgets to reflect realistic performance once operational. Thus, even for space telescopes, achieving a large $\boldsymbol{\Omega}_{\mathrm{FoV}}$ while maintaining a small $\boldsymbol{\Omega}_{\mathrm{PSF},\lambda}$ is a challenging optical design optimization.

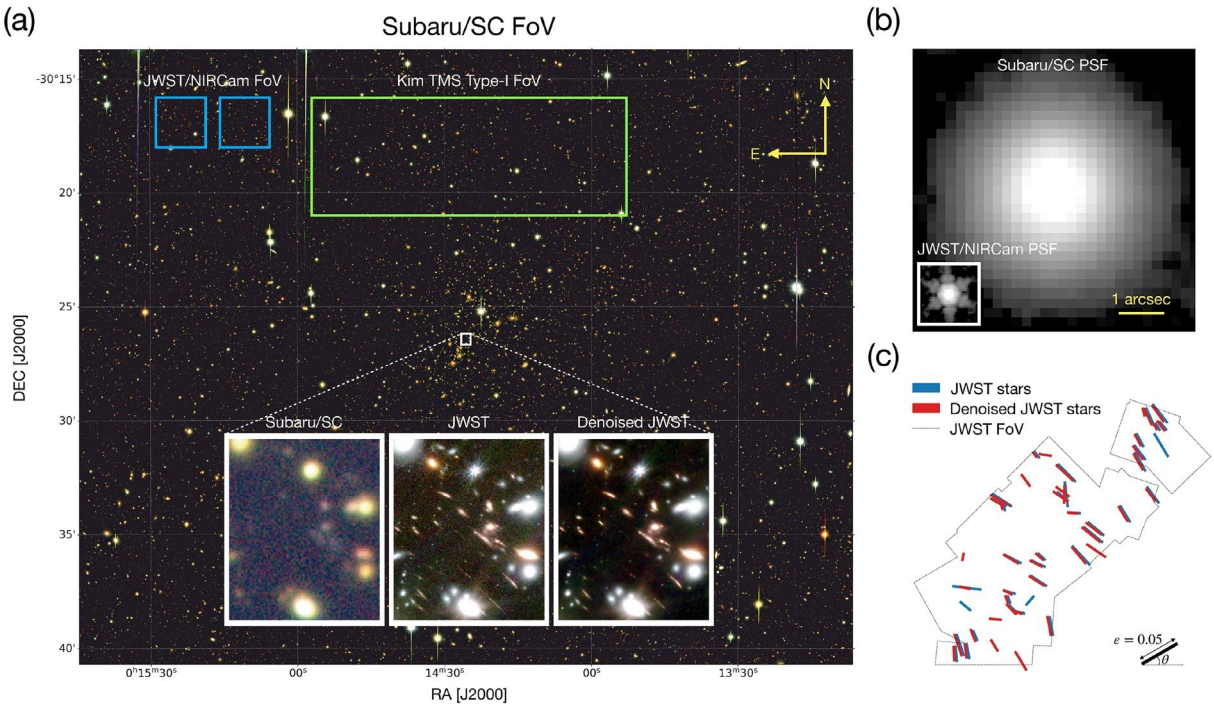


Figure 2. Comparisons of FoV, angular resolution, and PSF properties. (a) Subaru/SC image of the galaxy cluster Abell 2744 with the JWST/NIRCam [20] and Kim TMS Type-I FoV footprints overlaid. Insets show the same region observed with Subaru/SC and JWST/NIRCam, as well as in a "denoised" JWST image reconstructed using an efficient-Transformer network [21], approximating a deeper observation. (b) PSFs for Subaru/SC and JWST/NIRCam at the same angular scale and in logarithmic intensity. (c) Whisker plot of PSF ellipticity e and position angle θ across a JWST/NIRCam mosaic. Whiskers derived from the "denoised" image, shifted by a few pixels for clarity, retain the same pattern as the original, indicating that the field-dependent PSF variations remain even after image denoising.

$$\tilde{\mathcal{E}}_{RW,\lambda} = \left(\frac{A}{1\ \mathrm{m}^2}\right)\left(\frac{\Omega_{\mathrm{FoV}}}{1\ \mathrm{deg}^2}\right)\left(\frac{\Omega_{\mathrm{PSF},\lambda}}{1\ \mathrm{arcsec}^2}\right)^{-1}. \qquad (2)$$

Here, $\Omega_{\mathrm{PSF},\lambda}$ is expressed in arcsec^2, as commonly used in the literature for reporting PSF or beam sizes in arcseconds.

The RWE becomes especially important for space telescopes capable of achieving diffraction-limited performance over large FoVs. Figure 1 presents both standard and resolution-weighted étendues for two distinct comparison groups: optical/near-IR ground- and space-based facilities (filled symbols) and far-IR space-based facilities (open symbols). Under the conventional étendue definition, wide-field ground-based survey telescopes occupy the large-étendue regime, whereas space telescopes—and the next generation of extremely large-aperture telescopes (e.g., the European Extremely Large Telescope; E-ELT [15, 16]) as a notable exception that achieves near diffraction-limited performance from the ground despite its limited FoV—cluster at values of order unity or below. When diffraction-limited PSFs are taken into account through the RWE, this comparison is reframed, placing these high-resolution facilities within—and in some cases extending beyond—the region of design and optimization parameter space occupied by wide-field ground-based telescopes. The Kim Three-Mirror System (TMS) Type-I space telescope [17] and the ground-based E-ELT provide a representative example of this shift. Its large aperture and diffraction-limited PSF compensate for the relatively smaller FoV, yielding RWE values comparable to those of ground-based wide-field survey facilities such as Rubin/LSST.

In Figure 1, $\Omega_{\mathrm{PSF},\lambda}$ values for each facility are based on the delivered FWHM at representative spectral bands as defined above. For JWST, we utilize data from NIRCam F070W (0.70 μm; [5]) and MIRI F2550W (25.5 μm; [18]) to align with the nominal wavelengths of the Kim TMS Type-I and Type-II configurations discussed in Section 3, respectively. For Kim TMS Type-I and Type-II, the PSF FWHM is evaluated at the detector focal plane based on the reported nominal optical performance [17, 19]. In the absence of measured on-orbit jitter data, these values serve as the best-available estimate for the delivered PSF, providing a rigorous yet realistic benchmark for the RWE comparison. Minor variations in detector sampling or other system-level disturbances are treated as independent factors and do not affect the qualitative conclusions of this comparison.

The very large RWE of Kim TMS Type-I/II and E-ELT translates into increased survey depth, while providing a higher degree of resolved detail per pointing. This combination enables more efficient source detection within limited observing time and supports uniform sky coverage. Realizing these benefits, however, requires stringent control of temporal and spatial variations of the PSF across the

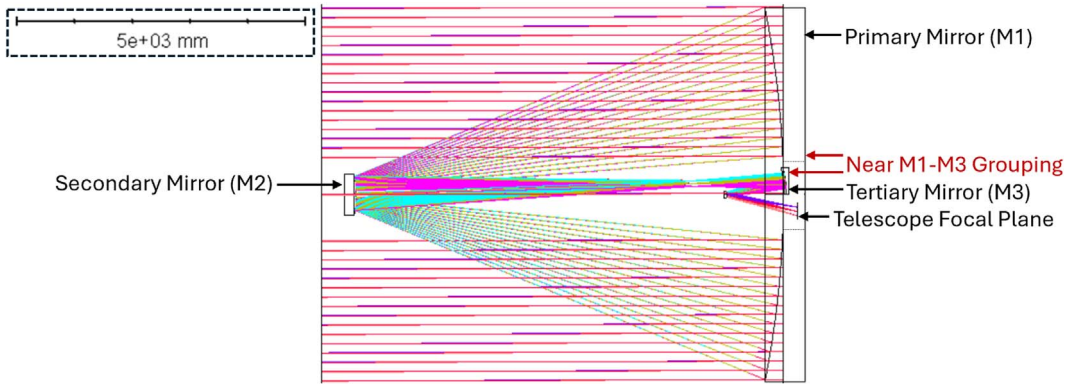


Figure 3. Optical design and ray-tracing of the Kim TMS Type-I configuration for a 6.5-m diameter compact telescope (adapted from [17]). The co-location of the active M1 and the M3 enables a very short total length of the telescope along the optical axis.

FoV, which is crucial for precision observational analyses as well as for space adaptive optics utilizing natural guide stars.

Figure 2 compares ground-based Subaru Suprime-Cam (SC) and space-based JWST/NIRCam observations, highlighting differences between wide-field, seeing-limited imaging and diffraction-limited imaging with a smaller FoV. We also consider a "denoised" diffraction-limited image reconstructed using an efficient-Transformer-based network [21]. We find that post-processing techniques such as deep-learning-based image restoration improve apparent image quality but do not mitigate the underlying spatial variations of the PSF in our data. These patterns limit PSF deconvolution accuracy, thereby impacting shape and photometry measurements and source deblending. As RWE increases, controlling PSF uniformity across the FoV becomes increasingly important at the optical design stage, with limitations that are difficult to fully address through post-processing alone.

3 Three mirror system categories based on tertiary mirror location diversity

Three-Mirror Anastigmat (TMA) optical systems and their variations—including those adopted for ground-based telescopes such as CMB-S4 [22, 23] and E-ELT and space-based survey facilities such as Euclid and Roman—span a broad range of FoV and aperture sizes, associated with different observational science priorities. The image quality in such designs is governed by well-known optical scalings: geometrical aberrations generally increase with aperture diameter D, while the diffraction-limited angular resolution, often characterized by the Airy disk size, scales inversely with D. These competing fundamental relationships provide context for considering alternative telescope design architectures in the RWE framework.

Beyond their optical performance, the physical configuration, including the size and relative placement of mirrors, profoundly influences the overall telescope architecture and its concept of operations. The strategic location of the tertiary mirror (M3) can enable diverse telescope functionalities. The M3 location relative to the primary (M1) and secondary (M2) mirrors offers novel telescope form factors. These configurations are particularly well-suited for achieving large RWE, a crucial characteristic for space telescopes designed for the demands of the era of precision cosmology and astrophysics utilizing big data.

In the Kim TMS Type-I configuration, the 6.5-m M1 is near M3. This M1–M3 grouping results in a highly compact optical design, with a total length along the optical axis of less than approximately 8 m, as illustrated in Figure 3 [17]. This design approach enables the packaging of a large-aperture telescope within a rocket's payload volume housed in its fairing. Furthermore, M1 is designed as an active deformable mirror, allowing real-time compensation of shape changes induced by gravitational and/or thermal deformations.

As another M3 location diversity option (Kim TMS Type-II), to facilitate an extremely large, deployable M1, such as an inflatable design, the remaining optical components (M2 and M3) are strategically positioned on the opposite side of the telescope from M1, as indicated by the dashed box in Figure 4 [19]. This configuration specifically incorporates a low-order deformable M3 grouped with M2. This M2–M3 grouping allows independent assembly and testing of the downstream optical train, enabling integration with a large, deployable primary. A comparative summary of the specific telescope design parameters and optical performance for both Kim TMS Type-I/II configurations is presented in Table 1.

In addition to the traditional large étendue telescope design parameter space, large RWE space telescopes must consider their form factors to realize large aperture sizes during launch and/or deployment on orbit. The tertiary mirror location diversity becomes an important optical design category and optimization parameter.

4 Science opportunities enabled by large-RWE space telescopes

Large-RWE facilities maximize scientific return by combining survey depth with stable, high-resolution imaging. This architecture minimizes source confusion and enhances measurement fidelity, enabling the following key applications.

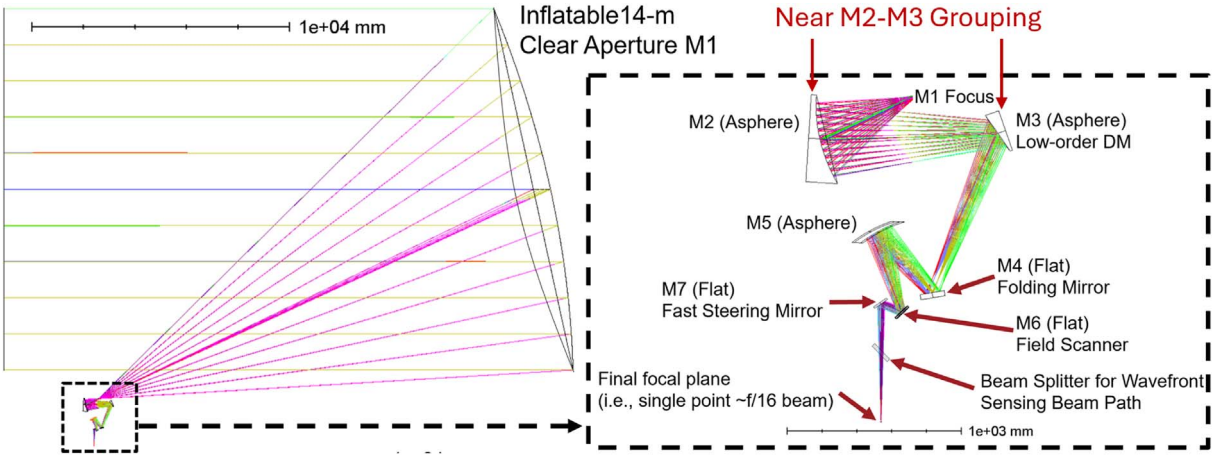


Figure 4. Optical design and ray-tracing of the off-axis Kim TMS Type-II configuration (adapted from [19]). The 14-m inflatable M1 is positioned far from the other optics (shown in the dashed box), allowing for independent deployment and inflation while the M2–M3 component can be assembled, tested, and packaged as a subsystem.

Table 1. Comparison of Kim TMS Type-I and Type-II telescope specifications.

Parameter	Kim TMS Type-I	Kim TMS Type-II
Reference	[17]	[19]
Nominal wavelength (μm)	0.650	30
Primary mirror diameter (m)	6.420	14
Central obscuration	On-axis optical design with 1.38 m inner diameter on the primary mirror	Off-axis optical design without obscuration
Full field of view (deg)*	$\pm 0.115 \times 0.043°$	$\pm 0.02 \times 0.02°$
Primary mirror surface shape	Ellipse (on-axis conic surface)	Parabola (off-axis conic surface)
Primary mirror type	Rigid mirror	Inflatable mirror
Telescope $F/\#$	15	16
Airy disk radius (μm)	11.97 (at 0.650 μm wavelength)	585 (at 30 μm wavelength)
Nominal strehl ratio†	∼0.90–0.99	∼0.82–0.94

*Refer to the cited references for detailed FoV descriptions and focal plane definitions.
†Refer to the corresponding references for further details regarding field-dependent performance.

Precision cosmology with weak and strong gravitational lensing: A large aperture space telescope with a stable PSF significantly enhances gravitational lensing analyses. In weak lensing, reduced blending increases the effective source number density and improves shape measurement accuracy and cosmic shear constraints. For strong lensing, superior resolution facilitates the identification of a much larger number of multiple-image systems and enhances sensitivity to lensing substructures, enabling unified weak-and-strong lensing analyses from homogeneous datasets.

Galaxy evolution across cosmic time and environment: Resolving internal galaxy structures (e.g., morphology, color gradients, stellar mass distributions) is critical for evolution studies but often degraded by blending, particularly in dense regions. Large RWE surveys reduce confusion noise, ensuring accurate photometry and classification for statistically powerful samples across diverse environments and redshifts.

Low-surface-brightness structures: Features such as intracluster light, tidal streams, and ultra diffuse galaxies are highly sensitive to PSF wings and spatial non-uniformity. A space telescope with large RWE and smooth, well-characterized PSFs enables robust separation of diffuse emission from compact sources across wide areas. Its improved resolution reduces contamination from unresolved background objects and allows accurate modeling of extended structures, providing insights into the halo assembly histories.

Complementarity with existing facilities: Large-RWE space missions fill the gap between seeing-limited ground-based surveys and narrow-field space telescopes. They

provide large, statistically robust samples, support calibration of survey measurements, and enhance the scientific return of the broader observational ecosystem.

5 Conclusion

We introduced RWE as a complementary and practical metric for characterizing resolved and deblended information collection efficiency across facilities spanning different FoV, apertures, and resolutions. Within this framework, two distinctly different tertiary mirror placement strategies were investigated. The Kim TMS Type-I design leads to a highly compact overall telescope envelope, ideal for applications requiring a minimized form factor. The Kim TMS Type-II design proves advantageous for enabling the architecture of telescopes with deployable, extremely large primary mirrors. The strategic placement of the tertiary mirror enables large RWE by linking aperture, resolution, and PSF uniformity and stability in wide-field space telescopes.

Acknowledgments

This work is based in part on data collected at Subaru Telescope and obtained from the SMOKA, which is operated by the Astronomy Data Center, National Astronomical Observatory of Japan. This work is based in part on observations made with the NASA/ESA/CSA JWST. The data were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations are associated with programs GO-2561, ERS-1324, and DD-2767. We made use of the publicly released UNCOVER Data Release 2 reduced NIRCam mosaics (F200W, F277W, and F444W) [20]. H.C. appreciates useful discussions with John P. Blakeslee and acknowledges support from the National Research Foundation of Korea (NRF) grants funded by the Korea government (MOE, RS-2022-NR070872; and MSIT, RS-2022-NR070525). M.J.J. acknowledges support for the current research from the NRF under the programs 2022R1A2C1003130 and RS-2023-00219959.

Funding

This research did not receive any specific funding.

Conflicts of interest

The authors declare that they have no competing interests to report.

Data availability statement

The research data are available on request from the authors.

Author contribution statement

Conceptualization, H.C. and D.K.; Formal Analysis, H.C. and H.P.; Methodology, H.C., M.J.J., and D.K.; Project Administration, D.K.; Supervision, D.K.; Visualization, H.C. and H.P.; Writing—Original Draft Preparation, H.C., H.P., M.J.J., and D.K.; Writing—Review & Editing, H.C., H.P., M.J.J., and D.K.

References

1 *National Academies of Sciences, Engineering, and Medicine, Pathways to Discovery in Astronomy and Astrophysics for the 2020s.* (The National Academies Press, Washington, DC, 2023). https://doi.org/10.17226/26141.
2 Saintonge A, Andersen AC, Catala C, Stark R, Editors. *The ASTRONET Science Vision and Infrastructure Roadmap 2022–2035* (ASTRONET, 2023). ISBN: 978-1-3999-5162-3. https://www.astronet-eu.org/.
3 Miyazaki S et al., Hyper Suprime-Cam: System design and verification of image quality, *Publ. Astron. Soc. Jpn.* **70** (SP1), S1 (2018). https://doi.org/10.1093/pasj/psx063.
4 Ivezić Ž et al., LSST: From science drivers to reference design and anticipated data products, *Astrophys. J.* **873**, 111 (2019). https://doi.org/10.3847/1538-4357/ab042c.
5 Rigby J et al., The science performance of JWST as characterized in commissioning, *Publ. Astron. Soc. Pac.* **135**, 048001 (2023). https://doi.org/10.1088/1538-3873/acb293.
6 McElwain MW et al., The James Webb Space Telescope mission: Optical telescope element design, development, and performance, *Publ. Astron. Soc. Pac.* **135**, 058001 (2023). https://doi.org/10.1088/1538-3873/acada0.
7 Euclid Collaboration et al., Euclid: I. Overview of the Euclid mission, *Astron. Astrophys.* **697**, A1 (2025). https://doi.org/10.1051/0004-6361/202450810.
8 Schlieder JE et al., Survey science with the Nancy Grace Roman Space Telescope wide field instrument, *Proc. SPIE.* **13092**, 130920S (2024). https://doi.org/10.1117/12.3020622.
9 Douglas ES et al., Approaches to lowering the cost of large space telescopes, *Proc. SPIE.* **12677**, 126770D (2023). https://doi.org/10.1117/12.2677843.
10 Chin G et al., Single aperture large telescope for universe studies: Science overview, *J. Astron. Telesc. Instrum. Syst.* **10**(4), 042310 (2024). https://doi.org/10.1117/1.JATIS.10.4.042310.
11 Tyson JA, Large Synoptic Survey Telescope: Overview, *Proc. SPIE.* **4836**, 10–20 (2002). https://doi.org/10.1117/12.456772.
12 Davies R et al., MICADO: the E-ELT adaptive optics imaging camera, *Proc. SPIE.* **7735**, 77352A (2010). https://doi.org/10.1117/12.856379.
13 Davies R et al., MICADO: first light imager for the E-ELT, *Proc. SPIE.* **9908**, 99081Z (2016). https://doi.org/10.1117/12.2233047.
14 Sturm E et al., The MICADO first light imager for the ELT: overview and current status, *Proc. SPIE.* **13096**, 1309611 (2024). https://doi.org/10.1117/12.3017752.
15 European Southern Observatory, The E-ELT Construction Proposal (ESO, 2011). https://www.eso.org/public/products/books/book_0046/.
16 European Southern Observatory, E-ELT Programme – Observatory Top Level Requirements, E-SPE-ESO-100-0260, Issue 2 (2012). https://www.eso.org/sci/facilities/eelt/docs/ESO-193696_2_Observatory_Top_Level_Requirements.pdf.
17 Kim D, Choi H, Douglas E, Compact three mirror anastigmat space telescope design using 6.5 m monolithic primary mirror, *Proc. SPIE.* **12677**, 126770E (2023). https://doi.org/10.1117/12.2682180.
18 Dicken D et al., JWST MIRI flight performance: Imaging, *Astron. Astrophys.* **689**, A5 (2024). https://doi.org/10.1051/0004-6361/202449451.

19 Kim D et al., Fourteen-meter aperture deployable off-axis far-infrared space telescope design for SALTUS observatory, *J. Astron. Telesc. Instrum. Syst.* **10**(4), 042309 (2024). https://doi.org/10.1117/1.JATIS.10.4.042309.
20 Bezanson R et al., The JWST uncover treasury survey: Ultradeep NIRSpec and NIRCam observations before the epoch of reionization, *Astrophys. J.* **974**, 92 (2024). https://doi.org/10.3847/1538-4357/ad66cf.
21 Park H et al., Deeper, sharper, faster: application of efficient transformer to galaxy image restoration, *Astrophys. J.* **972**, 45 (2024). https://doi.org/10.3847/1538-4357/ad5954.
22 Patricio A. Gallardo et al., Freeform three-mirror anastigmatic large-aperture telescope and receiver optics for CMB-S4, *Appl. Opt.* **63**, 310–321 (2024). https://doi.org/10.1364/AO.501744.
23 Patricio A. Gallardo et al., Overview of the optical design of the CMB-S4 large aperture telescopes and camera optics, *Proc. SPIE.* **13094**, 130942F (2024). https://doi.org/10.1117/12.3020608.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 19
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026011
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE OPEN ACCESS

Propagation of partially coherent fields radiated by sources with univariable cross-spectral density

Massimo Santarsiero[1], Rosario Martínez-Herrero[2], Juan Carlos González de Sande[3], Gemma Piquero[2,*], Olga Korotkova[4], and Franco Gori[1]

[1] DIIEM, Università Roma Tre, via V. Volterra 62, Rome 00146, Italy
[2] Departamento de Óptica, Fac. CC. Físicas, U.C.M., Ciudad Universitaria s/n, 28040 Madrid, Spain
[3] Universidad Politécnica de Madrid, ETSIS de Telecomunicación, Campus Sur, 28031 Madrid, Spain
[4] Department of Physics, University of Miami, 1320 Campo Sano Drive, Coral Gables, FL, 33146, USA

Received 18 December 2025 / Accepted 4 February 2026

Abstract. A new class of partially coherent light sources, the sources with uni-variable cross-spectral density (CSD), has been recently introduced. Their CSD is obtained starting from any function of a single complex argument having non-negative Taylor coefficients. This allows the conception of a virtually infinite number of physically realizable partially coherent sources. Here, the main characteristics of sources of this class are investigated through examples, with particular reference to the irradiance and coherence properties across the source plane and upon propagation, both in the near and in the far field. Furthermore, since the coherent modes of such sources present optical vortices, parameters quantifying the vortex structure of the field across the source plane are also evaluated for the presented cases.

Keywords: Coherence, Coherent modes, Propagation, Optical vortices, Orbital angular moment, Structured coherence.

1 Introduction

Structured light has attracted strong interest in recent years, largely driven by its wide range of applications and by rapid progress in generating and detecting tailored optical fields [1, 2]. Within this area, partially coherent sources are increasingly regarded not as a limitation but as a useful degree of freedom, since the engineering of spatial coherence can mitigate drawbacks of highly coherent beams. Recent reviews emphasize that partially coherent beams with structured coherence have been explored for concrete applications including: optical imaging [3, 4], free-space optical communications with improved robustness to atmospheric turbulence [5–8], coherence-based optical encryption and robust signal/information transmission through complex or scattering media [8, 9], optical manipulation [10, 11], and remote sensing [12]. These are some of the reasons why there is great interest in devising, realizing, and characterizing new light sources, both coherent and partially coherent, that allow the characteristics of the light radiated by these sources to be modified in a controlled way [13–38].

Interesting and promising results are derived from the presence of phase vortices, which confer unique capabilities to the beam radiated from the source [5, 6, 12, 39–54]. For example, they find application in optical manipulation, because they can exert torque on microscopic particles, enabling advanced optical tweezers and micromanipulation [10, 46]; in imaging and metrology, where they offer new possibilities for super-resolution or phase-sensitive detection [12, 44]; in optical communication, since they provide an extra degree of freedom, the orbital angular momentum (OAM), for multiplexing and increasing data capacity [5, 6]. Vortex beams are typically generated using elements such as spiral phase plates, computer-generated holograms, q plates, or metasurfaces that encode the desired OAM mode [47, 50, 51, 53].

Recently, a new type of partially coherent light source, namely, a source with uni-variable cross-spectral density (CSD), has been proposed [38]. The peculiarity of sources of this class is that their CSD [55] can be directly derived from a function of a single complex variable [38]. The only requirement this function has to fulfill in order for it to be associated with a well-defined CSD is that its Taylor expansion must involve only non-negative coefficients. Several analytical forms, in some cases very simple, of bona-fide CSD can be devised in this way. Sources with uni-variable CSD are defined within a finite circular region, related to the convergence domain of the above Taylor series, and

* Corresponding author: piquero@ucm.es

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

can be expressed as the superposition of coherent modes [55] carrying optical vortices. The analytical form of the vortex modes is quite simple, both across the source plane and in the far zone, enabling the study of the coherence features of the source, as well as those of the field radiated under Fraunhofer conditions [38]. One of the important features of univariable CSDs is their capability to customize at will the contributions from various coherent modes, serving, at the same rate, as the OAM modes. This renders them ideal for high-capacity information-transfer applications employing structured light.

The main aim of the present paper is, on one hand, to extend to the Fresnel regime the study of the propagation of beams radiated from sources with uni-variable CSD; on the other hand, to study the effects of truncating the number of the modes involved in the expansion. Examples will be presented concerning the so-called sZegö source, introduced in [38], which is obtained with an infinite number of Taylor terms, all with the same coefficient. Taking only a finite number of terms gives rise to the truncated sZegö source, introduced here, which represents a more realistic model for a partially coherent source, suitable to be physically realized. Finally, hints for devising other types of uni-variable CSDs are given for some cases in which the CSD takes a simple closed form.

2 Preliminaries

2.1 Uni-variable CSDs

A uni-variable CSD [38] is defined from any function g of a single complex variable ζ such that:

$$g(\zeta) = \sum_{n=0}^{\infty} c_n \zeta^n, \tag{1}$$

where

$$c_n \geq 0 \, (\forall n), \tag{2}$$

and ζ belongs to the convergence domain of the series. By suitably scaling the argument of g it is always possible to set such a domain as

$$|\zeta| < 1. \tag{3}$$

The CSD associated with g is obtained by letting

$$\zeta = \frac{r_1 r_2}{r_0^2} e^{i(\varphi_1 - \varphi_2)}, \tag{4}$$

where (r_1, φ_1) and (r_2, φ_2) are the polar coordinates of two points across the source plane ($\boldsymbol{r_1}$ and $\boldsymbol{r_2}$) and r_0 is the radius of the circular region where the source is defined. It should be noted that although the CSD depends on two points, for this class of sources, the dependence enters only through the single complex parameter ζ, following the terminology introduced previously in reference [38]. Therefore, using equations (1) and (4) we define, across the plane $z = 0$,

$$\begin{aligned} W(\boldsymbol{r_1}, \boldsymbol{r_2}, 0) &= g\left(\frac{r_1 r_2}{r_0^2} e^{i(\varphi_1 - \varphi_2)}\right) \\ &= \sum_{n=0}^{\infty} c_n \left(\frac{r_1 r_2}{r_0^2}\right)^n e^{in(\varphi_1 - \varphi_2)}, \\ (|r| < r_0). \end{aligned} \tag{5}$$

The sum in equation (5) can be read as the Mercer expansion of W_0, that is, [55]

$$W(\boldsymbol{r_1}, \boldsymbol{r_2}, 0) = \sum_{n=0}^{\infty} \lambda_n \Phi_n(\boldsymbol{r_1}, 0) \Phi_n^*(\boldsymbol{r_2}, 0), \tag{6}$$

with coherent modes (taking into account their normalization) given by

$$\begin{aligned} \Phi_n(\boldsymbol{r}, 0) &= \sqrt{\frac{n+1}{\pi r_0^2}} \left(\frac{r}{r_0}\right)^n \mathrm{circ}\left(\frac{r}{r_0}\right) e^{in\varphi}, \\ (n = 0, 1, ...), \end{aligned} \tag{7}$$

and eigenvalues

$$\lambda_n = c_n \frac{\pi r_0^2}{n+1}, \quad (n = 0, 1, ...), \tag{8}$$

where $\mathrm{circ}(r)$ is the characteristic function of a unit disk centered at the origin of the axes. The second argument of the modes Φ_n refers to the z-coordinate of the transverse plane where they are evaluated (in this case, across the source, that is, $z = 0$). The condition in equation (2), together with equation (8), guarantees that $\lambda_n > 0$ $(\forall n)$, so that the CSD turns out to be well defined [13, 14].

The modes in equation (7) are recognized at once as optical vortices [49] with charge n. They have the same structure as Laguerre–Gaussian beams [56] in which the Laguerre polynomial has order zero and where the Gaussian function has been replaced by a circ.

The number of genuine CSDs that can be obtained following the present approach is practically unlimited, the only constraint that the function g has to fulfill being the fact that the coefficients of its Taylor expansion are all non-negative. Examples will be given in the following, where simple forms of CSDs will be presented, together with the mean features of the beams they radiate.

2.2 Characterization of uni-variable CSDs in OAM space

Uni-variable CSD represents one of a few known model optical fields that carry OAM in multiple modes, either finite or infinite. For the characterization of radial-only correlations among the OAM modes in such fields, the coherence–orbital angular momentum (COAM) matrix $\overleftrightarrow{W}(\rho_1, \rho_2, z)$ and its derivatives can be used [57]. Each element of the COAM matrix is a scalar radial field correlation at a pair of OAM indices, given by the expression:

$$W_{\ell m}(r_1, r_2, z) = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} W(r_1, r_2, z)e^{-i(\ell\varphi_1 - m\varphi_2)}\mathrm{d}\varphi_1\mathrm{d}\varphi_2, \quad -\infty < \ell, m < \infty. \tag{9}$$

The COAM matrix is the counterpart of the CSD matrix defined in polarization space to the OAM space and carries similar properties (quasi-Hermiticity, non-negative definiteness, etc.). However, it can have any size, say $L \times L$, where L is the largest by magnitude OAM index involved. Also, for theoretical models, infinitely dimensional matrices are possible with certain convergence constraints. The uni-variable beam model is of this type.

Two derivatives of the COAM matrix are important and will be considered here. First, the OAM degree of coherence is the analog of the EM degree of coherence in the OAM space [58]:

$$o(r_1, r_2, z) = \frac{\sqrt{\mathrm{Tr}[\overleftrightarrow{W}^{\dagger}(r_1, r_2, z)\,\overleftrightarrow{W}(r_1, r_2, z)]}}{\prod\limits_{j=1,2}\sqrt{\mathrm{Tr}\,\overleftrightarrow{W}(r_j, r_j, z)}}. \tag{10}$$

It provides the cumulative information about the normalized correlations at a pair of radii for all pairs of the OAM modes, regardless of the OAM phase information contained in each mode.

Second, it was found that a harmonic beam carrying OAM in L modes traces an orbitalization ellipse in LD OAM space (polar Fourier space) at a fixed radius [59], analogously to the two/three component electric field tracing a polarization ellipse in real 2D/3D space. To extend this geometric description to partially coherent beams, the COAM matrix can be considered at a single radius $r_1 = r_2 = r$, and regarded as the $L \times L$ orbitalization matrix [60]:

$$\overleftrightarrow{O}(r, z) = \overleftrightarrow{W}(r, r, z), \tag{11}$$

the term steamming from its formal similarity to a polarization matrix. In LD the orbitalization matrix is shown to decompose into three parts: completely orbitalized, unorbitalized, and partially orbitalized, in which mixing occurs for mode couples, triples, …, up to $L - 1$. The completely orbitalized matrix is shown to factorize, representing a beam indistinguishable from harmonic, and, hence, described by an orbitalization ellipse. The contribution of the intensity of the completely orbitalized portion of the beam in the total beam, at a fixed radius, was introduced as the degree of orbitalization [60]:

$$O(r, z) = \frac{\sigma_0(r, z) - \sigma_1(r, z)}{\sum\limits_{n=0}^{\infty}\sigma_n(r, z)}, \tag{12}$$

where $\sigma_0(r,\ z) \geq \sigma_1(r,\ z) \geq \ldots \geq \sigma_{\mathrm{L}}(r,\ z) \geq \ldots$ is the ordered sequence of the eigenvalues of $\overleftrightarrow{O}$, all being real and nonnegative due to its Hermiticity and non-negative-definiteness. This quantity is analogous in nature to the degree of polarization and reduces to it in form for $L = 2$ and 3. In this work, we will thoroughly examine its behavior for uni-variable beams.

We note that for a univariable CSD the COAM matrix is diagonal and hence the calculations of both derivatives are relatively simple. For example, the OAM DOC becomes

$$o(r_1, r_2, 0) = \frac{\sqrt{\sum\limits_{n=0}^{\infty} c_n^2 (r_1 r_2 / r_0^2)^{2n}}}{\prod\limits_{j=1,2}\sqrt{\sum\limits_{n=0}^{\infty} c_n (r_j / r_0)^n}}. \tag{13}$$

As we show below, for uni-variable beams σ_0 and σ_1 may depend on different c_n at different radii and, hence, it appears impossible to write down an explicit single formula for the degree of orbitalization. Instead, a piecewise-continuous function will be used, with its interval endpoints determined numerically.

3 Beams radiated by uni-variable CSDs

The propagation from sources with uni-variable CSD, both in the near and in the far field, can be dealt with starting from their expansion in terms of coherent modes. In such a way, from equation (6) we have, at any plane $z = \text{const.}$,

$$W(\boldsymbol{r_1}, \boldsymbol{r_2}, z) = \sum_{n=0}^{\infty}\lambda_n \Phi_n(\boldsymbol{r_1}, z)\Phi_n^*(\boldsymbol{r_2}, z), \tag{14}$$

where the expression of the propagated modes $\Phi_{\mathrm{n}}(\boldsymbol{r},\ z)$ depends on the approximation we use.

Using the expression above, the spectral density and the degree of coherence of the radiated field can be evaluated as [55]

$$S(\boldsymbol{r}, z) \equiv W(\boldsymbol{r}, \boldsymbol{r}, z) = \sum_{n=0}^{\infty}\lambda_n |\Phi_n(\boldsymbol{r}, z)|^2, \tag{15}$$

and

$$\mu(\boldsymbol{r}_1, \boldsymbol{r}_2, z) \equiv \frac{W(\boldsymbol{r}_1, \boldsymbol{r}_2, z)}{\sqrt{S(\boldsymbol{r}_1, z)\,S(\boldsymbol{r}_2, z)}}, \tag{16}$$

respectively.

3.1 Coherent modes in paraxial regime

Within the paraxial approximation, the propagation of the modes can be studied through the Fresnel integral, which, for a typical field $V(r,\ \varphi,\ z)$, in polar coordinates reads [55]

$$V(r, \varphi, z) = -\frac{ik\mathrm{e}^{\mathrm{i}kz}}{2\pi z}\int_0^{\infty}\int_0^{2\pi} V(r', \varphi', 0)e^{\frac{ik}{2z}\left[r'^2 + r^2 - 2r'r\cos(\varphi' - \varphi)\right]} r'\mathrm{d}r'\mathrm{d}\varphi'. \tag{17}$$

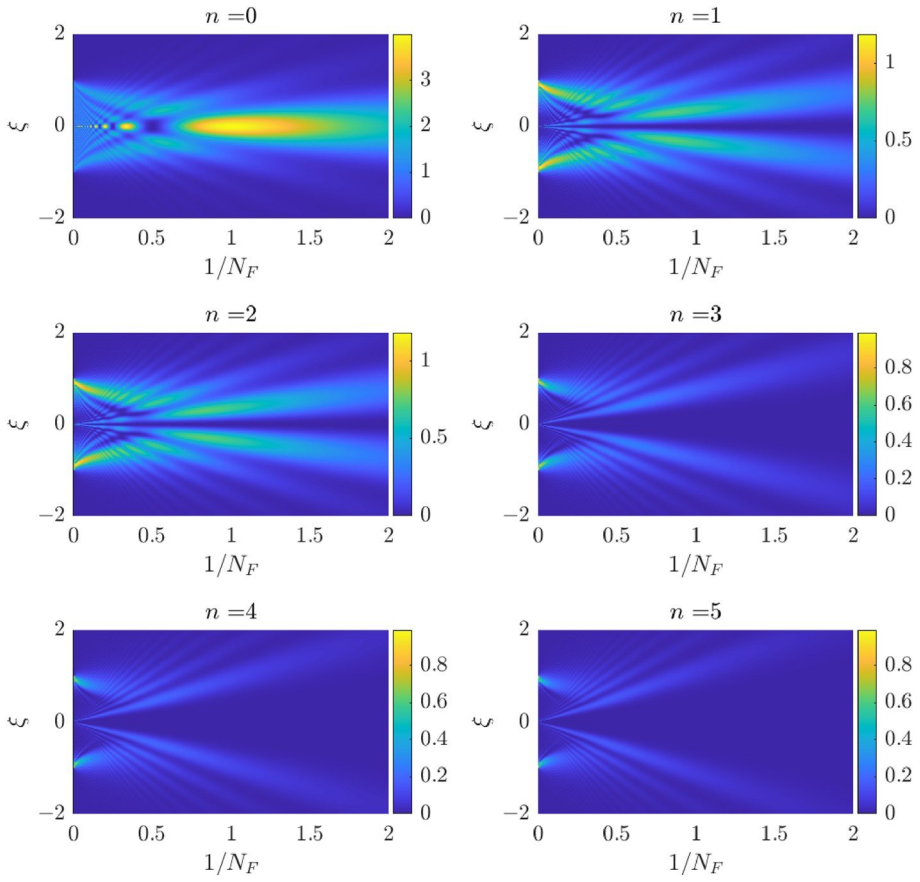


Fig. 1. Density plots of the spectral density of some modes, $|\Phi_n(\boldsymbol{r}, z)|^2$, across the plane (ξ, z) for $n = 0, 1, 2, 3, 4, 5$.

Using the equality

$$J_n(t) = \frac{1}{2\pi}\int_0^{2\pi} e^{in\psi - it\sin\psi}\mathrm{d}\psi, \tag{18}$$

after some manipulations the propagated modes take the following form:

$$\Phi_n(\boldsymbol{r}, z) = \frac{2N_F}{r_0}\sqrt{\pi(n+1)}e^{ikz}(-i)^{n+1} \times e^{in\varphi}e^{i\pi N_F\rho^2} \times \int_0^1 J_n(2\pi N_F\rho u)e^{i\pi N_F u^2}u^{n+1}\mathrm{d}u, \tag{19}$$

where the Fresnel number has been introduced as

$$N_F = \frac{kr_0^2}{2\pi z}, \tag{20}$$

together with the normalized transverse coordinate

$$\boldsymbol{\rho} = \frac{\boldsymbol{r}}{r_0} = (\xi, \eta). \tag{21}$$

It is seen that the vortex structure of the phase is preserved during propagation and that, once the order n has been set, the propagated field depends only on the Fresnel number, as happens for the diffraction of a plane wave by a circular hole (corresponding to $n = 0$) [61]. Therefore, the Fresnel number will be taken as the natural variable to represent the propagation distance. As examples, density plots of the spectral density of some modes across the plane (x, z), that is, the squared modulus of the expression in equation (19), are shown in Figure 1.

As a first remark, we note that the property for a CSD of being uni-variable across a certain plane is not preserved during propagation. In fact, when modes propagate, their analytical forms actually change, and, in general, the product of the propagated modes, $\Phi_n(\boldsymbol{r}_1, z)\,\Phi_n^*(\boldsymbol{r}_2, z)$, is no longer proportional to $r_1^n r_2^n \exp[in(\varphi_1 - \varphi_2)]$, as it is required for the series to be interpreted as a power expansion.

Second, as also appears from Figure 1, the on-axis irradiance is zero for all modes, except $n = 0$. Furthermore, the modes spread during propagation at a rate that increases with n, giving less and less significant contributions in the central zone of the transverse planes. In particular, we expect that, when superpositions of modes are considered, only a finite number of modes need to be taken into account, which depends on the extent of the region where the field has to be evaluated. For example, modes with order 5 or higher can be safely neglected if the field has to be evaluated at $N_F = 0.5$ within a circular region of radius $2r_0$ (see Fig. 1).

3.2 Coherent modes in the far field

Unlike mode propagation in the Fresnel regime, far-field conditions lead to simple analytical forms. When the Fraunhofer conditions are met, the expression of a typical field

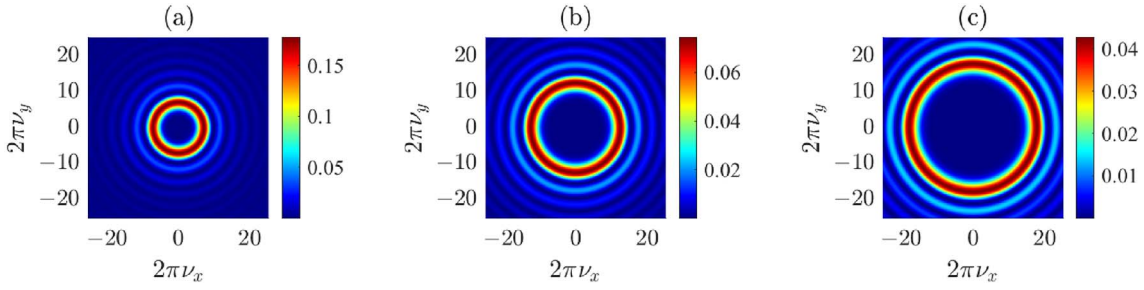


Fig. 2. Plots of the spectral density of the modes in the far field, equation (26) for $n = 5$ (a), 10 (b), 15 (c), with $r_0 = 1$.

(say, V_∞) is obtained from the Fourier transform, suitably scaled, of the field across the source [61]. For simplicity, the coordinate across the Fourier plane (say $\boldsymbol{v}$) will be taken as the spatial coordinate across a transverse plane in the far zone. Disregarding unessential proportionality factors and curvature terms, we simply write

$$V_\infty(\boldsymbol{v}) = \iint V(\boldsymbol{r}, 0) e^{-2\pi i \boldsymbol{v}\cdot\boldsymbol{r}} \mathrm{d}\boldsymbol{r}, \tag{22}$$

the integral extending over the whole $z = 0$ plane.

Using polar coordinates, the above integral reads

$$V_\infty(\boldsymbol{v}) = \int_0^\infty \int_0^{2\pi} V(r, \varphi, 0) e^{-2\pi i v r \cos(\vartheta - \varphi)} r \, \mathrm{d}r \, \mathrm{d}\varphi, \tag{23}$$

with v and ϑ being the polar coordinates of $\boldsymbol{v}$. For the case of the modes in equation (7), the above integral gives

$$\begin{aligned} \boldsymbol{\Phi}_{n,\infty}(\boldsymbol{v}) = {} & \frac{2(-i)^n}{r_0} \sqrt{\pi(n+1)} e^{in\vartheta} \\ & \times \int_0^{r_0} \left(\frac{r}{r_0}\right)^{n+1} J_n(2\pi v r) \mathrm{d}r, \end{aligned} \tag{24}$$

J_n being the Bessel function of the first kind and order n [62]. Exploiting the following relation obeyed by the Bessel functions [63]:

$$\tfrac{\mathrm{d}}{\mathrm{d}\eta}[\eta^{n+1} J_{n+1}(\eta)] = \eta^{n+1} J_n(\eta), \tag{25}$$

the expression of the modes propagated in the far zone takes the simple form

$$\boldsymbol{\Phi}_{n,\infty}(\boldsymbol{v}) = 2r_0(-i)^n \sqrt{\pi(n+1)} e^{in\vartheta} \frac{J_{n+1}(2\pi r_0 v)}{2\pi r_0 v}. \tag{26}$$

Note that for $n = 0$ the latter equation gives rise to the well-known Fraunhofer diffraction pattern from a circular hole [61]. Indeed, for $n = 0$ the vortex field coincides with that of an ordinary plane wave. For $n \neq 0$, we can read equation (26) as the diffraction pattern produced by a vortex field that impinges on a circular aperture [48].

Plots of the spectral density of the propagated modes of equation (26) are given in Figure 2 for $n = 5$, 10, 15. It appears that, in general, the function is very small until its argument reaches values roughly equal to n, and then slowly goes to zero in an oscillating way.

In light of the above result, it is useful to express the CSD of a beam radiated by a uni-variable source through its Mercer expansion, which reads

$$\begin{aligned} W_\infty(\boldsymbol{v}_1, \boldsymbol{v}_2) = {} & 4\pi r_0^2 \sum_{n=0}^\infty \lambda_n (n+1) e^{in(\vartheta_1 - \vartheta_2)} \\ & \times \frac{J_{n+1}(2\pi r_0 v_1)}{2\pi r_0 v_1} \frac{J_{n+1}(2\pi r_0 v_2)}{2\pi r_0 v_2}. \end{aligned} \tag{27}$$

In some cases, this expression even leads to closed forms for the degree of coherence in the far field (μ_∞). This happens, for example, if $\lambda_n \propto (n+1)$ and we limit ourselves to point along radial directions (i.e., $\vartheta_1 = \vartheta_2$), in which case it can be shown that [64]

$$\mu_\infty(v_1, \vartheta, v_2, \vartheta) = 2 \frac{J_1(2\pi r_0 (v_1 - v_2))}{2\pi r_0 (v_1 - v_2)}. \tag{28}$$

4 sZegö and truncated sZegö sources

The first example of a uni-variable CSD we are going to show is that of the so-called *sZegö source*. The features of such a source across the source plane and in the far zone were studied and commented on in detail in a previous paper [38]. Here, we limit ourselves to recalling some of the main results obtained there. The effects of paraxial propagation of the radiated field will be studied here on the basis of the expression of the propagated modes obtained in Section 3.1. However, we find it convenient to perform such an analysis starting from a different uni-variable CSD, namely, the *truncated* sZegö CSD. Unlike sZegö CSD, in fact, the latter requires a finite number of modes and reduces to sZegö CSD when the number of modes goes to infinity. In such a way, the propagation features of the beams they radiate can be evaluated exactly through finite sums. However, we want to stress that a CSD of this type is interesting in itself because it is suitable to be synthesized in the laboratory starting from the superposition of a finite number of coherent fields [20, 28, 32, 37].

4.1 The sZegö source

To obtain a sZegö CSD we let $c_n = I_0(\forall n)$ in equation (1), with I_0 a positive constant having dimensions of irradiance, so that

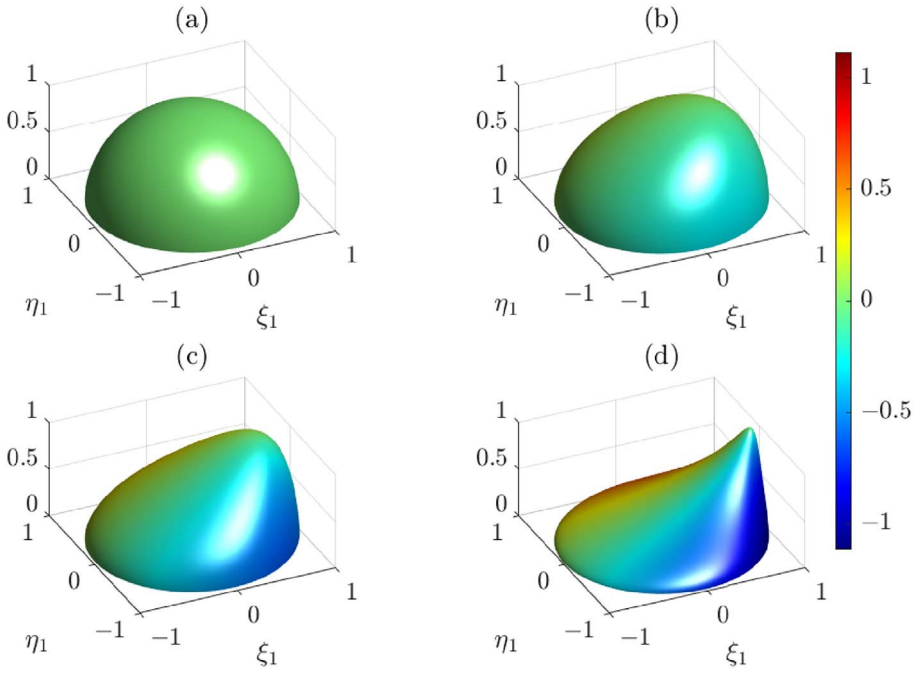


Fig. 3. Degree of coherence for a sZegö source at the source plane relative to a point located at (a) $\boldsymbol{\rho}_2 = (0, 0)$; (b) $\boldsymbol{\rho}_2 = (0.3, 0)$; (c) $\boldsymbol{\rho}_2 = (0.6, 0)$; and (d) $\boldsymbol{\rho}_2 = (0.9, 0)$. Absolute value is represented on the vertical axis, and the phase is coded in a color scale.

$$g(\zeta) = \frac{I_0}{1-\zeta}(|\zeta| < 1), \tag{29}$$

and the corresponding CSD turns out to be

$$W_S(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, 0) = \frac{I_0}{1-\rho_1\rho_2 e^{i(\varphi_1-\varphi_2)}}\,\mathrm{circ}(\rho_1)\ \mathrm{circ}(\rho_2). \tag{30}$$

Spectral density and degree of coherence across the source plane turn out to be, from equations (15) and (16),

$$S_S(\boldsymbol{\rho}, 0) = \frac{I_0}{1-\rho^2}\,\mathrm{circ}(\rho), \tag{31}$$

and

$$\mu_S(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, 0) = \frac{\sqrt{(1-\rho_1^2)(1-\rho_2^2)}}{1-\rho_1\rho_2 e^{i(\varphi_1-\varphi_2)}}\,\mathrm{circ}(\rho_1)\ \mathrm{circ}(\rho_2), \tag{32}$$

respectively. It can be seen that the spectral density grows to infinity as ρ approaches one. This somewhat anomalous behavior has to be ascribed to the fact that an infinite number of modes contribute to the CSD of the source. It will be solved in the next subsection, where the series that gives rise to sZegö source will be truncated. Plots of μ_S (absolute value and phase) are shown in Figure 3 as a function of $\boldsymbol{\rho_1}$ for fixed $\boldsymbol{\rho_2}$. The absolute value of μ_S is seen to vary from 1 (when $\boldsymbol{\rho_1} = \boldsymbol{\rho_2}$) to 0 (along the circle $\rho_1 = 1$). Furthermore, when $\boldsymbol{\rho_2}$ is in the center of the source, the phase is constant and equals to 0, while the range of variation of the phase increases with increasing ρ_2. The presence of a coherence vortex can also be observed, which has charge −1, centered at a point outside the source at a distance $1/\rho_2$ from the center [38].

The OAM degree of coherence at $z = 0$ can be derived using infinite geometric series. Together with the degree of orbitalization they take forms:

$$o_S(\rho_1, \rho_2, 0) = \sqrt{\frac{(1-\rho_1^2)(1-\rho_2^2)}{1-\rho_1^2\rho_2^2}};$$

$$O_S(\rho, 0) = \left(1-\rho^2\right)^2. \tag{33}$$

Note that $O_S(\rho, 0)$ has a particularly simple form since for this source $\lambda_n(\rho, 0) \geq \lambda_n + 1(\rho, 0)$ for all n and ρ, which is not generally true for other source classes. The color density plot of the OAM degree of coherence for this beam is shown in Figure 4a. As expected, this quantity takes the unity value at $\rho_1 = \rho_2 = 0$ but gradually decreases and reaches zero for $\rho_1 = 1$ or $\rho_2 = 1$. We also note that it is not generally unity at the coinciding radii $\rho_1 = \rho_2 \neq 0$, unlike the classic degree of coherence that would be at the coinciding points.

An explicit expression of the CSD of the field radiated in the far zone, $W_{S,\infty}$, can hardly be derived directly from the CSD in $z = 0$, but from its modal expansion, with the modes of equation (26) and the eigenvalues given by

$$\lambda_n = I_0\frac{\pi r_0^2}{n+1}, \quad (n = 0, 1, ...), \tag{34}$$

takes the form

$$W_{S,\infty}(\boldsymbol{v}_1, \boldsymbol{v}_2) \propto \sum_{n=0}^{\infty} \frac{J_{n+1}(2\pi r_0 v_1)}{2\pi r_0 v_1} \times \frac{J_{n+1}(2\pi r_0 v_2)}{2\pi r_0 v_2}\, e^{in(\vartheta_1-\vartheta_2)}, \tag{35}$$

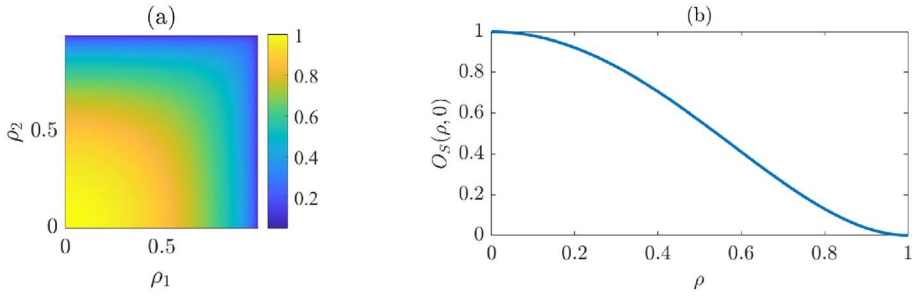


Fig. 4. (a) OAM degree of coherence $o_S(\rho_1, \rho_2)$ and (b) degree of orbitalization $O_S(\rho)$ for a sZegö source.

which can be used to evaluate $S_{S,\infty}$ and $\mu_{S,\infty}$ numerically. The latter quantities, evaluated by adding the contribution of the first 100 modes, are shown in Figures 5 and 6, respectively.

For the case of the spectral density, it was found that there are no significant differences, at least in the range $r_0 v < 1$, when calculating it using only the first 6 modes or the first 100 modes, so that the contribution of modes of order higher than 6 turns out to be negligible. Basically, the same conclusions hold for the degree of coherence, but in such a case, the first 10 modes have to be considered. The degree of coherence shows cylindrical symmetry when calculated in relation to the beam axis, with its absolute value reaching a maximum at the center of the beam and showing decreasing oscillations, while the phase alternates between 0 and π. When evaluated in relation to a direction outside the axis, the oscillations of the absolute value of the DOC are distorted, and the phase varies gradually. In this case also the presence of coherence vortices can be observed (with charge $+1$ and -1), corresponding to the zeros of $\mu_{S,\infty}$ [38]. The absolute value is represented on the vertical axis, and the phase is coded in a color scale. The first 100 modes were considered in the calculation.

The far-zone OAM degree of coherence of a field radiated by a sZegö source takes form

$$o_S(v_1, v_2) = \frac{2\sqrt{\sum_{n=0}^{\infty} J_{n+1}^2(2\pi r_0 v_1) J_{n+1}^2(2\pi r_0 v_2)}}{\sqrt{1 - J_0^2(2\pi r_0 v_1)}\sqrt{1 - J_0^2(2\pi r_0 v_2)}}, \qquad (36)$$

after using a summation formula for $J_{n+1}^2(2\pi r_0 v_j)$, $j = 1, 2$ in the denominator. Figure 7a shows this quantity as a function of v_1, v_2 with $r_0 = 1$, where we restrict ourselves to a region close to the axis. This distribution is structurally different from that in Figure 4a, since the coherence no longer decreases monotonically with the increasing radii but oscillates between wider and narrower distributions. It should also be noted that for $v_1 = v_2 = v$ the OAM coherence remains fairly high within the plotted region around the axis.

The degree of orbitalization becomes

$$O_S(v) = 2\frac{J_p^2(2\pi r_0 v) - J_q^2(2\pi r_0 v)}{1 - J_0^2(2\pi r_0 v)}, \qquad (37)$$

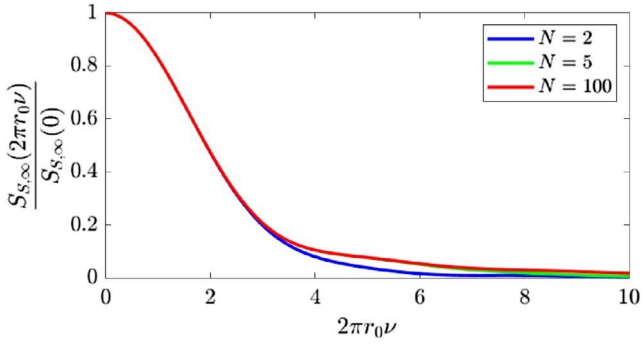


Fig. 5. Spectral density for a sZegö source in the far field (with $r_0 = 1$), calculated by adding the first 2, 5, and 100 modes, repectively.

where p and q are the indices of the largest and the second largest eigenvalues σ_n for a given radius v. We note that unlike in the source plane, where the zeroth and the first eigenvalues are the largest and second largest for all radii, in the far field such eigenvalues must be determined numerically. Figure 7b shows the degree of orbitalization as a function of v, with $r_0 = 1$, for $v \leq 1$. Unlike in the source plane, where this degree monotonically decreases from unity to zero, see Figure 4b, it shows a piecewise-continuous profile without smoothness at junctions. This is due to the fact that many OAM modes are competing for being the maximum and the second maximum. For example, close to the axis, modes $n = 0$ and $n = 1$ constitute this pair, but at $v \approx 0.419$, modes $n = 1$ and $n = 2$ become such. Hence, most of the oscillations are due to switching to higher indexes, with one exception: in the region $0.817 \leq v \leq 0.894$, the $n = 0$ mode has a crest becoming the second largest eigenvalue again.

4.2 Truncated sZegö sources

In this case, we let $c_n = I_0$ if $0 \leq n < N$ and $c_n = 0$ if $n \geq N$, so that equation (1) gives

$$g_T(\zeta) = I_0 \sum_{n=0}^{N-1} \zeta^n = I_0 \frac{1 - \zeta^N}{1 - \zeta} \quad (|\zeta| < 1), \qquad (38)$$

and the corresponding CDS turns out to be

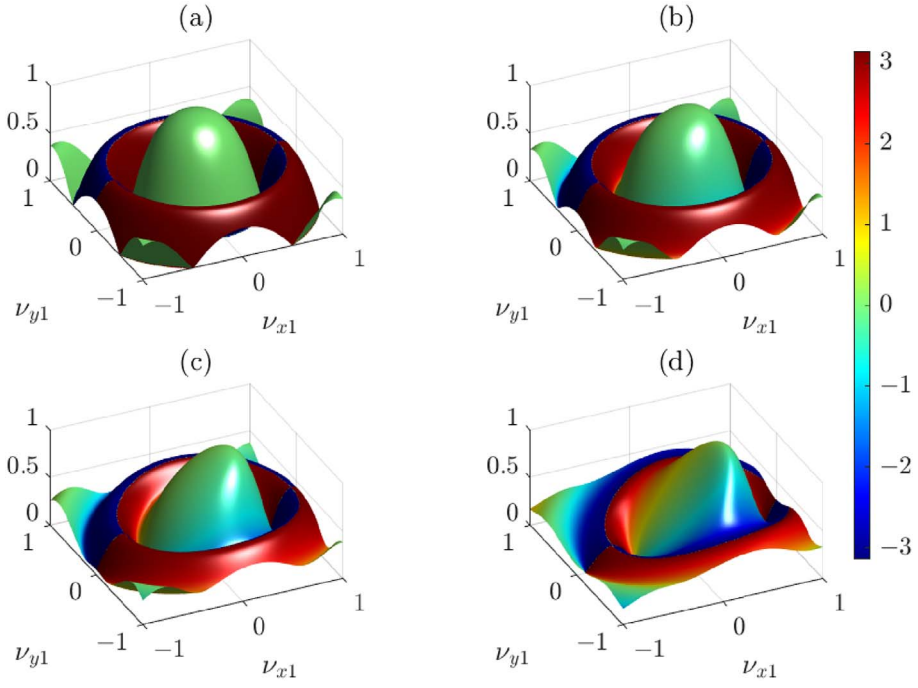


Fig. 6. Degree of coherence for a sZegö source across a plane in the far zone relative to a point located at (a) $\boldsymbol{v}_2 = (0, 0)$; (b) $\boldsymbol{v}_2 = (0.16, 0)$; (c) $\boldsymbol{v}_2 = (0.32, 0)$; and (d) $\boldsymbol{v}_2 = (0.48, 0)$. Absolute value is represented on the vertical axis, and the phase is coded in a color scale. The first 100 modes were considered in the calculation.

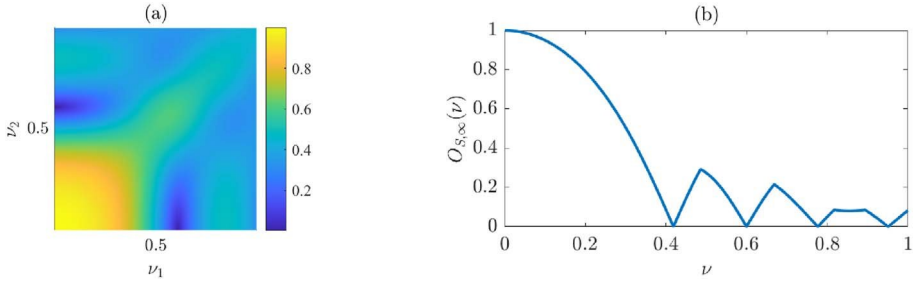


Fig. 7. (a) OAM degree of coherence $o_S(v_1, v_2)$; (b) degree of orbitalization $O_S(v)$ for a far field radiated by sZegö source.

$$W_T(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, 0) = I_0 \frac{1 - (\rho_1\rho_2)^N e^{iN(\varphi_1-\varphi_2)}}{1 - \rho_1\rho_2 e^{i(\varphi_1-\varphi_2)}}; \quad (0 \le \rho_1, \rho_2 < 1). \tag{39}$$

Spectral density and degree of coherence across the source plane turn out to be, from equations (15) and (16),

$$S_T(\boldsymbol{\rho}, 0) = I_0 \frac{1 - \rho^{2N}}{1 - \rho^2} \mathrm{circ}(\rho), \tag{40}$$

and

$$\mu_T(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, 0) = \sqrt{\frac{(1-\rho_1^2)(1-\rho_2^2)}{(1-\rho_1^{2N})(1-\rho_2^{2N})}} \times \frac{1 - (\rho_1\rho_2)^N e^{iN(\varphi_1-\varphi_2)}}{1 - \rho_1\rho_2 e^{i(\varphi_1-\varphi_2)}} \mathrm{circ}(\rho_1)\ \mathrm{circ}(\rho_2), \tag{41}$$

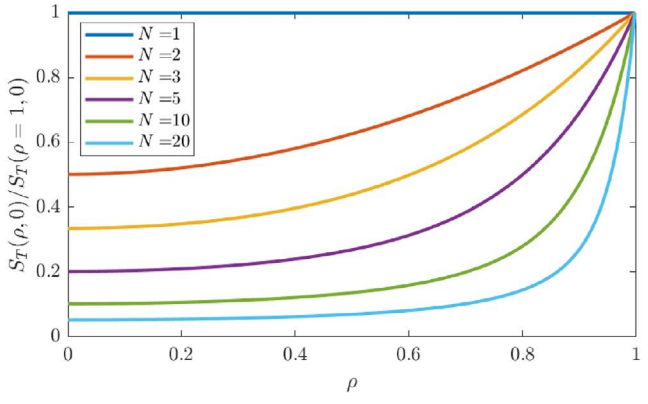


Fig. 8. Spectral density at the source (normalized to the maximum) for a truncated sZegö CSD with $N = 1, 2, 3, 5, 10, 20$.

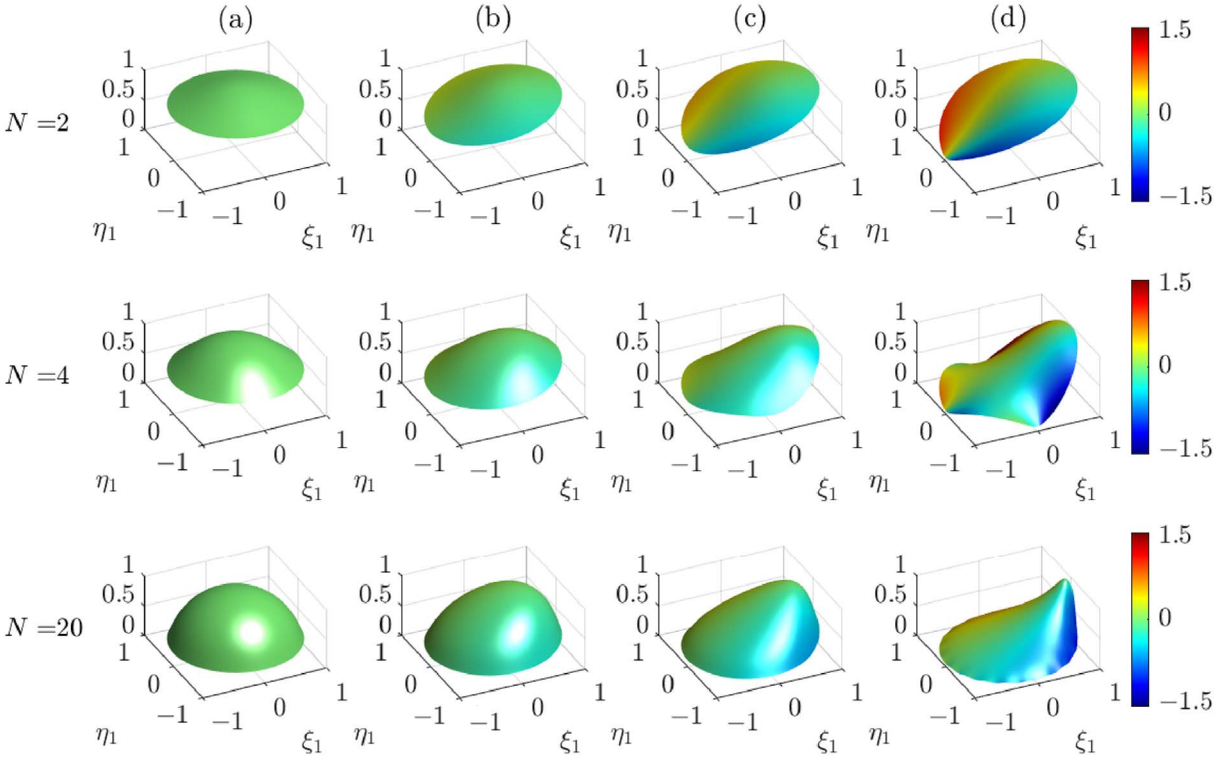


Fig. 9. Degree of coherence at the source plane, for a truncated sZegö CSD with $N = 2, 4, 20$ (from top to bottom), relative to a point located at (a) $\boldsymbol{\rho}_2 = (0, 0)$; (b) $\boldsymbol{\rho}_2 = (0.3, 0)$; (c) $\boldsymbol{\rho}_2 = (0.6, 0)$; and (d) $\boldsymbol{\rho}_2 = (0.9, 0)$. Absolute value is represented on the vertical axis, and the phase is coded in a color scale.

respectively. The plots of S_T and μ_T are shown in Figures 8 and 9, respectively. It can be seen that, as the number of modes increases, the spectral density becomes more and more similar to that of the sZegö source but never diverges, as expected. On the other hand, the DOC is highly dependent on the number of modes considered.

Two aspects deserve to be noticed. First, the coherence area decreases as N increases, as expected. The second aspect concerns the phase structure of μ_T. It can be highlighted in a simpler way if we focus our attention on the second fraction in equation (41), provided the other terms are positive. Using again the variable $\zeta = \rho_1\rho_2 \exp[i(\varphi_1 - \varphi_2)]$, this term can be written as

$$\frac{1-\zeta^N}{1-\zeta} = -\frac{\prod_{n=1}^{N}(\zeta-\zeta_n)}{1-\zeta}, \tag{42}$$

where ζ_n $(n = 1, \ldots, N)$ are the solutions of the equation $\zeta^N = 1$, namely, $\zeta_n = \exp(i2\pi n/N)$. Therefore, since $\zeta_N = 1$, we have

$$-\frac{\prod_{n=1}^{N}(\zeta-\zeta_n)}{1-\zeta} = \prod_{n=1}^{N-1}(\zeta-\zeta_n), \tag{43}$$

which for fixed values of ρ_2 and φ_2 can be written as

$$\prod_{n=1}^{N-1}(\zeta-\zeta_n) = \rho_2^{(N-1)} e^{-i(N-1)\varphi_2} \times \prod_{n=1}^{N-1}\left[\rho_1 e^{i\varphi_1} - \frac{1}{\rho_2} e^{i(\varphi_2+2\pi n/N)}\right], \tag{44}$$

representing the product of $N-1$ vortices, each having charge $+1$, centered on $N-1$ vertices of a regular enneagon, at a distance $1/\rho_2$ from the center.

To better understand this point, let us take $\varphi_2 = 0$. Then, the nth term of the product takes the form

$$\rho_1 e^{i\varphi_1} - \frac{1}{\rho_2} e^{i2\pi n/N} = \left[\xi_1 - \frac{1}{\rho_2}\cos\left(\frac{2\pi n}{N}\right)\right] + i\left[\eta_1 - \frac{1}{\rho_2}\sin\left(\frac{2\pi n}{N}\right)\right], \tag{45}$$

which actually represents a vortex with charge $+1$ and center at

$$\boldsymbol{\rho}_1^{(n)} = \left(\frac{1}{\rho_2}\cos\left(\frac{2\pi n}{N}\right), \frac{1}{\rho_2}\sin\left(\frac{2\pi n}{N}\right)\right). \tag{46}$$

The phase structure of the DOC is shown in Figure 10 for $\boldsymbol{\rho}_2 = (0.9, 0)$ and several values of N.

The OAM degree of coherence for this source at $z = 0$ can be derived using finite geometric series. This quantity and the degree of orbitalization take form

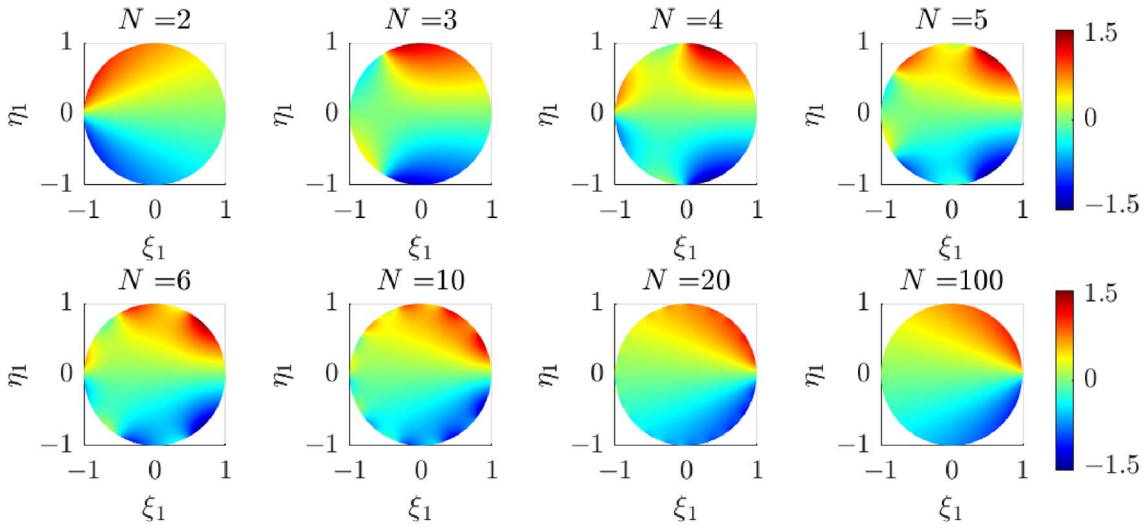


Fig. 10. Phase of the degree of coherence at the source plane, for a truncated sZegö CSD relative to the point $\boldsymbol{\rho}_2 = (0.9, 0)$ for several values of N.

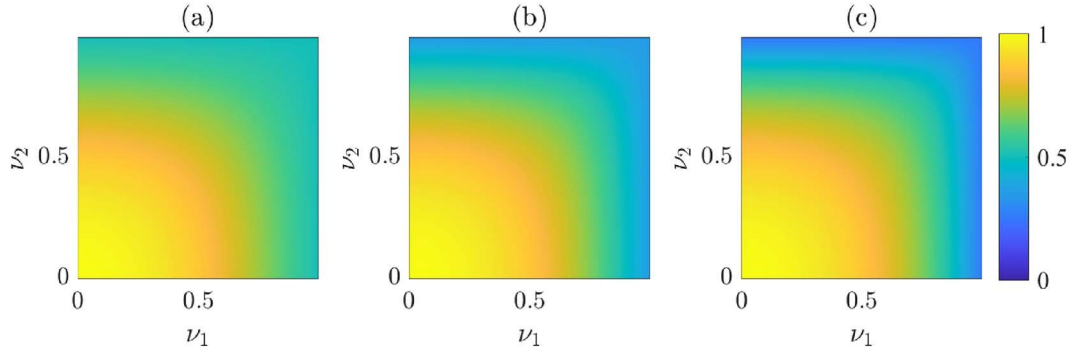


Fig. 11. OAM degree of coherence $o_T(\rho_1, \rho_2)$ for truncated sZegö sources with (a) $N = 4$; (b) $N = 8$; (c) $N = 16$.

$$o_T(\rho_1, \rho_2, 0) = \sqrt{\frac{(1-\rho_1^2)(1-\rho_2^2)}{1-\rho_1^2\rho_2^2}} \times \sqrt{\frac{1-\rho_1^{2N}\rho_2^{2N}}{(1-\rho_1^{2N})(1-\rho_2^{2N})}};$$
$$O_T(\rho, 0) = \frac{(1-\rho^2)^2}{1-\rho^{2N}}. \tag{47}$$

The plots of the OAM degree of coherence for this source are shown in Figures 11a–11c, for several values of N. Unlike for the sZegö sources the distributions do not decrease to zero for $\rho_1 = 1$ or $\rho_2 = 1$ but rather take on values in the interval (0, 1) decreasing with increasing N. The plots of the degree of orbitalization are shown in Figure 12 for several values of N. The case $N = \infty$ is also shown corresponding to the un-truncated sZegö source. While all the curves have monotonic behavior taking on unit value at $\rho = 0$ and decreasing to zero at $\rho = 1$ only the curve for $N = 1$ is convex on the whole interval, all the others change convexity once.

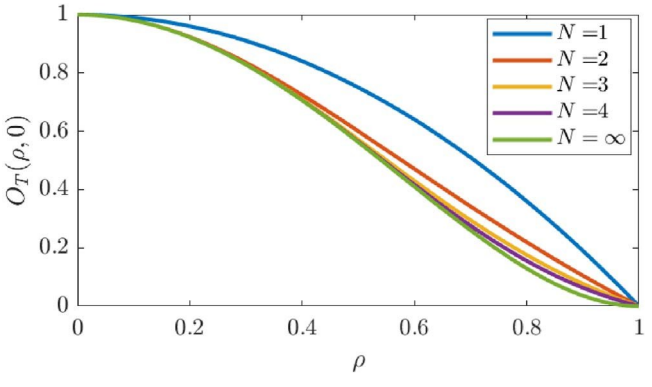


Fig. 12. Degree of orbitalization $O_T(\rho)$ for truncated sZegö sources with different values of N. The case $N = \infty$ corresponds to the non-truncated sZegö source.

To evaluate the propagated CSD in the near and in the far field equations (14) and (19), or (26) are used. Equations (38) and (8) give, for the eigenvalues,

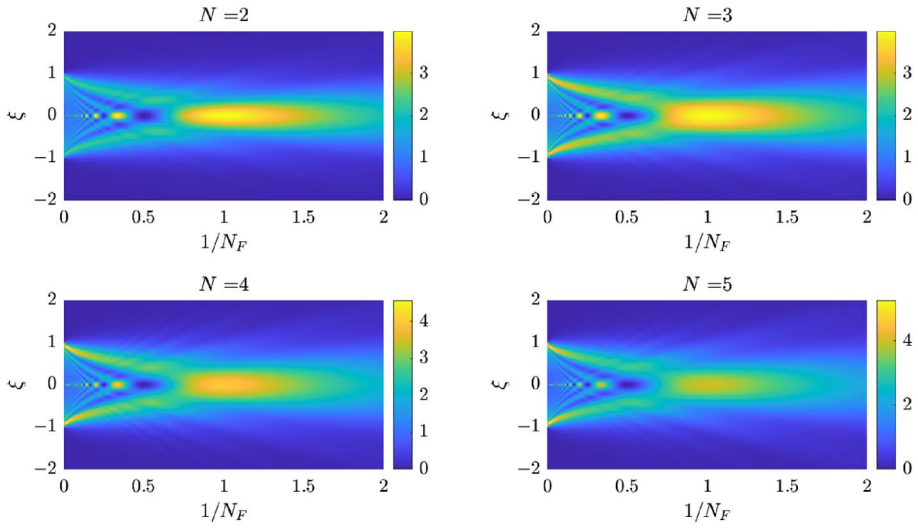


Fig. 13. Spectral density versus the inverse of the Fresnel number for several truncated sZegö sources with N modes.

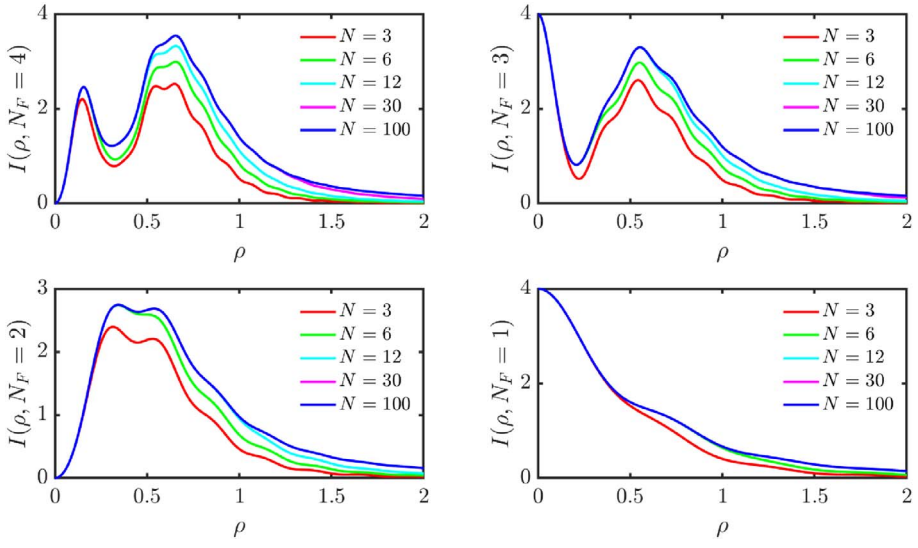


Fig. 14. Spectral density versus the inverse of the Fresnel number for different truncated sZegö sources.

$$\lambda_n = \begin{cases} I_0 \frac{\pi r_0^2}{n+1} & (0 \le n < N); \\ \\ 0 & (n \ge N). \end{cases} \tag{48}$$

Figure 13 shows examples of the spectral density, across the (x, z) plane, of the field propagated from sZegö sources of various orders. It can be observed that the main contribution for large propagation distances is due to the *0*th order mode (compare Fig. 13 with Fig. 1 for $n = 0$).

As commented above, the addition of more and more modes modifies the distribution of the spectral density only for small propagation distances. This is confirmed by the plots in Figure 14 where the spectral density profiles are calculated at several propagation distances for sZegö truncated sources with different numbers of modes. For a Fresnel number $N_F = 4$, the five profiles are different, although those corresponding to $N = 30$ and $N = 100$ are practically the same up to a normalized distance $\rho = 1$. For $N_F = 1$ the profiles practically coincide for all truncated sZegö sources with $N \ge 12$. This profile should be the same as the one for sZegö source.

The same approach can be used to evaluate the degree of coherence across different planes (Fig. 15). The analogous calculations across the source plane and in the far field [38] showed the presence of coherence vortices (in both cases).

The far-zone OAM degree of coherence radiated by a truncated sZegö source takes the form

$$o_T(v_1, v_2) = \frac{\sqrt{\sum_{n=0}^{N-1} J_{n+1}^2(2\pi r_0 v_1) J_{n+1}^2(2\pi r_0 v_2)}}{\prod_{j=1,2} \sqrt{\sum_{n=0}^{N-1} J_{n+1}^2(2\pi r_0 v_j)}}. \tag{49}$$

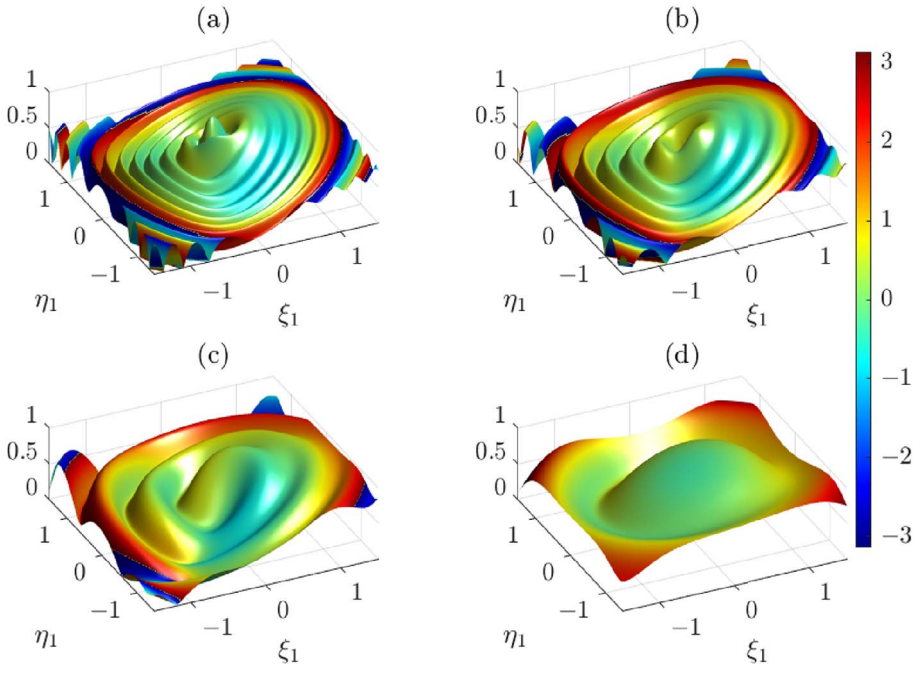


Fig. 15. Degree of coherence relative to a point located at $\boldsymbol{\rho}_2 = (0, 0.5)$ for a truncated sZegö source with $N = 4$ at several propagation distances (a) $N_F = 3$; (b) $N_F = 2$; (c) $N_F = 1$; (d) $N_F = 0.5$. The absolute value is represented on the vertical axis, and the phase is coded in a color scale.

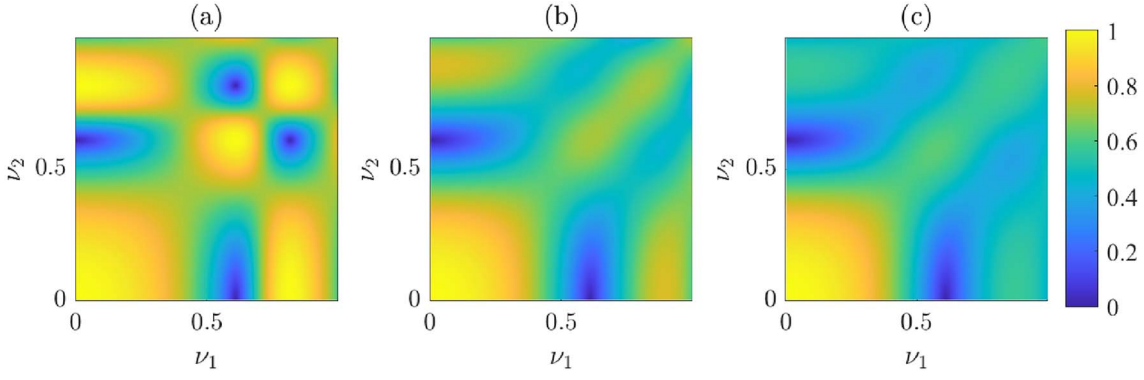


Fig. 16. OAM degree of coherence $o_T(\nu_1, \nu_2)$ for truncated sZegö far fields with (a) $N = 2$; (b) $N = 3$; (c) $N = 4$.

Figure 16 shows this quantity for several values of the summation index: (a) $N = 2$, (b) $N = 3$, and (c) $N = 4$. As N increases, the distributions start resembling that for the non-truncated model, if compared to Figure 7a. This practically occurs already for $N = 4$, as higher-order modes deliver much smaller contributions to the sums.

For truncated sZegö far fields the degree of orbitalization becomes

$$O_T(v) = \frac{J_p^2(2\pi r_0 v) - J_q^2(2\pi r_0 v)}{\sum_{n=0}^{N-1} J_{n+1}^2(2\pi r_0 v)}, \tag{50}$$

where, as above, p and q are the indices of the largest and the second largest eigenvalues. Analogously to the non-truncated model, the determination of the largest and the second largest eigenvalues for different radii was carried out numerically.

Figure 17 shows formation of the degree of orbitalization for (a) $N = 2$ and (b) $N = 3$. For $N = 2$ the degree exhibits strictly oscillatory nature with smooth maxima and non-smooth minima, all at zeroes. This is due to the fact that only two modes compete with each other for being maximum and second maximum and can only alternate. This behavior is similar to that of the degree of polarization, also involving competition of two electric field components. Starting from $N = 3$ both maxima and minima are not smooth, since the switches are determined by three pairs of modes. Figure 17c summarizes the behavior for cases $N = 2$, $N = 3$ and $N \to \infty$ (non-truncated model), same as in Figure 7b. In parts where the curves are almost the same, the same modes act as the largest and second largest

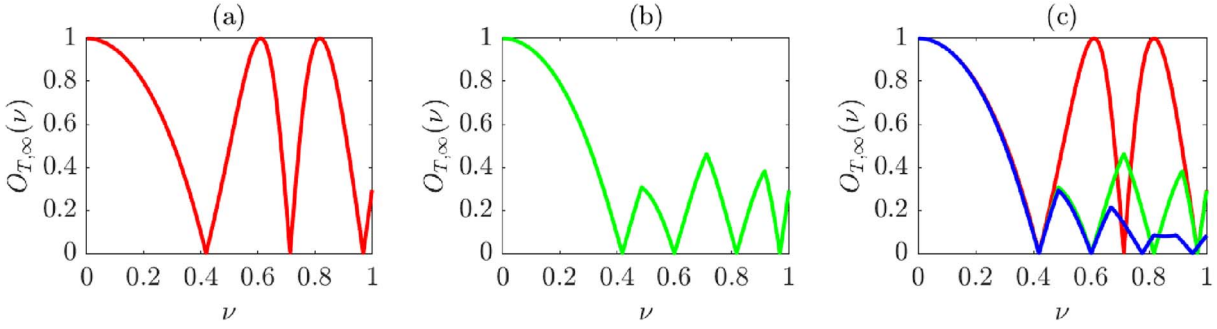


Fig. 17. Degree of orbitalization $O_T(v)$ for truncated sZegö far fields with (a) $N = 2$; (b) $N = 3$; (c) combination of $N = 2$, $N = 3$, and non-truncated model.

eigenvalues resulting in the same numerator in equation (17), while the denominator has an increasing value with increasing N. This always occurs at small v as the 0th mode always dominates close to the axis.

5 Further examples of uni-variable sources

Different uni-variable CSDs can be defined by choosing different sets of c_n coefficients, and in many cases they are given in closed form, as happened for the case in the previous section. Actually, using handbooks of mathematical formulas, dozens of other CSDs can be defined from formulas involving power series. Some of them lead to closed forms even when the series is truncated. This is particularly useful because, as we saw for sZegö and truncated sZegö sources, in such a case the effects of truncation can be evaluated quite simply. Furthermore, in any practical realization of partially coherent sources based on the superimposition of perfectly coherent and mutually uncorrelated fields, the number of modes must be necessarily truncated.

Among others, we quote the following (formula 5.2.2.4 of [64]):

$$g(\zeta) = I_0 \sum_{n=1}^{\infty} n\zeta^n = I_0 \frac{\zeta}{(1-\zeta)^2}, \tag{51}$$

together with its truncated version (formula 4.1.7.2 of [64])

$$g_T(\zeta) = I_0 \sum_{n=1}^{N} n\zeta^n = I_0 \frac{\zeta + (N\zeta - N - 1)\,\zeta^{N+1}}{(1-\zeta)^2}, \tag{52}$$

with integer N.

Another interesting example of a source obtained with a finite number of modes is that of the *binomial source*, for which (formula 4.2.3.1 of [64])

$$g(\zeta) = I_0 \sum_{n=0}^{N} \binom{N}{n} \zeta^n = I_0 (1+\zeta)^N. \tag{53}$$

As a last example, we quote the following, which will be discussed in a little more detail. If we take (formula 5.2.7.2 of [64])

$$g(\zeta) = I_0 \sum_{n=0}^{\infty} \frac{(\gamma^2\zeta)^n}{n!} = I_0 e^{\gamma^2\zeta}, \tag{54}$$

with γ^2 a positive dimensionless parameter, the corresponding uni-variable CSD turns out to be

$$W_G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = I_0 \exp\left\{\gamma^2 \rho_1 \rho_2 \exp\left[i(\varphi_1 - \varphi_2)\right]\right\} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2), \tag{55}$$

and the corresponding spectral density has the form

$$S_G(\rho) = I_0 e^{\gamma^2\rho^2} \operatorname{circ}(\rho). \tag{56}$$

The interest in such sources comes from what follows. If we express W_G through the vectors $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$, since $\boldsymbol{\rho}_j = (\xi_j, \eta_j)$ and $\rho_j \exp(i\varphi_j) = \xi_j + i\eta_j$, $(j = 1, 2)$, equation (55) can also be written as

$$\begin{aligned} W_G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) &= I_0 e^{\gamma^2(\xi_1\xi_2 + \eta_1\eta_2)} e^{-i\gamma^2(\xi_1\eta_2 - \xi_2\eta_1)} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2) \\ &= I_0 e^{\gamma^2 \boldsymbol{\rho}_1\cdot\boldsymbol{\rho}_2}\, e^{-i\gamma^2(\boldsymbol{\rho}_1\times\boldsymbol{\rho}_2)_z} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2) \end{aligned} \tag{57}$$

or, after some manipulations,

$$W_G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = I_0 e^{\gamma^2(\rho_1^2 + \rho_2^2)/2} e^{-\gamma^2|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|^2/2} e^{-i\gamma^2(\boldsymbol{\rho}_1\times\boldsymbol{\rho}_2)_z} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2), \tag{58}$$

which apparently has the form of a twisted Gaussian Schell-model source with saturated twist [40], the only difference being the sign in the argument of the Gaussian amplitude term. In particular, the degree of coherence turns out to be

$$\mu_G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = e^{-\gamma^2|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|^2/2} e^{-i\gamma^2(\boldsymbol{\rho}_1\times\boldsymbol{\rho}_2)_z} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2), \tag{59}$$

whose modulus is

$$|\mu_G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)| = e^{-\gamma^2|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|^2/2} \operatorname{circ}(\rho_1)\operatorname{circ}(\rho_2). \tag{60}$$

The eigenvalues of the Mercer expansion of W_G turn out to be

$$\lambda_n = I_0 \gamma^{2n} \frac{\pi r_0^2}{(n+1)!}, \tag{61}$$

from which the propagation features of the radiated beams could be evaluated.

In this case, too, the truncated version of the series can be expressed in closed form because (formula 4.1.7.10 of [64])

$$g_T(\zeta) = I_0 \sum_{n=0}^{N} \frac{(\gamma^2\zeta)^n}{n!} = I_0 \frac{\Gamma(N+1, \gamma^2\zeta)}{N!} e^{\gamma^2\zeta}, \qquad (62)$$

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function [63].

6 Conclusions

Uni-variable CSDs can be derived from any function of a single complex argument whose Taylor series involves only non-negative coefficients. The convergence range of the series determines the spatial extent of the source. The Taylor series can be directly related to the Mercer expansion of the CSD of the source in such a way that the coherent modes turn out to be optical vortices, restricted to a circle at the source plane, and the corresponding eigenvalues are proportional to the Taylor coefficients.

Although the coherence features of light across the source plane can be directly derived from the analytical form of the CSD, the analogous properties for the radiated beam can hardly be evaluated in closed form. Nevertheless, from the knowledge of modes and eigenvalues of the CSD across the source, the coherence features of the radiated beam can be evaluated both in the Fresnel regime and in the Fraunhofer regime.

In this paper, the tools for studying the properties of the propagated field have been provided and applied to a test case, that is, the sZegö source. Furthermore, the effects, on the coherence properties of the source and the radiated field, of truncating the Taylor series have also been investigated. This is crucial from an experimental point of view, because limiting the Taylor series (i.e., the coherent-mode expansion of the CSD) to a finite number of terms becomes necessary whenever sources of this kind have to be synthesized through the superposition of mutually uncorrelated perfectly coherent fields.

Since the coherent modes of uni-variable CSDs possess vortex-like structures at various OAM indices, the overall fields can be regarded as carrying OAM in multiple states. Hence, besides the spectral density and the degree of coherence, which carry information about correlations in the physical space, some recently introduced measures, namely, the OAM degree of coherence and the degree of orbitalization, have also been invoked to characterize the radial correlations of the investigated sources in the OAM space, i.e., the polar Fourier space. Although the technique presented here has been applied to the sZegö CSD and to its truncated version (the truncated sZegö CSD), it can be applied to any source of the uni-variable type. The number of uni-variable CSD that can be envisaged starting from the Taylor expansion of a single-argument function is actually huge, and this work lays the foundation for the design and tayloring of sources with required coherence characteristics, both on the source plane and in propagation.

Funding

This work has been supported by the Spanish Ministerio de Economía y Competitividad under project PID2023-148021NB-I00 and Project FASLIGTH (RED2022-134391-T).

Conflicts of interest

This work has no financial or non-financial competing interests.

Data availability statement

Data will be made available on request.

Author contribution statement

All authors have contributed equally to all aspects of this work, including conceptualization, methodology, formal analysis, and the writing of the original draft and final manuscript.

References

1 Yu J, Zhu X, Wang F, Chen Y, Cai Y, Research progress on manipulating spatial coherence structure of light beam and its applications, *Prog. Quantum Electron.* **91–92**, 100486 (2023). https://doi.org/10.1016/j.pquantelec.2023.100486.
2 Rosales-Guzmán C, Rodríguez-Fajardo V, A perspective on structured light's applications, *Appl. Phys. Lett.* **125**, 200503 (2024). https://doi.org/10.1063/5.0236477.
3 Liang C, Wu G, Wang F, Li W, Cai Y, Ponomarenko SA, Overcoming the classical Rayleigh diffraction limit by controlling two-point correlations of partially coherent light sources, *Opt. Express* **25**, 28352 (2017). https://doi.org/10.1364/OE.25.028352.
4 Shen Y, Sun H, Peng D, Chen Y, Cai Q, Wu D, Wang F, Cai Y, Ponomarenko SA, Optical image reconstruction in 4f imaging system: Role of spatial coherence structure engineering, *Appl. Phys. Lett.* **118**, 181102 (2021). https://doi.org/10.1063/5.0046288.
5 Gibson G, Courtial J, Padgett MJ, Vasnetsov M, Pas'ko V, Barnett SM, Franke-Arnold S, Free-space information transfer using light beams carrying orbital angular momentum, *Opt. Express* **12**, 5448 (2004). https://doi.org/10.1364/OPEX.12.005448.
6 Wang J, et al., Terabit free-space data transmission employing orbital angular momentum multiplexing, *Nat. Photonics* **6**, 488 (2012). https://doi.org/10.1038/nphoton.2012.138.
7 Liu X, Shen Y, Liu L, Wang F, Cai Y, Experimental demonstration of vortex phase-induced reduction in scintillation of a partially coherent beam, *Opt. Lett.* **38**, 5323 (2013). https://doi.org/10.1364/OL.38.005323.
8 Gbur G, Partially coherent beam propagation in atmospheric turbulence, *J. Opt. Soc. Am. A* **31**, 2038 (2014). https://doi.org/10.1364/JOSAA.31.002038.
9 Lin S, Zhu X, Shen Y, Wang F, Chen X, Gbur G, Cai Y, Yu J, Statistically stationary random light for high-security encryption, *Optica* **12**, 1261 (2025). https://doi.org/10.1364/OPTICA.546899.
10 Zhao C, Cai Y, Trapping two types of particles using a focused partially coherent elegant Laguerre–Gaussian beam, *Opt. Lett.* **36**, 2251 (2011). https://doi.org/10.1364/OL.36.002251.

11 Zhang Z, Liu X, Wang H, Liang C, Cai Y, Zeng J, Flexible optical trapping and manipulating Rayleigh particles via the cross-phase modulated partially coherent vortex beams, *Opt. Express* **32**, 35051 (2024). https://doi.org/10.1364/OE.539069.

12 Guo J, Ming S, Wu Y, Chen LQ, Zhang W, Super-sensitive rotation measurement with an orbital angular momentum atom-light hybrid interferometer, *Opt. Express* **29**, 208 (2021). https://doi.org/10.1364/OE.409964.

13 Gori F, Santarsiero M, Devising genuine spatial correlation functions, *Opt. Lett.* **32**, 3531 (2007). https://doi.org/10.1364/OL.32.003531.

14 Martínez-Herrero R, Mejías PM, Gori F, Genuine cross-spectral densities and pseudo-modal expansions, *Opt. Lett.* **34**, 1399 (2009). https://doi.org/10.1364/OL.34.001399.

15 Raghunathan SB, van Dijk T, Peterman EJG, Visser TD, Experimental demonstration of an intensity minimum at the focus of a laser beam created by spatial coherence: application to the optical trapping of dielectric particles, *Opt. Lett.* **35**, 4166 (2010). https://doi.org/10.1364/OL.35.004166.

16 Gbur G, Visser TD, The structure of partially coherent fields, *Prog. Opt.* **55**, 285 (2010). https://doi.org/10.1016/B978-0-444-53705-8.00005-9.

17 Wu G, Cai Y, Detection of a semirough target in turbulent atmosphere by a partially coherent beam, *Opt. Lett.* **36**, 1939 (2011). https://doi.org/10.1364/OL.36.001939.

18 Lajunen H, Saastamoinen T, Propagation characteristics of partially coherent beams with spatially varying correlations, *Opt. Lett.* **36**, 4104 (2011). https://doi.org/10.1364/OL.36.004104.

19 Cai Y, Chen Y, Wang F, Generation and propagation of partially coherent beams with nonconventional correlation functions: a review, *J. Opt. Soc. Am. A* **31**, 2083 (2014). https://doi.org/10.1364/JOSAA.31.002083.

20 Rodenburg B, Mirhosseini M, Magaña-Loaiza OS, Boyd RW, Experimental generation of an optical field with arbitrary spatial coherence properties, *J. Opt. Soc. Am. B* **31**, A51 (2014). https://doi.org/10.1364/JOSAB.31.000A51.

21 Divitt S, Novotny L, Spatial coherence of sunlight and its implications for light management in photovoltaics, Optica **2**, 95 (2015). https://doi.org/10.1364/OPTICA.2.000095.

22 Chen Y, Ponomarenko SA, Cai Y, Experimental generation of optical coherence lattices, *Appl. Phys. Lett.* **109**, 061107 (2016). https://doi.org/10.1063/1.4960966.

23 M.W. Hyde IV MW, Bose-Pillai SR, Wood RA, Synthesis of non-uniformly correlated partially coherent sources using a deformable mirror, *Appl. Phys. Lett.* **111**, 101106 (2017). https://doi.org/10.1063/1.4994669.

24 Santarsiero M, Martínez-Herrero R, Maluenda D, de Sande JCG, Piquero G, Gori F, Partially coherent sources with circular coherence, *Opt. Lett.* **42**, 1512 (2017). https://doi.org/10.1364/OL.42.001512.

25 Santarsiero M, Martínez-Herrero R, Maluenda D, de Sande JCG, Piquero G, Gori F, Synthesis of circularly coherent sources, *Opt. Lett.* **42**, 4115 (2017). https://doi.org/10.1364/OL.42.004115.

26 Ding C, Koivurova M, Turunen J, Pan L, Self-focusing of a partially coherent beam with circular coherence, *J. Opt. Soc. Am. A* **34**, 1441 (2017). https://doi.org/10.1364/JOSAA.34.001441.

27 Piquero G, Santarsiero M, Martínez-Herrero R, de Sande JCG, Alonzo M, Gori F, Partially coherent sources with radial coherence, *Opt. Lett.* **43**, 2376 (2018). https://doi.org/10.1364/OL.43.002376.

28 Chen X, Li J, Rafsanjani SMH, Korotkova O, Synthesis of I_m-Bessel correlated beams via coherent modes, *Opt. Lett.* **43**, 3590 (2018). https://doi.org/10.1364/OL.43.003590.

29 Wu D, Wang F, Cai Y, High-order nonuniformly correlated beams, *Opt. Laser Technol.* **99**, 230 (2018). https://doi.org/10.1016/j.optlastec.2017.09.007.

30 Martínez-Herrero R, Piquero G, de Sande JCG, Santarsiero M, Gori F, Besinc pseudo-Schell model sources with circular coherence, *Appl. Sci. Basel* **9**, 2716 (2019). https://doi.org/10.3390/app9132716.

31 de Sande JCG, Martínez-Herrero R, Piquero G, Santarsiero M, Gori F, Pseudo-Schell model sources, *Opt. Express* **27**, 3963 (2019). https://doi.org/10.1364/OE.27.003963.

32 Zhu X, Yu J, Chen Y, Wang F, Korotkova O, Cai Y, Experimental synthesis of random light sources with circular coherence by digital micro-mirror device, *Appl. Phys. Lett.* **117**, 121102 (2020). https://doi.org/10.1063/5.0024283.

33 Martínez-Herrero R, Santarsiero M, Piquero G, de Sande JCG, A new type of shape-invariant beams with structured coherence: Laguerre-Christoffel-Darboux beams, *Photonics* **8**, 134 (2021). https://doi.org/10.3390/photonics8040134.

34 Santarsiero M, Martínez-Herrero R, Piquero G, de Sande JCG, Gori F, *Modal analysis of pseudo-Schell model sources*, *Photonics* **8**, 449 (2021). https://doi.org/10.3390/photonics8100449.

35 Martínez-Herrero R, Piquero G, Santarsiero M, Gori F, de Sande JCG, A class of vectorial pseudo-Schell model sources with structured coherence and polarization, *Opt. Laser Technol.* **152**, 108079 (2022). https://doi.org/10.1016/j.optlastec.2022.108079.

36 Korotkova O, Gbur G, in A Tribute to Emil Wolf. Progress in Optics, Vol. 65, edited by T.D. Visser (Elsevier, 2020), pp. 43–104.

37 Moreno-Acosta P, Rickenstorff-Parrao C, Ramirez-San-Juan JC, Rosales-Guzman C, Ramos-Garcia R, Experimental generation of partially coherent besinc pseudo-Schell model sources using a digital micromirror device, *J. Opt. Soc. Am. B* **42**, 1804 (2025). https://doi.org/10.1364/JOSAB.564315.

38 Gori F, Santarsiero M, Martínez-Herrero R, Uni-variable cross-spectral densities, *Opt. Laser Technol.* **180**, 111511 (2025). https://doi.org/10.1016/j.optlastec.2024.111511.

39 Indebetouw G, Optical vortices and their propagation, *J. Mod. Opt.* **40**, 73 (1993). https://doi.org/10.1080/09500349314550101.

40 Simon R, Mukunda N, Twisted Gaussian Schell-model beams, *J. Opt. Soc. Am. A* **10**, 95 (1993). https://doi.org/10.1364/JOSAA.10.000095.

41 Gori F, Santarsiero M, Borghi R, Vicalvi S, Partially coherent sources with helicoidal modes, *J. Mod. Opt.* **45**, 539 (1998). https://doi.org/10.1080/09500349808231913.

42 Ponomarenko SA, A class of partially coherent beams carrying optical vortices, *J. Opt. Soc. Am. A* **18**, 150 (2001). https://doi.org/10.1364/JOSAA.18.000150.

43 Gbur G, Visser TD, Coherence vortices in partially coherent beams, *Opt. Commun.* **222**, 117 (2003). https://doi.org/10.1016/S0030-4018(03)01606-7.

44 Tamburini F, Anzolin G, Umbriaco G, Bianchini A, Barbieri C, Overcoming the Rayleigh criterion limit with optical vortices, *Phys. Rev. Lett.* **97**, 163903 (2006). https://doi.org/10.1103/PhysRevLett.97.163903.

45 Franke-Arnold S, Allen L, Padgett M, Advances in optical angular momentum, *Laser Photonics Rev.* **2**, 299 (2008). https://doi.org/10.1002/lpor.200810007.
46 Padgett M, Bowman R, Tweezers with a twist, *Nat. Photonics* **5**, 343 (2011). https://doi.org/10.1038/nphoton.2011.81.
47 Mirhosseini M, Magaña-Loaiza OS, Chen C, Rodenburg B, Malik M, Boyd RW, Rapid generation of light beams carrying orbital angular momentum, *Opt. Express* **21**, 30196 (2013). https://doi.org/10.1364/OE.21.030196.
48 Ambuj A, Vyas R, Singh S, Diffraction of orbital angular momentum carrying optical beams by a circular aperture, *Opt. Lett.* **39**, 5475 (2014). https://doi.org/10.1364/OL.39.005475.
49 Gbur G, Singular Optics, 1st edn. (CRC Press, 2016), ISBN 978-0521642224.
50 Zeng J, Lin R, Liu X, Zhao C, Cai Y, Review on partially coherent vortex beams, *Front. Optoelectron.* **12**, 229 (2019). https://doi.org/10.1007/s12200-019-0901-x.
51 Zhang Y, Cai Y, Gbur G, Control of orbital angular momentum with partially coherent vortex beams, *Opt. Lett.* **44**, 3617 (2019). https://doi.org/10.1364/OL.44.003617.
52 Liu X, Zeng J, Cai Y, Review on vortex beams with low spatial coherence, *Adv. Phys. X* **4**, 1626766 (2019). https://doi.org/10.1080/23746149.2019.1626766.
53 Zhang H, Lu X, Wang Z, Konijnenberg AP, Wang H, Zhao C, Cai Y, Generation and propagation of partially coherent power-exponent-phase vortex beam, *Front. Phys.* **9**, 781688 (2021). https://doi.org/10.3389/fphy.2021.781688.
54 Hyde IV MW, Korotkova O, Spencer MF, Partially coherent sources whose coherent modes are spatiotemporal optical vortex beams, *J. Opt.* **25**, 035606 (2023). https://doi.org/10.1088/2040-8986/acba2d.
55 Mandel L, Wolf E, Optical Coherence and Quantum Optics (Cambridge University Press, 1995), ISBN 9780521417112.
56 Siegman AE, Lasers (University Science Books, 1986), ISBN 0935702113.
57 Korotkova O, Gbur G, Unified matrix representation for spin and orbital angular momentum in partially coherent beams, *Phys. Rev. A* **103**, 023529 (2021). https://doi.org/10.1103/PhysRevA.103.023529.
58 Korotkova O, Pokharel S, OAM degree of coherence, *Opt. Lett.* **49**, 5103 (2024). https://doi.org/10.1364/OL.528291.
59 Korotkova O, Orbitalization ellipse of a light beam, *Opt. Lett.* **50**, 391 (2025). https://doi.org/10.1364/OL.545613.
60 Korotkova O, Orbitalization structure of random light beams, *J. Opt.* **27**, 065606 (2025). https://doi.org/10.1088/2040-8986/addc7c.
61 Born M, Wolf E, Principles of Optics, 7th (corrected) edn. (Cambridge University Press, 1999), ISBN 978-0521642224.
62 Gradshteiin IS, Ryzhik IM, Table of Integrals, Series, and Products, 4th edn. (Academic Press, 1965).
63 Abramowitz M, Stegun I, eds., Handbook of mathematical functions (Dover Publications Inc, 1972).
64 Prudnikov A, Brychkov Y, Marichev O, Integrals and Series, Vol. 1: Elementary Functions (Gordon and Breach Science Publisher, 1986).

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 58
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026051
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE

OPEN ACCESS

Coherence control and wavelength correction in multi-wavelength digital holography

Balasubrahmaniyam Mukundakumar* and Shimpei Matsuura*

Mitutoyo Research Center Europe B.V., Best, The Netherlands

Received 27 January 2026 / Accepted 15 June 2026

Abstract. Accurate surface measurement and alignment are increasingly vital in industrial manufacturing, where both extended axial range and sub-micrometer precision are essential. Multi-wavelength Digital Holography (MDH) achieves this by generating synthetic phase maps through digital combination of multiple laser wavelengths. This work presents two key innovations that enhance the robustness and practicality of MDH for industrial surface metrology. First, we introduce a coherence control technique using a dynamic mode-mixing process with a phase-randomizing mirror and bent multimode fiber to suppress coherent noise by tailoring the illumination coherence length. Second, we develop a post-processing algorithm that detects and compensates for wavelength shifts using an order parameter derived from residual phase maps, enabling accurate phase unwrapping even with unstable or low-cost laser sources. Together, these contributions enable MDH to deliver reliable, phase-jump-free heightmaps with sub-micrometer accuracy over mesoscopic ranges in realistic manufacturing environments.

Keywords: Digital holography, Coherence control, Wavelength correction, Surface metrology, Phase unwrapping, Synthetic wavelength.

1 Introduction

Accurate topological measurement and alignment of complex surfaces are becoming increasingly essential in modern industrial manufacturing, where precision directly influences product performance, reliability, and yield. In advanced sectors such as semiconductor packaging, MEMS, integrated photonics, and precision optics, the demand for characterizing surface topographies with both mesoscopic-range depth and sub-micrometer-scale accuracy is growing rapidly. For example, in semiconductor packaging, features like micro-bumps, through-silicon vias, and redistribution layers span tens to hundreds of micrometers in height, yet require nanometer-level precision for proper alignment and electrical connectivity. Traditional optical metrology techniques often struggle to balance range, resolution, and throughput, particularly in non-contact, high-throughput environments.

Multi-wavelength Digital Holography (MDH) offers a compelling solution to this challenge by extending the axial measurement range of digital holography while preserving its high resolution and throughput [1, 2]. MDH works by digitally "beating" phase signals from two or more laser wavelengths to generate synthetic phase maps at longer effective wavelengths [3, 4]. A carefully selected array of laser wavelengths is used to construct a hierarchy of synthetic wavelengths, where each longer synthetic wave unwraps the phase of the preceding shorter one. This cascading unwrapping process culminates in the longest synthetic wavelength, which defines the maximum unambiguous axial height that can be measured. As a result, MDH enables microscopic accuracy over mesoscopic ranges, making it ideally suited for applications that require both fine detail and extended depth coverage. Despite its advantages, implementing MDH under realistic environmental conditions presents two major challenges:

Sensitivity to Coherent Noise: Optical disturbances such as internal reflections, speckle, and diffraction from particles can degrade the phase signal, leading to measurement inaccuracies and reduced reliability.

Laser Instability and Wavelength Uncertainty: Accurate estimation of synthetic wavelengths is essential for reliable phase unwrapping. Instabilities in laser modes or fluctuations in wavelength can introduce errors in the scaling factors, resulting in phase discontinuities and incorrect height reconstructions. These errors become more pronounced at longer synthetic wavelengths, where even minor deviations can significantly impact measurement fidelity.

* Corresponding authors: b.mukundakumar@mitutoyo.nl; s.matsuura@mitutoyo.nl

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To overcome the first challenge, we propose a coherence control method using a phase-randomizing mirror (PRM) combined with a bent multimode fiber to induce efficient mode mixing. This configuration allows us to control the effective coherence length of the illumination – keeping it shorter than the cavity formations in the optical path to suppress multiple reflections, yet longer than the axial range to be measured. Along with its higher operational frequency, this approach provides stable, uniform extended laser illumination, effectively suppressing coherent noise with shorter exposure times, without compromising throughput or robustness. To address the second challenge, we developed a post-processing algorithm that detects and compensates for wavelength shifts. The algorithm uses the residual phase map – originating from the misestimation of spatial frequencies, to predict the correct order parameter and to quantify wavelength errors. This order parameter helps identify and localise phase inconsistencies, which are then used to correct the scaling factors and synthetic phase maps. The result is a set of accurate, phase-jump-free heightmaps, making MDH feasible even under environmental wavelength shifts and with more affordable, less stable laser sources.

2 Methodology

2.1 Coherence control for noise suppression

To mitigate sensitivity to coherent noise – such as speckle and parasitic fringes caused by internal reflections – we implemented a coherence control technique [5]. The system utilizes a Phase-Randomizing Mirror (PRM) [6] coupled with a bent multimode fiber to induce efficient mode mixing. This configuration allows for the precise tailoring of the illumination's effective coherence length.

The PRM operates at MHz frequencies, introducing phase modulations on timescales significantly shorter than the detector's integration time [6]. The rapid mirror oscillation imposes a time-varying optical path modulation on the reflected beam, so that within each camera exposure the illumination phase cycles through many independent states and coherent noise is suppressed by temporal averaging. This rapid modulation, combined with the bent multimode fiber acting as a mode scrambler, effectively averages out the speckle patterns and homogenizes the beam. Mode redistribution is achieved separately by the multiple bends in the multimode fiber: each bend introduces coupling between guided modes, progressively distributing optical power uniformly across all modes to produce the top-hat output profile. The multimode fiber output exhibits a "top-hat" intensity profile, which ensures uniform illumination across the field of view. The fiber end-face serves as a spatially extended source placed in the illumination arm; the object is illuminated in a free-space Köhler-like geometry in which the source is not re-imaged onto the object surface.

Crucially, this dynamic mode-mixing process (PRM along with Bent multi mode fiber) also reshapes the source's coherence profile. According to the Van Cittert-Zernike theorem [7, 8], the complex degree of coherence for a finite-size incoherent source is given by:

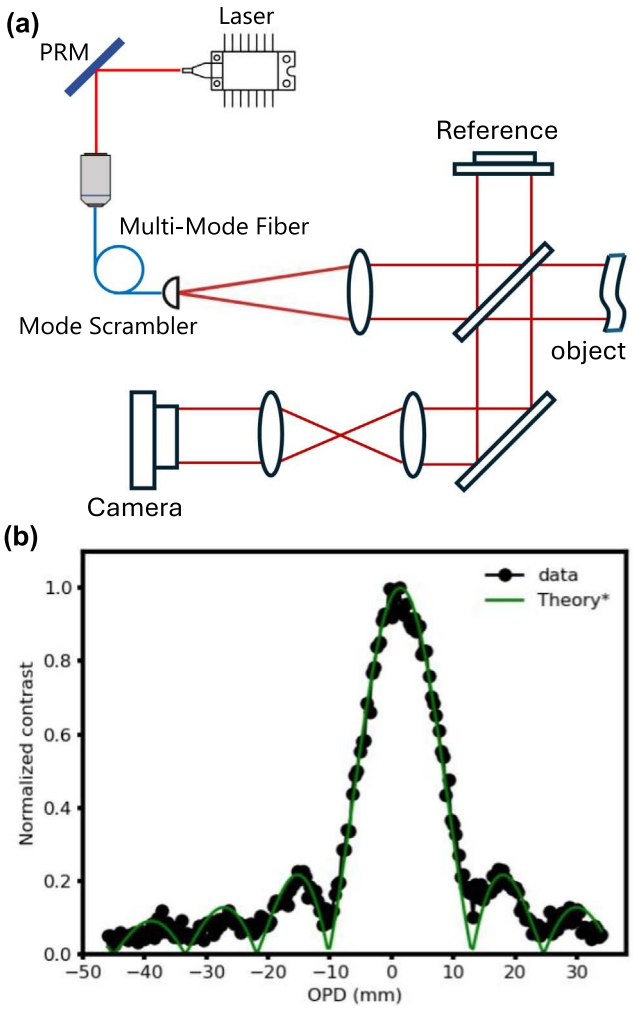


Fig. 1. Coherence control implementation and characterization. (a) Schematic of the digital holography setup featuring the Phase-Randomizing Mirror (PRM) and bent multimode fiber acting as a mode scrambler. (b) Measured interference contrast as a function of optical path length difference. The green curve represents the fitted coherence profile, demonstrating the tailored finite coherence length.

$$j(P_1, P_2) = \frac{2J_1(v)}{v} \tag{1}$$

where J_1 is the Bessel function of the first kind, and $v = k\beta^2 \cdot \mathrm{OPD}$ with $k = 2\pi/\lambda$ being the wave number and β the angular diameter of the source. For our system operating at $\lambda = 630$ nm with an effective source radius of approximately 300 µm (angular diameter $\beta \approx 0.01$ rad), the coherence function exhibits its characteristic Airy-disk profile with a main lobe width (coherence length) of $L_c \approx 10$ mm. This relationship allows precise control of the coherence length by tailoring the fiber's modal distribution through the bend radius [9]. In contrast to classical speckle averaging (e.g. a moving diffuser), where it is difficult to define or design the required coherence function, the well-defined top-hat profile produced by modal scrambling gives the source a stable coherence function with a predetermined

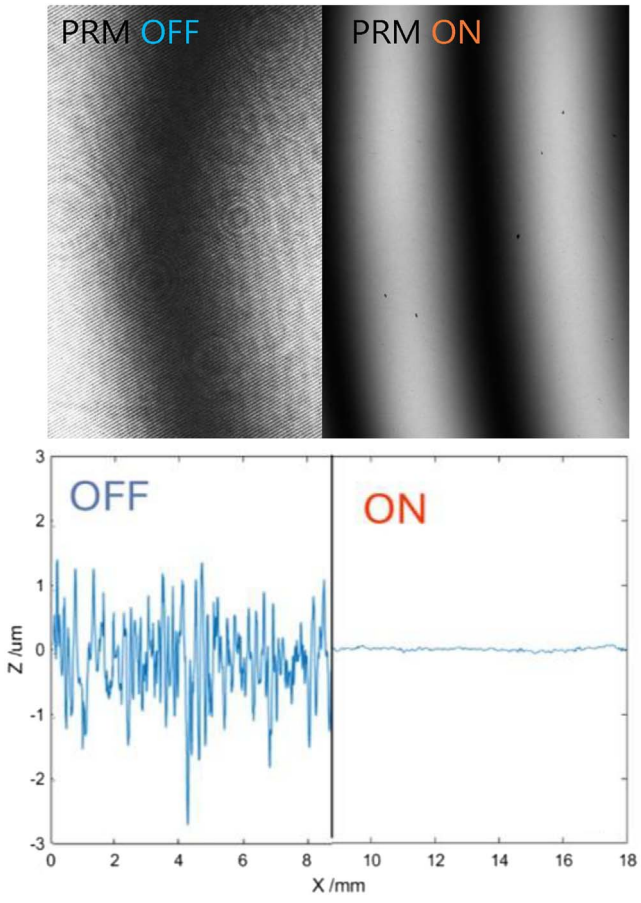


Fig. 2. Impact of coherence control on measurement quality. Comparison of interferograms (top) and extracted height map cross-sectional profiles (bottom) with PRM off (left) and PRM on (right), demonstrating effective suppression of coherent noise.

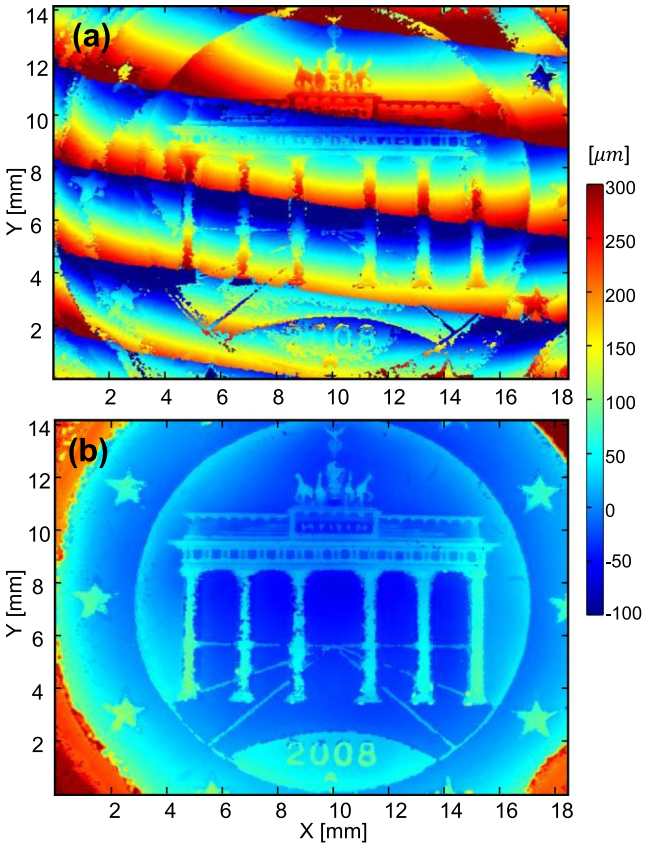


Fig. 3. Wavelength shift compensation results. (a) Height map without compensation, showing step discontinuities caused by wavelength instability. (b) Height map with residue-based compensation, demonstrating artifact-free reconstruction.

coherence length set by the MMF's NA and core diameter. Consequently, fringe contrast is fully preserved for the reference–object OPD while being suppressed to zero for all parasitic-cavity OPDs that exceed L_c. As illustrated in Figure 1b, the measured interference contrast reveals this tailored coherence function.

The design goal is to maintain a coherence length that is strictly shorter than the optical path differences of parasitic cavities (thereby suppressing multiple reflections) yet sufficiently longer than the axial measurement range of the sample. This "finite coherence" approach provides stable, uniform extended laser illumination. As shown in Figure 2, this method effectively suppresses coherent noise. The top row of Figure 2 compares the interferogram with the regular laser source, i.e., PRM off (left) and with PRM on (right). With PRM off, high-frequency fringes due to cavity formation within the optical system and bull's-eye patterns caused by dust particles on optical surfaces are clearly visible, resulting in significant high-frequency noise that degrades measurement quality. These coherent disturbances are effectively suppressed when the PRM is activated, as the reduced coherence length prevents interference from parasitic reflections and scattering centers. The bottom row shows extracted height cross-sectional profiles corresponding to the interferograms above, quantitatively demonstrating the dramatic reduction in noise amplitude and improved signal quality with coherence control enabled. This enables high-quality topography retrieval without compromising the system's throughput or robustness, which is often a problem with complex denoising algorithms in post-processing.

2.2 Wavelength shift compensation algorithm

The second major challenge in MDH is the accurate estimation of synthetic wavelengths, which is critically susceptible to laser instability. Understanding the hierarchical unwrapping strategy and its vulnerabilities is essential to appreciating our residue-based correction method.

Motivation for Hierarchical Unwrapping: In MDH, we construct at least one synthetic wavelength (often more) from two or more laser sources: a long synthetic wavelength Λ and a short wavelength (regular or synthetic) λ_s, where $\Lambda > \lambda_s$. The long synthetic wavelength is chosen such that Λ exceeds the maximum axial height of the object being measured. This critical design choice means the synthetic phase Φ_Λ is inherently unambiguous – it requires no unwrapping since the object height never exceeds one fringe period.

However, phase noise scales proportionally with the synthetic wavelength. While Φ_Λ provides unambiguous height information, it suffers from high noise levels. Conversely, ϕ_{λ_s} offers superior precision due to its shorter wavelength but is ambiguous – the measured phase wraps multiple times across the object's height range. The solution is to use the noisy-but-unambiguous Φ_Λ to determine the integer fringe order N of the precise-but-ambiguous ϕ_{λ_s}:

$$N = \left\lfloor \frac{\alpha\Phi_\Lambda - \phi_{\lambda_s}}{2\pi} \right\rceil \tag{2}$$

where $\alpha = \Lambda/\lambda_s$ is the scaling factor. The unwrapped phase is then:

$$\phi_{\text{unwrap}} = 2\pi\, N + \phi_{\lambda_s} \tag{3}$$

This approach determines integer fringe orders independently at each pixel [3, 4], fundamentally different from point-to-point spatial unwrapping methods [10] which are inherently unstable and cannot resolve absolute integer ambiguity.

The Wavelength Shift Problem: The hierarchical unwrapping relies on accurate knowledge of the scaling factor α. When a wavelength shift $\delta\lambda$ occurs, it predominantly affects the long synthetic wavelength Λ (since Λ depends on small wavelength differences) [11, 12]. The scaling factor becomes mis-estimated, and consequently the quantity $(\alpha\Phi_\Lambda - \phi_{\lambda_s})/2\pi$ is no longer an integer. This causes incorrect floor/rounding operations, producing step discontinuities in the unwrapped phase, as illustrated in Figure 3a.

Crucially, wavelength shifts affect shorter wavelengths proportionally less – the same scaling behavior exhibited by noise [13]. In our system, three lasers are used: λ_1 and λ_2 form the long synthetic wavelength $\Lambda = \lambda_1\,\lambda_2/|\lambda_1 - \lambda_2|$, while λ_1 and λ_3 form the short synthetic wavelength $\lambda_s = \lambda_1\,\lambda_3/|\lambda_1 - \lambda_3|$. Since $\delta\lambda \ll |\lambda_1 - \lambda_3|$, the short synthetic wavelength λ_s remains reliable and serves as a reference.

The Residue-Based Solution: Our method detects and corrects wavelength instability using the residual of the order parameter. This also allows for an explicit calculation of the deviation in wavelength (though not required for correcting height map). The complete algorithm is presented in Algorithm 1, and the compensation result is shown in Figure 3b. The key steps are detailed below.

Detailed Explanation: We define a complex phasor P that encodes the phase relationship:

$$P = e^{i\left(\alpha\Phi_\Lambda - \phi_{\lambda_s}\right)} \tag{6}$$

When wavelengths are correct (i.e. $\alpha = \Lambda/\lambda_s$ is accurate), the argument of P remains near zero across the image, with variations dominated by the noise. However, under wavelength shift – when the actual scaling factor deviates from the assumed α – the phasor P develops systematic spatial modulation. This modulation arises because an error in α directly translates to an error in the phase relationship encoded in P. Crucially, such wavelength-induced errors scale in a way that introduces systematic spatial frequency content: the mean spatial frequency (and higher cumulants) of the Fourier spectrum $|F\{P\}|$ shift away from zero.

Algorithm 1: Residue-based wavelength correction.

Input: Phase maps Φ_Λ (long, unambiguous) and ϕ_{λ_s} (short, precise; synthetic or regular); wavelengths are such that, $\Lambda > \lambda_s$.

Output: Corrected order parameter $N_{\text{corrected}}$; Unwrapped phase $\phi_{\text{unwrap}}\ (x,y)$

Step 1: Inputs (phases)
Begin with two phases aned their wavelengths:

$$\Phi_\Lambda\ (\text{long, unambiguous}),\ \phi_{\lambda_s}\ (\text{short, precise;}$$
$$\text{regular or synthetic}); \Lambda, \lambda_s$$

Example (synthetic construction):

$$\left(\phi_\Lambda = \phi_1 - \phi_2, \Lambda = \tfrac{\lambda_1\lambda_2}{|\lambda_1 - \lambda_2|}\right);$$
$$\left(\phi_{\lambda_s} = \phi_1 - \phi_3, \lambda_s = \tfrac{\lambda_1\lambda_3}{|\lambda_1 - \lambda_3|}\right);$$

Step 2: Scaling and order calculation
Calculate scaling factor and initial order parameter:

$$\alpha = \Lambda/\lambda_s, \quad N = \left(\alpha\Phi_\Lambda - \phi_{\lambda s}\right)/2\pi$$

Note: N is not rounded here; it contains both integer and fractional parts.

Step 3: Phasor construction
Construct the complex phasor encoding phase relationship:

$$P = e^{i\left(\alpha\Phi_\Lambda - \phi_{\lambda_s}\right)}$$

Step 4: Reference frequency extraction
Compute the intensity-weighted average spatial frequency from the Fourier spectrum:

$$k_0 = \frac{k_\parallel\ |F\{P\}|}{\langle |F\{P\}| \rangle} \tag{4}$$

where $\boldsymbol{k}_{||}$ denotes the in-plane spatial-frequency vector, $F\{\cdot\}$ denotes the Fourier transform, and $\langle\cdot\rangle$ represents averaging over the spectral domain.
The residue R is then calculated by removing this systematic modulation:

$$R = \arg\left\{P\ e^{-ik_0\cdot r}\right\} + k_0\cdot r$$

The magnitude $|R|$ indicates wavelength instability severity.

Step 6: Order correction and phase unwrapping
Correct the order parameter and compute unwrapped phase:

$$N_{\text{corrected}} = N - R\ (\text{integers})$$
$$\phi_{unwrap} = 2\pi\ N_{corrected} + \phi_{\lambda_s}$$

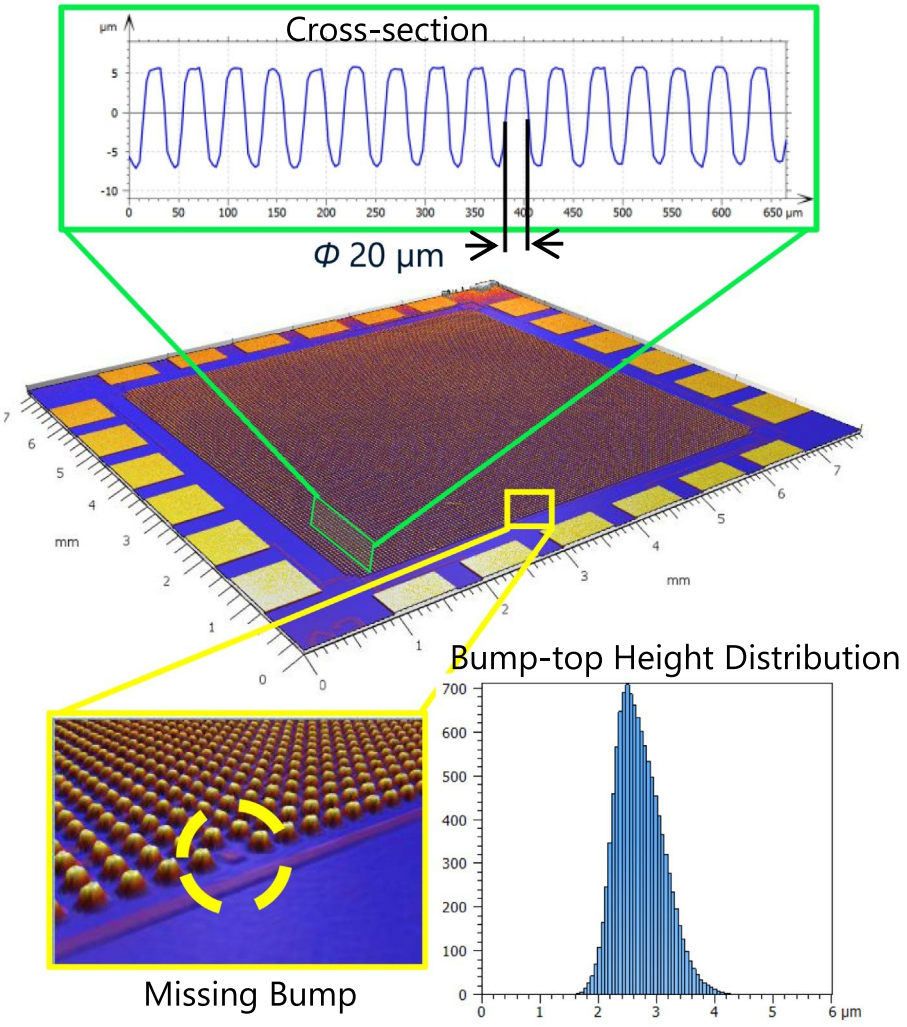


Fig. 4. Semiconductor micro-bump array measurement demonstrating system capability. 3D topography of >20,000 bumps (diameter ~20 µm) with missing bump detection (circled), bump-top coplanarity <0.6 µm, topography map, height distribution, and bump location identification.

Quantifying this spatial frequency deviation via the Fourier spectrum of P yields the exact correction needed for the order parameter, enabling accurate phase unwrapping despite wavelength instability.

The reference spatial frequency $\boldsymbol{k}_0$ is estimated from the Fourier spectrum of P using a weighted average:

$$\boldsymbol{k}_0 = \frac{\boldsymbol{k}_{\|}\, F\{P\}}{\langle F\{P\}\rangle} \tag{7}$$

The residue R is then calculated by removing this systematic modulation:

$$R = arg\{P\, e^{-i\boldsymbol{k}_0\cdot\boldsymbol{r}}\} + \boldsymbol{k}_0 \cdot \boldsymbol{r} \tag{8}$$

The magnitude $|R|$ directly flags the presence of wavelength instability. More importantly, the residue provides the exact correction needed for the order parameter:

$$N_{\text{corrected}} = N - R \tag{9}$$

where $N_{\text{corrected}}$ are now proper integers.

The corrected, wrap-free phase is obtained as:

$$\phi_{\text{unwrap}} = 2\pi\, N_{\text{corrected}} + \phi_{\lambda_s} \tag{10}$$

This approach maintains the high accuracy of the short synthetic wavelength while eliminating phase-jump artifacts caused by wavelength instability. As illustrated in Figure 3b, this compensation ensures the production of accurate, phase-jump-free heightmaps even when employing cost-effective laser sources that exhibit spectral instabilities. The residue-based correction directly adjusts the order parameter without requiring explicit wavelength recalculation, providing rapid and robust compensation suitable for high-throughput industrial applications.

3 Conclusion

Multi-wavelength Digital Holography (MDH) offers the unique capability of combining mesoscopic axial range with sub-micrometer precision – essential for industrial surface

metrology in semiconductor packaging, MEMS, and precision optics. However, two fundamental challenges have hindered its practical deployment: (1) coherent noise from internal reflections and scattering degrades phase quality, and (2) laser wavelength instabilities cause catastrophic phase-jump artifacts in the hierarchical unwrapping process.

We have addressed both challenges. Our coherence control technique using the dynamic mode-mixing process tailors the illumination coherence length ($L_c \approx 10$ mm) to suppress parasitic interference while preserving measurement capability – eliminating spurious fringes and bull's-eye patterns. Our residue-based wavelength correction algorithm exploits two key insights: first, wavelength shifts affect shorter synthetic wavelengths proportionally less, providing a stable reference; second, these shifts manifest as spatial frequency deviations in the phasor spectrum, which we quantify through the residue to directly correct the order parameter. The algorithm flags and compensates for laser wavelength instabilities (up to ∼20% of the longer synthetic wavelength) in post-processing, effectively correcting heightmap errors due to phase jumps without iterative optimization.

Figure 4 demonstrates these innovations on a semiconductor micro-bump array: over 20,000 bumps (diameter ∼20 µm) measured with coplanarity precision <0.6 µm across a millimeter-scale field, with automatic missing bump detection. By ensuring accurate, phase-jump-free heightmaps despite environmental wavelength shifts, our techniques make MDH more accessible and viable for inline industrial environments – previously requiring expensive wavelength-stabilized lasers, such measurements are now achievable with cost-effective sources.

Acknowledgments

The authors thank NTS Optel B.V. for their support of this research.

Funding

This research was funded by Mitutoyo Corporation. No external academic grants were received for this specific study.

Conflicts of interest

B. Mukundakumar and S. Matsuura are both employed by Mitutoyo Research Center Europe, which is a subsidiary of Mitutoyo Corporation. This work was conducted as part of their regular employment duties. Both authors declare that a patent application has been filed based on findings presented in this manuscript. Both authors certify that they have no other financial conflicts of interest (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements outside their employment) in connection with this article.

Data availability statement

The experimental data and phase maps supporting the results reported in this study are available upon reasonable request from the corresponding authors.

Author contribution statement

B. Mukundakumar and S. Matsuura contributed equally to this work. Both authors conceived and designed the study. Both authors contributed to manuscript writing. Both authors have read and approved the final version of the manuscript.

References

1 Osten W, Reingand N, *Optical Imaging and Metrology: Advanced Technologies*, (Wiley-VCH, 2014). https://doi.org/10.1002/9783527648443
2 Kim MK, Principles and techniques of digital holographic microscopy, *SPIE Rev.* **1**, 018005 (2010). https://doi.org/10.1117/6.0000006
3 Gass J, Dakoff A, Kim MK, Phase imaging without 2π ambiguity by multiwavelength digital holography, *Opt. Lett.* **28**, 1141–1143 (2003). https://doi.org/10.1364/OL.28.001141
4 Parshall D, Kim MK, Digital holographic microscopy with dual-wavelength phase unwrapping, *Appl. Opt.* **45**, 451–459 (2006). https://doi.org/10.1364/AO.45.000451
5 Bianco V, Memmolo P, Leo M, Montresor S, Distante C, Paturzo M, Picart P, Javidi B, Ferraro P, Strategies for reducing speckle noise in digital holography, Light: Sci. *Appl.* **7**, 48 (2018). https://doi.org/10.1038/s41377-018-0050-9
6 Shevlin F, Phase randomization for spatiotemporal averaging of unwanted interference effects arising from coherence, *Appl. Opt.* **57**, E6–E10 (2018). https://doi.org/10.1364/AO.57.0000E6
7 Goodman JW, *Statistical Optics*, (Wiley, 2015).
8 Wolf E, *Introduction to the Theory of Coherence and Polarization of Light* (Cambridge University Press, 2007).
9 Auksorius E, Borycki D, Wojtkowski M, Multimode fiber enables control of spatial coherence in Fourier-domain full-field optical coherence tomography for in vivo corneal imaging, *Opt. Lett.* **46**, 1413–1416 (2021). https://doi.org/10.1364/OL.417178
10 Ghiglia DC, Pritt MD, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software* (Wiley, 1998).
11 Mann CJ, Bingham PR, Paquit VC, Tobin KW, Quantitative phase imaging by three-wavelength digital holography, *Opt. Express* **16**, 9753–9764 (2008). https://doi.org/10.1364/OE.16.009753
12 Kuhn J, Colomb T, Montfort F, Charriere F, Emery Y, Cuche E, Marquet P, Depeursinge C, Real-time dual-wavelength digital holographic microscopy with a single hologram acquisition, *Opt. Express* **15**, 7231–7242 (2007). https://doi.org/10.1364/OE.15.007231
13 Carl D, Fratz M, Pfeifer M, Giel DM, Hoefler HJ, Multi-wavelength digital holography with autocalibration of phase shifts and artificial wavelengths, *Appl. Opt.* **48**, H1–H8 (2009). https://doi.org/10.1364/AO.48.000H1

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 26
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026029
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Study of tapered microfibers by line-field confocal optical coherence tomography

Arnaud Dubois, Aloïs Baudry, and Sylvie Lebrun*

Université Paris-Saclay, Institut d'Optique Graduate School, CNRS, Laboratoire Charles Fabry, 91127, Palaiseau, France

Received 19 January 2026 / Accepted 14 March 2026

Abstract. We present for the first time to our knowledge measurements of the evolution of the core and cladding diameters in tapered silica microfibers using LC-OCT. The results could help refine models of propagation of optical modes in tapers.

Keywords: Silica microfibers, Optical tapers, Line-field confocal optical coherence tomography, Silica optical fibers.

1 Introduction and motivations

Optical microfibers are typically produced by tapering standard silica fibers – commonly used in telecommunications – until their diameters become comparable to or smaller than the wavelength of the propagating light. The uniform section of the resulting structure, referred to as the microfiber, is connected to the untapered fiber parts through two conical sections known as tapers (see Fig. 1). These devices have been extensively used for over three decades in both scientific research and engineering, serving as fundamental optical devices that are inherently compatible with all-fiber networks [1]. Owing to their unique properties, microfiber-based technologies have proven to be versatile across a wide range of applications, from fundamental studies to advanced implementations such as quantum information systems [2, 3], remote sensing devices [4], and nonlinear optics in the silica itself [5, 6] or in the surrounding medium via the evanescent field [7, 8].

A key challenge in microfiber fabrication is ensuring the adiabaticity of the taper regions. An adiabatic taper is characterized by a sufficiently small taper angle to suppress coupling between the guided fundamental mode and unwanted higher-order modes, thereby preserving high transmission efficiency of the desired mode. The standard adiabaticity criterion is based on the comparison of two characteristic lengths. The first one is the beat length z_b between the fundamental LP_{01} mode and the nearest higher-order mode with the same azimuthal symmetry, namely the LP_{02} mode. This beat length is obtained from the difference between the longitudinal propagation constants of the two modes $\beta_{01}(z)$ and $\beta_{02}(z)$ along the taper and is given by $z_b = \frac{2\pi}{\beta_{01}(z)-\beta_{02}(z)}$.

The second characteristic length is the local taper length z_t, which can be approximated as $z_t \approx \frac{\Omega(z)}{r(z)}$, where $r(z)$ denotes the local cladding radius and $\Omega(z)$ the local taper angle [9]. When the condition $z_t \gg z_b$ is satisfied at every position z along the taper – implying negligible intermodal coupling – the taper can be considered adiabatic for the fundamental mode. Consequently, designing an adiabatic taper profile requires calculating the propagation constants of the relevant modes throughout the taper as functions of z, implying a precise knowledge of the refractive index profile.

In this study, we consider conventional step-index optical fibers consisting of a Ge:SiO_2-doped core with refractive index n_{core}, surrounded by a pure SiO_2 cladding with refractive index n_{clad}. Current theoretical models for the calculation of the propagation constants in such fibers rely on the following assumptions: (1) the ratio between the cladding and core diameters remains constant throughout the taper; (2) the refractive indices of both the core and the cladding are uniform along the taper; (3) the core effectively vanishes at a critical diameter, beyond which light guidance transitions from the core-cladding interface to the cladding–external medium interface. In general, for step-index telecommunication fibers, the critical cladding diameter is typically calculated to be around 40 µm [9]. Typically, to design optimized tapers, three steps should be considered: the decrease of the diameter should be firstly sharp until reaching the critical diameter, then much smoother when passing this diameter to limit the coupling with higher order modes, then sharp again. This highlights the importance to determine also the position and the value of this diameter.

In this article, for the first time to our knowledge, we report on measurements of the tapered section of a silica microfiber performed using Line-field Confocal Optical

* Corresponding author: sylvie.lebrun@institutoptique.fr

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

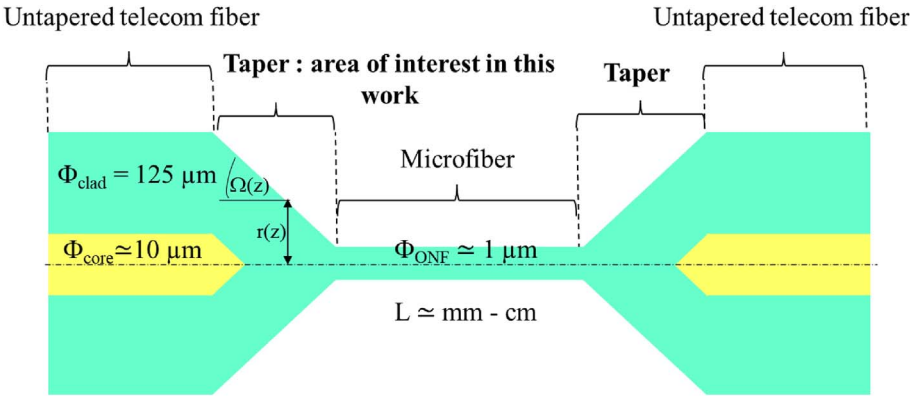


Fig. 1. Schematic diagram of a microfiber and its two tapers linking the uniform part to the untapered fiber. Φ_{clad} and Φ_{core} denote the cladding and core diameters, respectively, as indicated here for conventional single-mode step-index fibers used in telecommunications.

Coherence Tomography (LC-OCT), aiming to assess the validity of the existing models.

Observing the evolution of both the core and cladding over distances of a few centimeters with high resolution and sensitivity requires sophisticated techniques tailored to semi-transparent media. Compared with nano- and microtomography methods, which offer higher sensitivity [10], LC-OCT provides significantly faster acquisition times (on the order of tens of minutes to scan a ∼2 cm-long taper). Moreover, LC-OCT enables microscale characterization of macroscopic objects, whereas the above-mentioned methods are typically limited to lengths of only a few millimeters. Quantitative phase microscopy can also be used to obtain phase maps of the taper; however, unlike the approach presented here, the resulting images are not directly amenable to physical interpretation [11]. Altogether, these advantages make LC-OCT a complementary and powerful technique for investigating the evolution of the core and cladding in the tapered section of a microfiber, with a micrometer-scale resolution and high detection sensitivity.

This article is organized as follows:

After describing the experimental setup, we report results obtained from two different tapered sections fabricated from step-index fibers designed for telecommunications (Alcatel Clear Incolore and Corning SMF28e). Our measurements show that the assumptions commonly made for numerical modelling need to be refined for a more accurate description of the evolution of propagation modes in the tapered section.

2 Experimental setup

LC-OCT is an optical imaging technique that combines confocal detection with low-coherence interferometry to provide real-time, cross-sectional images of semi-transparent samples [12]. The LC-OCT system used in this work, shown schematically in Figure 2, is routinely employed for cellular-level skin imaging [13, 14]. Micrometer-scale lateral and axial resolution is achieved using two identical high numerical aperture (0.5) immersion microscope objectives and broadband light with a central wavelength of 750 nm. Illumination is provided by a fiber-coupled supercontinuum light source, whose output is collimated and shaped into a line by a cylindrical lens before entering the interferometer via a beam splitter. The line of light is focused onto the sample in the object arm of the interferometer and onto a glass plate in the reference arm. The interference signal is detected by a line-scan camera, and a vertical cross-sectional image is reconstructed from a stack of lines acquired by the camera as the depth is scanned using a piezoelectric translation stage [12, 14]. The refractive index step $\Delta n = n_{core} - n_{clad}$ between the fiber core and the cladding being very small (typically a few 10^{-3}), the resulting optical reflection to be detected is very weak ($\sim 10^{-5}$). To enhance the interferometric signal contrast and thereby improve detection sensitivity, the reflection coefficient of the interferometer reference surface was reduced to 0.03% by immersing it in water.

The microfibers are drawn using the classical pull-and-brush technique with a butane flame, producing high-optical-quality microfibers with symmetrical tapers in a very reproducible manner as was shown in previous studies [6, 7, 15]. Two silica step-index single-mode fibers designed for telecommunications are studied (Alcatel Clear Incolore and Corning SMF28e, respectively named fiber 1 and fiber 2). The initial cladding diameter of both fibers is 125 µm. The characteristics of fiber 2 are well-known. Its initial core diameter is 8.2 µm, and its step-index is $\Delta n = 5.2 \times 10^{-3}$ [11]. Although the parameters of fiber 1 (diameters and step-index) are less well characterized, they are known to be similar. In both cases, the targeted microfiber diameter is 1.71 µm. The length of the uniform part of the microfibers is 1 cm, and the length of the tapers is 1.8 cm. The microfiber and its two tapers are immersed in a tank filled with silicone oil (refractive index ≃1.45) to attenuate optical reflections at the fiber surface (see Fig. 2). Consequently, the contrast in optical reflection between the cladding surface and the core surface is reduced. A series of cross-sectional LC-OCT images of the taper is acquired at positions separated by 0.5 mm. Each cross-sectional image

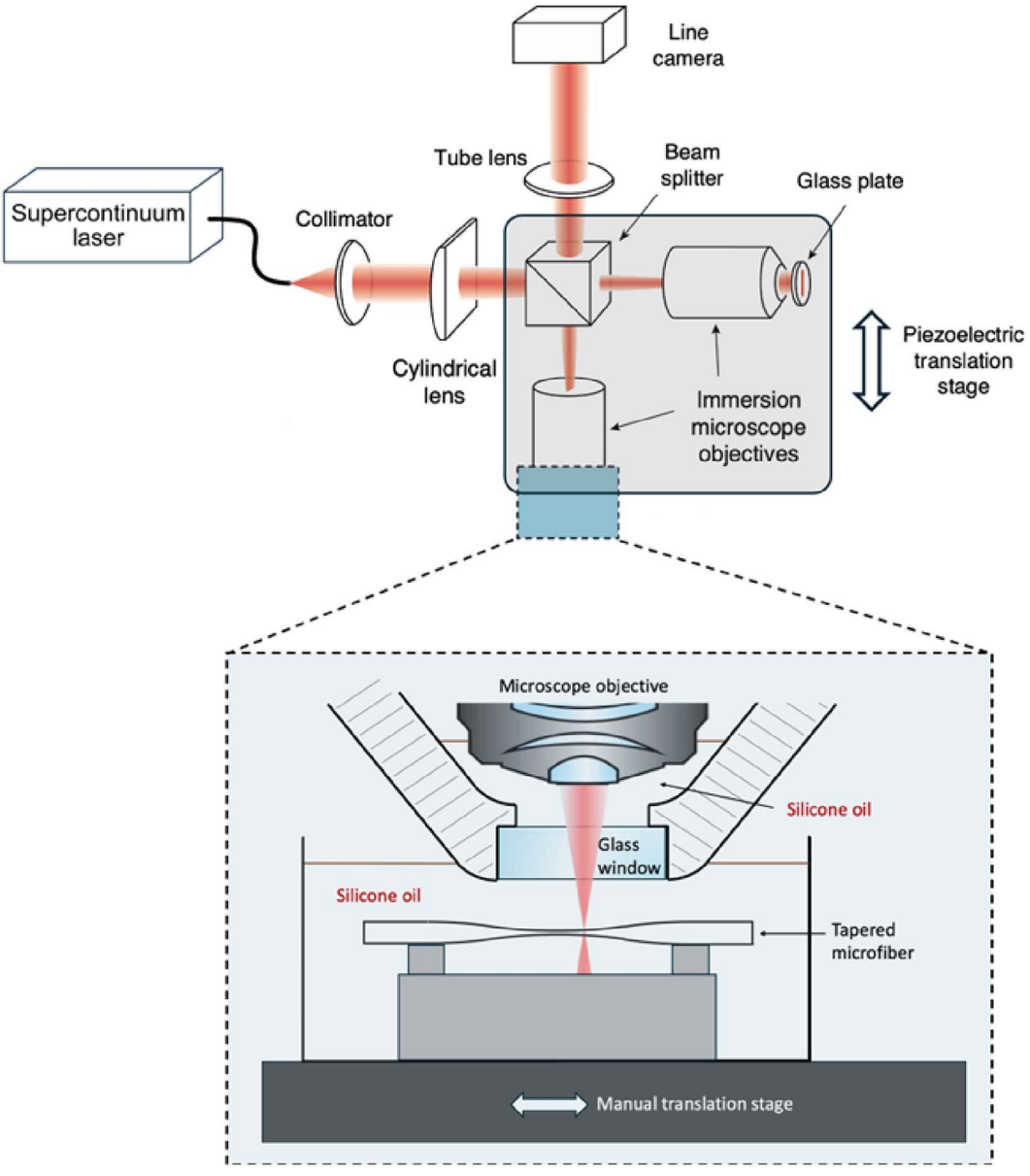


Fig. 2. LC-OCT setup. The tapered microfiber under study is not shown to scale and is enlarged for illustration purposes.

is generated by averaging 100 frames to reduce photon shot noise. The device is mechanically stable, ensuring that this averaging does not introduce blur that could degrade spatial resolution.

3 Results and discussion

Figures 3a–3e show cross-sectional images of the tapered section of fiber 1 for different cladding diameters. The white traces indicate the cladding, and the small points at the center of the images in Figures 3a–3c indicate the core. Because the cladding and core surfaces are smooth, the optical reflection is predominantly specular. Only light reflected from surface regions that are nearly perpendicular to the illumination direction is detected. Fitting the visible cladding boundary in the image with a circle confirms its circular geometry, although only a portion of it is observed. From the initial diameter of 125 µm down to a cladding diameter of approximately 67 µm, the cladding/core interface could still be observed despite the small step index. The resolution of the LC-OCT device enabled to measure the diameter of the thinnest part, namely the microfiber (Figs. 3e and 3f), yielding a diameter value of 1.6 ± 0.1 µm. This measurement is in very good agreement with the value obtained by conventional optical microscopy using an immersion microscope objective (see Fig. 3g), giving a diameter of 1.5 ± 0.1 µm. Both measurements are also in good agreement with the targeted diameter of 1.71 µm. Performing measurements by Scanning Electron

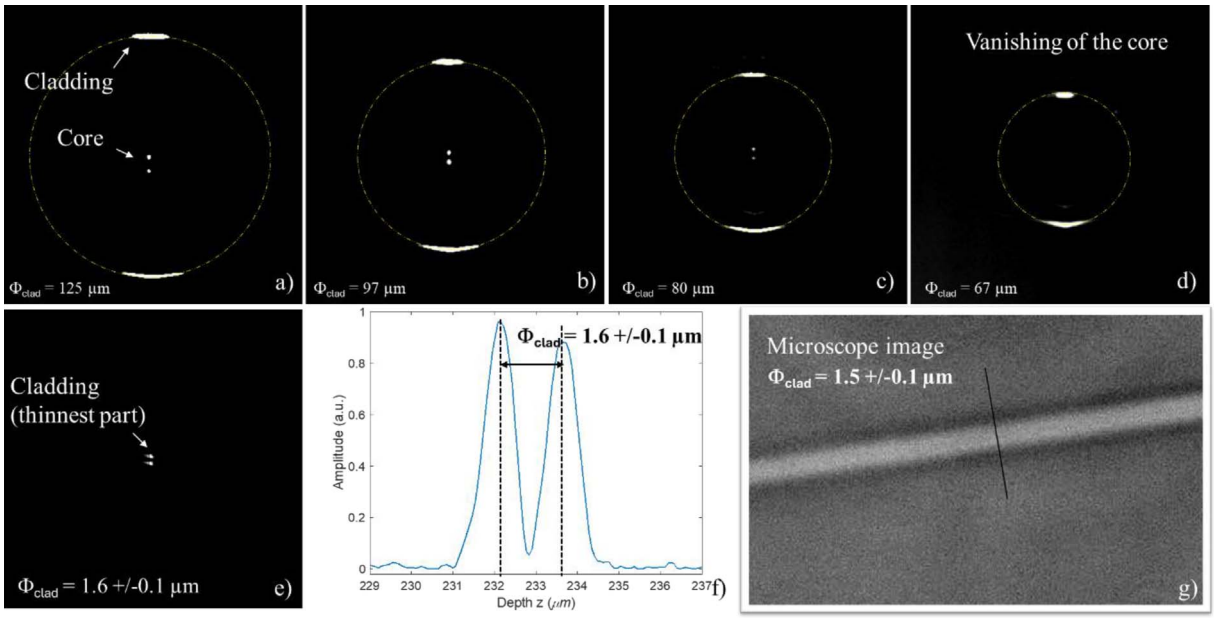


Fig. 3. From (a) to (e). Cross-sectional images of the tapered section of fiber 1 for 5 different cladding diameters (125, 97, 80, 67 and 1.6 µm). The yellow dotted lines represent a fit of the external boundary of the fiber with a circle. (f) Measurement of the thinnest part diameter obtained by LC-OCT. (g) Optical microscopy image of the thinnest part of the fiber.

Microscopy (SEM) provides much higher resolution but is far more complex to implement than this optical method. Moreover, in our previous works, we have performed SEM measurements that have shown that our drawing machine yields very good results in terms of diameter accuracy and homogeneity [7, 15].

In Figure 4a, we report the evolution of the cladding and core diameters along the taper. As the diameters decrease, the reflectivities also decrease due to the progressive diffusion of the Ge:SiO_2 core during the tapering process, leading to a reduction of the step index value. In particular, the smallest core diameter that could be measured is 1.85 µm for a cladding diameter of 67 µm. Below this value, the detection limit of LC-OCT is reached [12], corresponding to a minimum reflectivity $R_{min} = \left(\frac{\Delta n}{n_{\text{core}} + n_{\text{clad}}}\right)^2 \simeq 10^{-9}$, which is obtained for a refractive-index step of $\Delta n \simeq 10^{-4}$.

In Figure 4b, we plot the ratio of the core to cladding diameters, $R = \frac{\Phi_{\text{core}}}{\Phi_{\text{clad}}}$. This ratio is not constant, varying from 0.064 to 0.05 over the first 4.5 mm in the taper. It then decreases abruptly to 0.027 for the last point that could be measured.

Our hypothesis is that there are two distinct effects occurring during the tapering process: the geometrical reduction of the initial core size and the expansion of the region of higher refractive index, due to the diffusion of Ge into the cladding in the vicinity of the initial core. We still name this region "core". These two combined effects could explain the behavior observed in Figure 4b, that can be separated into three parts:

- At the beginning, until $z = 2.5$ mm, the increase in core size due to diffusion is negligible compared to the reduction in core size from the fiber stretching, giving an almost constant ratio R.
- After $z = 2.5$ mm, there is a more pronounced increase in the ratio R until $z = 3.5$ mm because of a more important diffusion reducing the core diameter decrease speed.

Then after $z = 3.5$ mm, R decreases sharply because of both the reduction of the geometrical size of the core and the fact that the index difference Δn is flattened due to diffusion and becomes lower than the detection limit.

The same measurements were performed with a taper drawn from fiber 2. The results are shown in Figure 5.

Even for large cladding diameters, the core was more difficult to observe than for fiber 1, which we attribute to a smaller initial step index. The step index for the untapered fiber 2 being known ($\Delta n = 5.2 \times 10^{-3}$), we deduce that the step index of untapered fiber 1 should be higher. Determining its exact value would require a calibration of the LC-OCT system, which we were not able to do for practical reasons. As for fiber 1, the vanishing of the core corresponds to the limit of detection of the LC-OCT device. We conclude that the range of variation of the step index we were able to measure for fiber 2 extends from 5.2×10^{-3} to $\sim 10^{-4}$. In Figure 5a, the core vanishes for a cladding diameter of 91 µm, and the smallest core diameter that could be measured is 2.7 µm. The smallest cladding diameter measured is 1.8 µm, in good agreement with the targeted diameter of 1.71 µm. In Figure 5b, we report the ratio of the core and cladding diameters. As for fiber 1, this ratio decreases, from 0.072 to 0.023. The decrease appears more monotonic than for fiber 1, with a two step behavior without any increase of the ratio R, suggesting that the diffusion of Ge does not counter balance the effect of the geometrical reduction of the core size. Further investigations

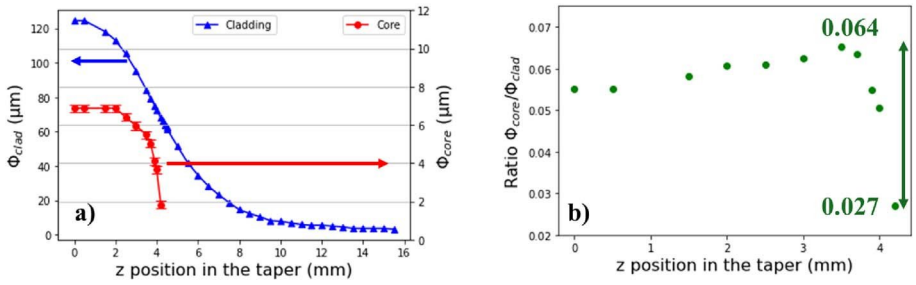


Fig. 4. (a) Evolution of the cladding (blue triangles) and core diameters (red dots) versus the z position in the taper drawn from fiber 1. Error bars on Φ_{core} are ± 0.1 µm (smaller than the size of the cladding marks). (b) Evolution of the ratio of the core and cladding diameter $R = \Phi_{\text{core}}/\Phi_{\text{clad}}$ versus the z position in the taper in the area where the core diameter is measurable. Error bars are smaller than the size of the marks (± 0.1%).

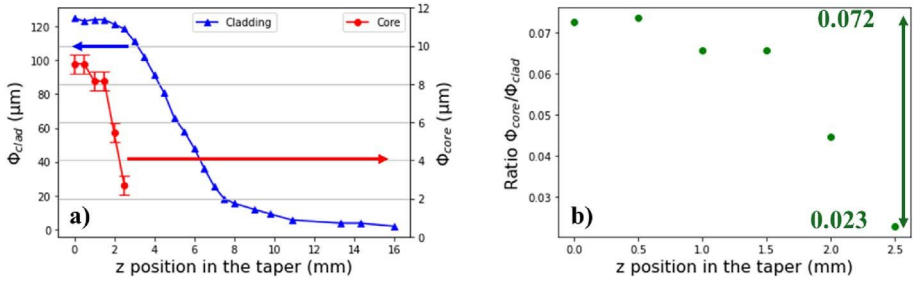


Fig. 5. (a) Evolution of the cladding (blue triangles) and core diameters (red dots) versus the z position in the taper drawn from fiber 2. Error bars on Φ_{core} are ± 0.5 µm (smaller than the size of the cladding marks). (b) Evolution of the ratio of the core and cladding diameter $\Phi_{\text{core}}/\Phi_{\text{clad}}$ versus the z position in the taper in the area where the core diameter is measurable. Error bars are smaller than the size of the marks (± 0.1%).

based on measurement of the chemical composition of the taper (synchrotron or X-rays for instance) would be necessary to confirm this assumption. Moreover, fewer data points were acquired, which prevents a definitive comparison. Similar measurements were performed on 3 identical microfibers drawn from fiber 1 and 2 from fiber 2, showing the same results.

4 Conclusion and perspectives

To our knowledge, we report for the first time measurements of the evolution of the cladding and core diameters along the tapered section of a microfiber fabricated using the classical pull-and-brush technique, obtained by LC-OCT. This imaging technique is particularly attractive, as it does not require complex preparation of the samples and enables fast high-sensitivity measurements with micrometer-scale resolution of semi-transparent objects such as silica tapers.

The results show that, contrary to assumptions commonly used in numerical models, the ratio of the core to cladding diameters is not constant along the taper. We attribute this behavior to both the fiber origin and the fabrication process. Moreover, the step-index is found to decrease, at least down to a value of $\sim 10^{-4}$, corresponding to the detection limit of LC-OCT. At last, for both fibers, we measured that the critical diameter is not obtained for a diameter of 40 µm as currently used but depends on the single mode fiber: for fiber 1, this diameter is 67 µm whereas for fiber 2, it is 91 µm. These results can be used to optimize the shape of the tapers and to calculate more precisely the effective index of the propagation mode in the taper.

Future work will focus on collecting additional data to obtain reliable statistics and on investigating other types of optical fibers (e.g., polarization-maintaining fiber, visible single-mode fibers, fibers with other silica dopings, ...). These results will enable the refinement of numerical models of mode propagation in tapers, particularly for the design and optimization of adiabatic taper profiles.

Funding

The work was supported by internal fundings.

Conflicts of interest

The authors declare that there are no conflicts of interest related to this article.

Data availability statement

This article has no associated data generated.

Author contribution statement

A. Dubois and S. Lebrun collected and processed the data and wrote the article. A. Baudry and S. Lebrun fabricated the nanofibers. A. Dubois, A. Baudry and S. Lebrun discussed the results and reviewed the article.

References

1 Tong LM et al., Subwavelength-diameter silica wires for low-loss optical wave guiding, *Nature* **426**, 816–819 (2003) https://doi.org/10.1038/nature02193.

2 Berroir J et al., Ultralow-power single-pass all-optical photon router, *Optica* **12** (8), 1250–51 (2025). https://doi.org/10.1364/OPTICA.569736.

3 Horikawa S et al., Low-loss telecom-band nanofiber cavity for interfacing Yb atomic qubits, *Opt. Lett.* **50** (17), 5294–5298 (2025) https://doi.org/10.1364/OL.570362.

4 Zhang L, et al., Micro-/nanofiber optics: merging photonics and material science on nanoscale for advanced sensing technology, *iScience* **23**(1), 100810 (2020). https://doi.org/10.1016/j.isci.2019.100810.

5 Birks TA, et al., Supercontinuum generation in tapered fibers, *Opt. Lett.* **25**(19), 1415–1417 (2000). https://doi.org/10.1364/ol.25.001415. PMID: 18066233.

6 Delaye P, et al., Continuous-wave generation of photon pairs in silica nanofibers using single-longitudinal and multilongitudinal-mode pumps, *Phys. Rev. A* **104**, 063715 (2021). https://doi.org/10.1103/PhysRevA.104.063715.

7 Bouhadida M et al., Highly efficient and reproducible evanescent Raman converters based on a silica nanofiber immersed in a liquid, *Appl. Phys. B* **125**, 228 (2019). https://doi.org/10.1007/s00340-019-7340-7.

8 Fanjoux G et al., Demonstration of the evanescent Kerr effect in optical nano fibers, *Opt. Express* **27**, 29460–29470 (2019). https://doi.org/10.1364/OE.27.029460.

9 Love JD et al., *IEE Proc. Pt. J. Optoelect.* **138** (5), 343–354 (1991). https://doi.org/10.1049/ip-j.1991.0061.

10 Crocco MC et al., Multiscale X-ray computed tomography of standard optical fibers, *Tomogr. Mater. Struct.* **9**, 100078 (2025) https://doi.org/10.1016/j.tmater.2025.100078.

11 Belanger E et al., Comparative study of quantitative phase imaging techniques for refractometry of optical waveguides, *Opt. Express* **26** (13), 17498 (2018) https://doi.org/10.1364/OE.26.017498.

12 Dubois A et al., Line-field confocal time-domain optical coherence tomography with dynamic focusing, *Opt. Express* **26**, 33534–33542 (2018). https://doi.org/10.1364/OE.26.033534.

13 Dubois A et al., Line-field confocal optical coherence tomography for high-resolution noninvasive imaging of skin tumors, *J. Biomed. Opt.* **23**, 106007 (2018). https://doi.org/10.1117/1.JBO.23.10.106007.

14 Latriglia F et al., Line-field confocal optical coherence tomography (LC-OCT) for skin imaging in dermatology, *Life* **13**, 2268 (2023). https://doi.org/10.3390/life13122268.

15 Shan L et al., Stimulated Raman scattering in the evanescent field of liquid immersed tapered nanofibers, *Appl. Phys. Lett.* **102** (20), 201110 (2013). https://doi.org/10.1063/1.4807170.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 23
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026016
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

Review Article OPEN ACCESS

Fano-like resonance in optical fiber

Vinicius Piaia, Paulo Robalinho, Susana Silva, and Orlando Frazão*

INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, 4169-007 Porto, Portugal

Received 24 November 2025 / Accepted 18 February 2026

Abstract. Fano resonances emerge from the coherent interference between a discrete resonant state and a continuum of propagating modes, giving rise to a characteristically asymmetric spectral line shape with heightened sensitivity to minute variations in the underlying physical parameters. This study provides a chronological perspective on the development and increasing implementation of Fano-type effects in fiber-optic technology. Their implementation in fiber Bragg gratings (FBGs) through numerical simulations of resonant Fano behavior. Specifically, a Fano-like response in an FBG can be designed by introducing a tailored phase shift into the grating structure; the magnitude of this phase discontinuity constitutes an effective control parameter for tuning the interference condition and, consequently, the resulting spectral asymmetry. The resonance produces a distinctive, asymmetrical spectral response that is attractive for a broad range of photonic functionalities. Key applications include fiber-integrated sensing – where the enhanced spectral sensitivity supports the detection of changes in refractive index, temperature, pressure, or biomolecular interactions – as well as narrowband filtering and spectral shaping for telecommunications and optical signal-processing systems.

Keywords: Fano-like resonances, Optical fibers, Gratings structures and interferometry.

1 Introduction

In the literature, Fano resonances have been extensively reported across a wide range of photonic systems, underscoring their fundamental importance in controlling light–matter interactions and enabling novel optical functionalities. Representative examples include artificial photonic molecules [1], Fano lasers [2], metamaterials [3], nanophotonic structures [4], photonic crystals [5], semiconductor nanostructures [6], plasmonic nanoantennas [7], optical switches [8], all-optical modulators [9], dielectric metasurfaces [10], and topological photonic systems [11], among others. The ubiquity of Fano-type interference across such diverse platforms highlights both the universality and tunable nature of this phenomenon, which continues to attract significant interest for applications in optical filtering, switching, nonlinear photonics, and high-resolution sensing.

Despite these remarkable advances, many existing implementations rely on intricate geometries or complex fabrication techniques, which can restrict scalability and practical deployment. In this context, fiber-based platforms offer unique advantages, combining the inherent merits of optical fibers – such as low transmission loss, compact form factor, mechanical flexibility, and compatibility with existing telecommunication infrastructures – with the enhanced spectral sensitivity and selectivity enabled by Fano resonances.

This review traces the evolution of research on Fano resonance in fiber-optic systems, revealing a clear trajectory from the first fundamental demonstrations in the mid-2000s to the emergence of advanced, tunable, and application-driven devices after 2020. It then presents the main structures used to induce Fano resonance in optical fiber devices, such as whispering-gallery-mode resonators, microstructured optical fibers, interferometric devices, and FBGs. To exploit FBG applications, a theoretical model of a Fano-FBG is introduced, which captures the spectral evolution as either the FBG reflectivity or the fiber-end environment reflectivity is varied. Finally, the model is demonstrated in a self-referenced refractive-index sensor as one of the possible applications.

2 Fano concept

A deeper understanding of the wave-like behavior of light has enabled the development of advanced optical technologies. Among the most significant properties exploited in these innovations is resonance, which has played a central role in many of the scientific community's major achievements. Resonant cavities constitute the foundation of widely used optical devices, including lasers, Fabry–Pérot

* Corresponding author: orlando.frazao@inesctec.pt

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

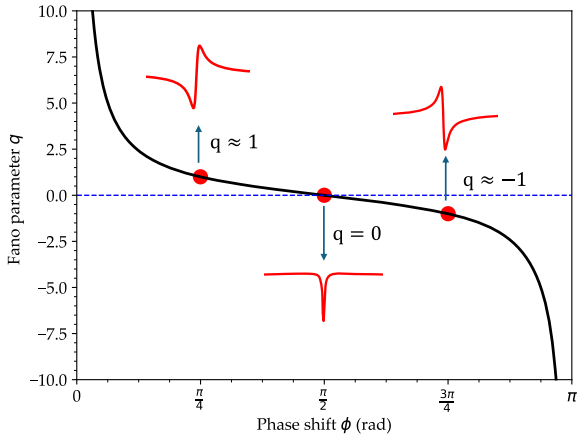


Fig. 1. Dependence of the Fano parameter q on the phase shift of the continuum mode, with the resulting line shapes: asymmetric for $q \approx \pm 1$ and quasi-Lorentzian for $q = 0$.

interferometers (FPI), and Bragg gratings. Fano resonance was first demonstrated by Ugo Fano in his investigation of inelastic electron scattering by helium [12], revealing an asymmetric line shape with a high-quality factor. It was established that this resonance arises when a discrete mode spectrum interferes with a continuous mode spectrum. This interaction is described by the spectrum $\sigma(E)$, as expressed in equation (1) [13].

$$\sigma(E) = 4\sin^2\phi \frac{(q+\Omega)^2}{1+\Omega^2}. \quad (1)$$

Here, $q = \cot\phi$ is the Fano parameter, with ϕ the continuum phase shift and E the spectral energy. The dimensionless variable $\Omega = 2(E - E_0)/\Gamma$ is defined by the resonance energy E_0 and linewidth Γ. Originally introduced for absorption spectra, the model also describes transmission, reflection, and scattering, highlighting its versatility in optical devices. The parameter q governs spectral asymmetry: for $q \approx \pm 1$, strong discrete–continuum coupling produces asymmetric line shapes, while for $|q| \gg 1$ or $q \approx 0$, weaker coupling yields Lorentzian-like profiles. In Figure 1 shows the dependence of the Fano parameter q on the phase shift.

3 Historical overview

Between 2005 and 2015, only a limited number of papers were published in this field, reflecting the exploratory nature of Fano resonance at the time and the early emergence of its application in fiber-based structures. Figure 2 presents a timeline of the annual number of articles on Fano resonance in fiber-optic systems published between 2005 and 2025.

Chiba et al. [14] reported that Fano resonances can be induced in a multimode tapered fiber waveguide coupled with a high-Q microspherical cavity, without requiring any additional elements such as reflectors or delay optics. Steinvurzel et al. [15] demonstrated that the scattering resonances of a single cylinder correspond to Fano resonances, which indeed appear in the loss spectra of solid-core photonic bandgap fibers (PBFs). Totsuka et al. [16] performed Fano resonance to yield coupled-resonator-induced transparency. Vincetti and Setti [17] reported that Fano resonances between hollow-core modes and higher-order dielectric modes occur within the transmission bands. Li et al. [18] observed Fano resonances in a single whispering-gallery microresonator. Furthermore, Vincetti and Setti [19] showed that the additional losses in Kagome fiber designs arise from Fano resonances between core modes and cladding modes exhibiting strong spatial dependence. Chang and Solgaard [20] presented Fano resonances in integrated silicon Bragg reflectors, highlighting their potential for sensing applications. Finally, Yu et al. [21] reported the control of Fano resonances in photonic crystal structures and demonstrated their application in ultrafast optical switching. Yang et al. [22] demonstrated an aerostatically tuned microbubble whispering-gallery resonators to obtain an electro-magnetic-induced transparency and Fano-like lineshapes to sensing application. Over the last decade, Figure 3 highlights a marked increase in the use of resonant waveguide structures in optical fibers, underscoring a strong potential that remains not yet fully explored in this field.

In 2016, Miao et al. [23] reported the evolution of Fano resonance in a thin fiber taper–coupled cylindrical microcavity. Unlike enclosed microcavities, the cylindrical configuration supported both localized whispering gallery modes (WGMs) and delocalized radiation modes. Zhao et al. [24] demonstrated a simple method to achieve Fano-like resonance by coupling a microsphere resonator with a fiber Mach–Zehnder interferometer (FMZI). In their configuration, a tapered microfiber was integrated into one arm of the interferometer to enable evanescent coupling with the microsphere resonator. Gu et al. [25] investigated a fiber loop laser stabilized by Fano resonance in a metallic grating–coupled resonator, demonstrating improved spectral stability. Liao et al. [26] developed a technique to reduce the spectral density of a microbottle resonator (MBR) by depositing UV-curable adhesive droplets near the resonator's central region, effectively degrading the Q-factors of high-order bottle modes and enhancing resonance selectivity.

In 2017, several studies expanded the understanding and applications of Fano resonance in optical systems. Lin et al. [27] proposed a novel microstructured optical fiber (MOF) for multichannel sensing based on Fano resonance among different WGMs propagating through the fiber. The proposed MOF comprised multiple capillary channels of varying diameters arranged within a tubular framework. Miao et al. [28] demonstrated a ring fiber laser structure based on a thin fiber taper–coupled microcylinder system functioning as a transmissive optical filter that selectively allowed narrowband modes to propagate. Li et al. [29] carried out theoretical and numerical investigations of three distinct types of Fano resonances, each arising from different physical mechanisms, in a plasmonic resonator composed of two circular cavities. Zhang et al. [30] proposed a cone-shaped inner-wall coupler to efficiently excite WGMs in a microsphere resonator, while Zhang et al. [31] reported

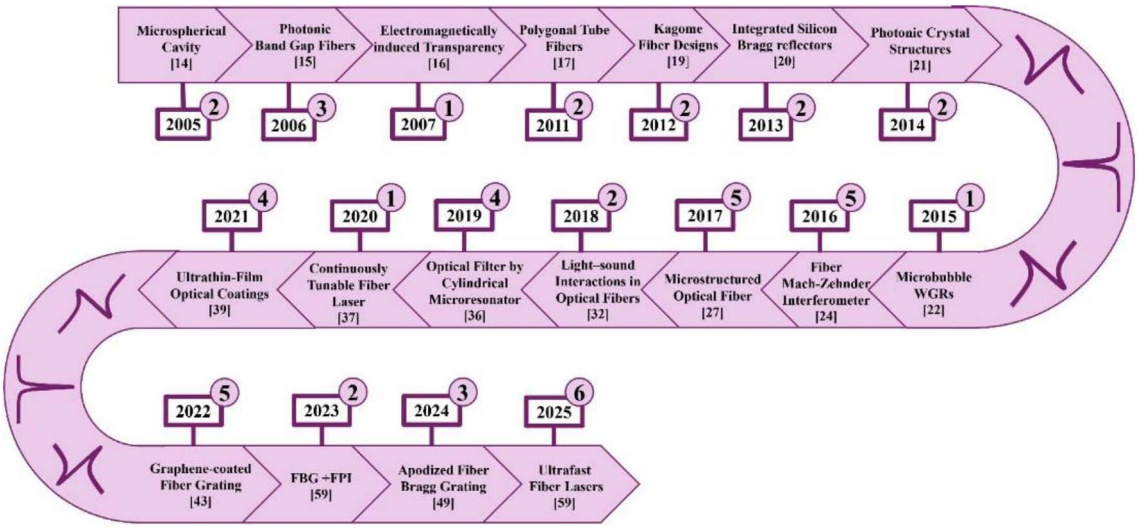


Fig. 2. The timeline on Fano resonance in fiber-optic structures illustrates a significant increase in the annual number of articles published between 2005 and 2025.

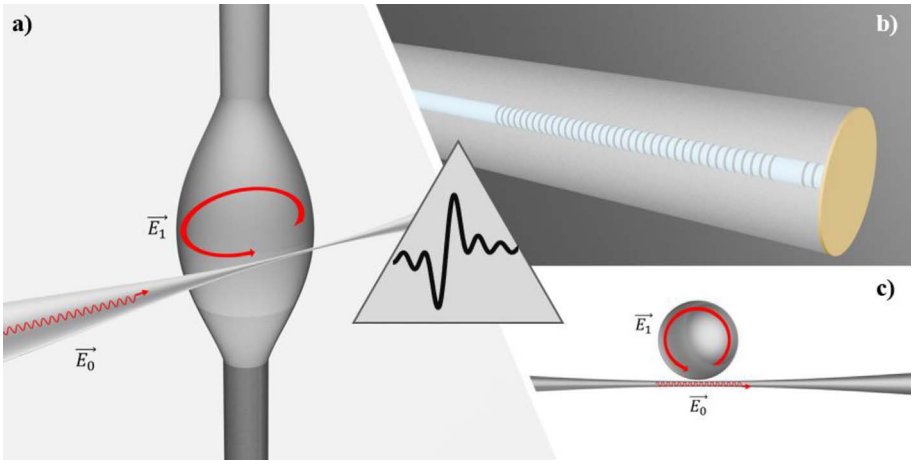


Fig. 3. Optical fiber structures that enable a Fano-like spectrum: a) a microbubble coupled to a taper fiber [22], b) a Bragg grating etalon-based optical fiber [52], c) a microsphere coupled to a taper fiber as a whispering-gallery resonators [14].

enhanced Fano resonance in a non-adiabatic tapered fiber coupled with a microresonator, demonstrating its potential for high-sensitivity optical sensing and laser stabilization.

In 2018, Pant et al. [32] reported wideband excitation of Fano resonances and induced transparency through coherent interactions between multiple Brillouin resonances, revealing a new mechanism for controlling light–sound interactions in optical fibers. In the same year, Zhang et al. [33] proposed an in-line fiber Michelson interferometer designed to enhance the Q-factor of cone-shaped in-wall capillary–coupled resonators, providing improved mode confinement and sharper resonance features.

In 2019, Jiang et al. [34] demonstrated Fano-like resonance in an all-fiber structure, achieving high spectral asymmetry and tunability through precise control of fiber coupling conditions. Lu et al. [35] reported tunable oscillating Fano spectra in a fiber taper–coupled conical microresonator, showing dynamic control of resonance profiles by adjusting the coupling gap and taper diameter. Wang et al. [36] proposed an ultranarrow optical filter based on Fano resonance in a single cylindrical microresonator, enabling single-longitudinal-mode operation in fiber lasers with high spectral purity. Tsai et al. [37] simulated a numerically method to minimize Fresnel reflections with an inscribed FBG at the fiber tip.

In 2020, Li et al. [38] introduced a continuously tunable fiber laser based on a Fano resonance filter using a thin fiber taper–coupled conical microresonator. This configuration allowed stable single-mode lasing and continuous wavelength tuning, demonstrating the potential of Fano-resonant devices for narrow-linewidth laser design and precision optical sensing.

In 2021, Li et al. [39] realized and modulated Fano-like lineshapes in fiber Bragg gratings, providing new insight

into dynamic control of asymmetric spectral responses via refractive index (RI) modulation. ElKabbash et al. [40] investigated Fano-resonant ultrathin-film optical coatings, enabling enhanced light–matter interactions and tunable spectral asymmetry in nanoscale photonic structures. Hu et al. [41] reported a laser-controlled Fano resonance sensing mechanism based on WGM coupling in eccentric-hole fibers integrated with azobenzene materials, achieving optically tunable sensitivity through photochemical modulation. Li et al. [42] developed a fiber Fabry–Perot interferometer-based Fano resonance coupler for whispering-gallery-mode resonators, improving coupling efficiency and enhancing the Q-factor through hybrid cavity interactions.

In 2022, Sun et al. [43] proposed a quasi-3D Fano resonance cavity integrated on the end-facet of an optical fiber for high signal-to-noise ratio "dip-and-read" surface plasmon sensing. This configuration provides a compact and robust sensing platform capable of real-time refractive index monitoring, ideal for biosensing and chemical detection with excellent signal fidelity. Jiang et al. [44] introduced an electrically tunable Fano-like resonance in a graphene-coated fiber grating, enabling fast, reversible modulation through electrical biasing – a step toward reconfigurable photonic systems for sensing and communication. Zhu and Yin [45] investigated Fano-resonance-based optical fiber characteristics for blood glucose detection, leveraging sharp asymmetric spectral features for high RI sensitivity in biomedical sensing. Aman et al. [46] combined the Vernier effect with Fano resonance to realize an ultra-sensitive all-optical sensor, where the Vernier amplification and Fano asymmetry jointly enhanced detection precision. Kudashkin et al. [47] developed a self-stabilized whispering gallery mode resonator in active-core fibers using negative feedback to maintain spectral stability under environmental fluctuations.

In 2023, Gan et al. [48] reported a Fano-like spectrum with quasi-independent slope ratio tuning by coupling higher-order HE modes with WGMs, enabling fine control of spectral asymmetry and linewidth. Finally, Sakhabutdinov et al. [49] proposed a hybrid FBG–FP structure producing composite Fano-type resonances with sharp, tunable spectra, suitable for precision strain, temperature, and pressure measurements.

In 2024, Wang et al. [50] reported an all-optical modulation of Fano-like resonances in an apodized fiber Bragg grating using Er/Yb-doped fiber. The nonlinear gain dynamics of the doped medium enabled effective modulation of the Fano profile without external electronics, demonstrating a new route toward all-optical control and tunable photonic filtering. Chai et al. [51] investigated Fano resonance in a microcylinder–taper coupling system for liquid RI sensing based on the axial separation method. By precisely adjusting the taper–cavity distance, the system achieved strong interference between localized and continuum modes, allowing high-resolution detection of RI variations in liquids. La et al. [52] demonstrated a Bragg etalon-based optical fiber for optoacoustic detection. Robalinho et al. [53] introduced a phase-shifted FBG (PS-FBG), which was fabricated using selective pitch slicing. This enabled tuning of the Fano-like asymmetry spectrum, resulting in a wide-range vibrational sensor. Feng et al. [54] proposed a simple scheme to generate multiple ultra-high-slope Fano-like resonances via Mach–Zehnder interferometers (MZIs) cascaded with fiber gratings. The configuration achieved enhanced spectral discrimination and steep transmission slopes, suitable for high-resolution spectral sensing and wavelength filtering.

In 2025, Wu et al. [55] presented a method for the synchronous and controllable realization of Lorentzian, Fano, and electromagnetically induced transparency (EIT) lineshapes in an all-fiber configuration. This work demonstrated full optical tunability between distinct resonance profiles within a single platform, significantly expanding the flexibility of all-fiber photonic filters and sensors. Li et al. [56] explored Fano resonance in a whispering gallery mode microsphere resonator coupled with a femtosecond laser-drilled tapered fiber. The laser-drilled taper provided enhanced coupling precision and mechanical stability, resulting in sharper resonance lines and improved sensing repeatability. Wang et al. [57] developed a modulation-free laser frequency stabilization technique based on balanced detection of a common-path Fano resonance in an all–polarization-maintaining (PM) fiber cavity. This approach provided a compact and vibration-insensitive means of laser locking, eliminating the need for active modulation or complex feedback electronics. Li et al. [58] proposed a mechanism to enhance the dynamic range in microcavity sensing through nonlinear harmonic generation of Fano resonances. The introduction of higher-order nonlinear interactions broadened the measurable range while preserving spectral sensitivity, enabling versatile performance in high-intensity or nonlinear photonic environments. Finally, Zou et al. [59] investigated the resonantly driven nonlinear dynamics of soliton molecules in ultrafast fiber lasers. The study revealed that soliton molecule formation and evolution could be strongly influenced by Fano-like interference effects within the laser cavity, providing new insights into the control of ultrafast nonlinear dynamics and pulse shaping in fiber lasers. Figure 3 summarizes the principal fiber-optic configurations employed with this technique.

4 Fano resonance on fiber structures

As outlined in the historical overview, various structures were designed to induce Fano resonance through modal coupling within the fiber. This section will address these architectures and their operational principles. The three most common approaches are whispering gallery modes, microstructured fibers and interferometric systems. Recently, new configurations using FBG structures were published to excite the discrete mode.

4.1 Whispering gallery mode microcavities

Whispering gallery modes are waves capable of propagating around a concave surface and were first presented as acoustic waves in 1910 [60]. In 1961, an electromagnetic analog was also demonstrated [61]. The microcavities used for WGM present advantages such as a high Q-factor and a comparatively reduced mode volume [62]. When coupled

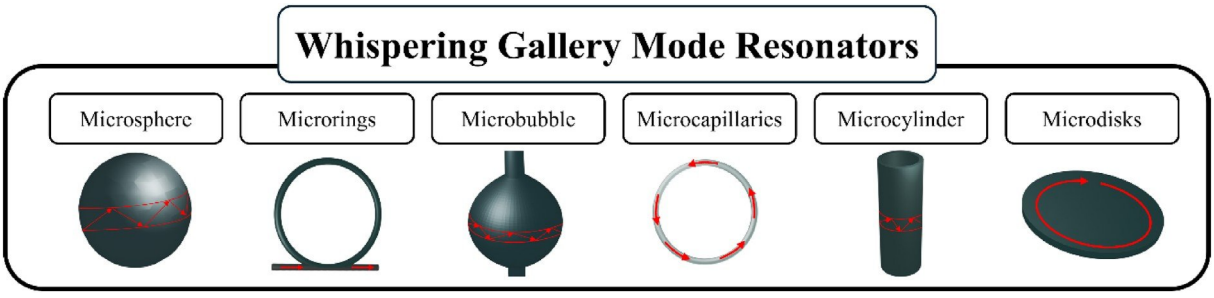


Fig. 4. Overview of WGM resonators used for induced Fano resonance coupled into optical fibers.

with an optical fiber, considered to have a low Q-factor and to propagate a continuum mode, the capacity to induce Fano resonance is enhanced due to its high Q-factor, which represents the discrete mode [63]. Two mathematical methodologies have been developed for modeling the modal interference between both. The classical transfer matrix method analyzes the exchange of optical power between the resonator and the waveguide [64]. The coupled mode theory, proposed by Pierce at Bell Laboratories, accounts for different fields that perturb the system to calculate the spectrum [65].

In literature, a wide range of structures have been developed to propose a WGM. These include microspheres [66], microrings [67], microbubble [68], microdisks [69], microcapillaries [70], microcylinder [71], microtoroid [72] and others. The coupled waveguide exhibits a low Q-factor and operates within a limited frequency range, thereby inducing the Fano phenomenon. The structure can be obtained through various methods, which can be classified into two distinct groups. Single-cavity coupling, defined as a coupling configuration in which only one cavity is involved, can be achieved through unilateral coupling, a process in which a resonator is coupled to the side of a single waveguide. For example, a WGM microcavity coupled to a tapered fiber [73] exhibits the advantages of simple fabrication, high coupling efficiency, and elevated Q-value in the order of 10^6 [74] and a maximum obtained of 10^8 [75]. However, it is considered a fragile system. Alternatively, bilateral coupling typically involves positioning the resonator between two waveguides [76]. Another approach is the in-fiber coupling structure, which enhances system robustness by embedding the resonator within the fiber waveguide [77, 78]. However, the Q-factor presents lower values compared to microresonators coupled to tapered fibers, being limited to 10^4 [33]. This demands further investigations of this method in order to fully exploit the potential of WGM resonators and the robustness of optical fibers. An overview of the resonator's structures mentioned is illustrated in Figure 4.

Another common approach is multi-cavity coupling, which comprises the same coupling types showed above: unilateral, bilateral, and in-fiber; however, it involves more than one microcavity interacting through resonance [79–83]. In comparison, multi-cavity coupling more readily excites resonances between a low-Q mode and a high-Q mode due to the resonant interactions among the microcavities coupled to the fiber. Figure 5 presents typical configurations of these resonators with an optical fiber, which can be unilateral, bilateral, or in-fiber.

4.2 Microstructured optical fibers

Microstructured optical fibers are structures that differ from the classic single-mode fiber, in which different structures are developed within the fiber to guide light along the direction of propagation. Typically, instead of a central core with a higher refractive index, air holes are introduced within the fiber. Geometry and spatial organization are determinant for the confined and propagated modes [84]. The Fano resonance is induced when the mode confinement geometry is designed to enable interaction between a high-Q mode and a low-Q mode.

An investigation was conducted on how the phase relationship between each hollow-core cylinder in a photonic bandgap fiber induces Fano-like resonances in the scattering spectrum [15]. In that study, the number of cylinders and their spacing were identified as critical parameters in coupling terms and Fano appearance. Another confinement mechanism is the Inhibited Coupling (IC), which provide a wider transmission bandwidth than PBG [85]. Distinct MOFs were designed to achieve IC, and this mechanism has been used to obtain Fano resonance in Kagome fibers [86] and polygonal tube fibers [17, 19].

4.3 Interferometric strutures

Interferometric structures have been widely used in photonics for a range of applications. In the context of Fano resonance, this phenomenon has been observed in a Mach-Zender interferometer (MZI), and a theorical model was developed by Miroshinichenko et al. [87]. Several applications have been demonstrated using such configurations, including sensing with a tapered optical fiber [88], an in-line MZI incorporating an FBG inscribed in a D-shaped fiber [34], and microrings exhibiting a discrete comb-like spectrum coupled to arm of an MZI [55]. MZI-based Fano resonances have also been explored for optical bistability [89], and for tunable control of the Fano asymmetry using FBGs [44, 50, 54, 90], demonstrating a versatile approach to inducing Fano resonance in all-fiber systems.

Michelson interferometer configurations have been demonstrated in an in-fiber integrated setup with a microsphere WGM for sensing and lasing applications [33], as well as in a ring-cavity-coupled scheme for optical bistabil-

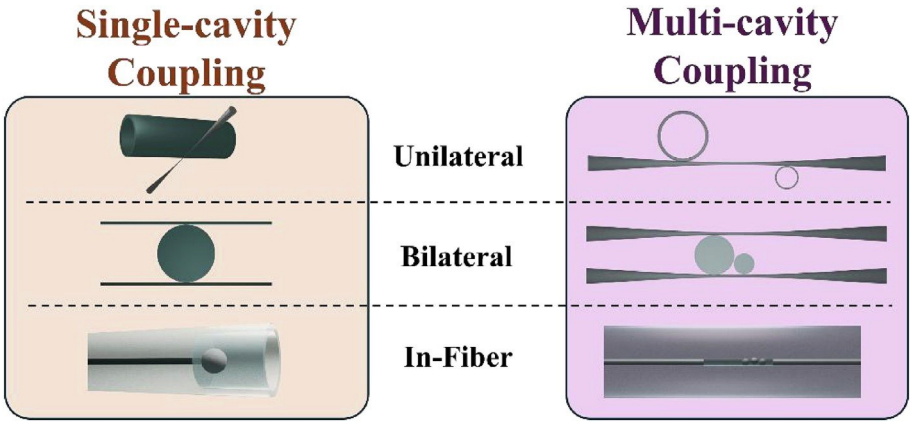


Fig. 5. Typical configurations of WGM resonators coupled to optical fibers to excite Fano resonance.

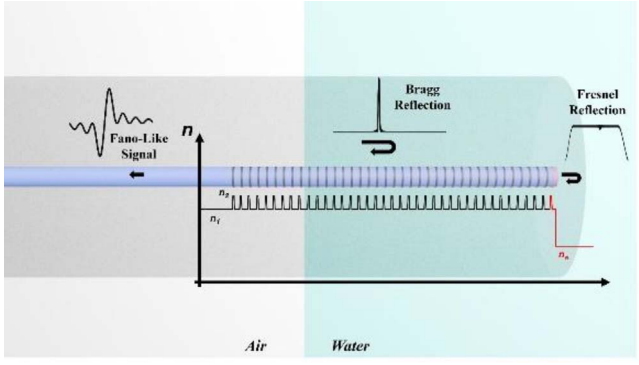


Fig. 6. Dynamic behavior of the Fano-FBG structure, where the interference between Fresnel and Bragg reflections results in the emergence of a Fano-like reflected spectrum.

ity [91]. Another approach relies on Fabry–Pérot interferometers (FPIs) coupled to a WGM microsphere [42], or combined with FBGs [49, 92, 93], where the interaction between the two elements induces a Fano-like resonance. FPI-based schemes have also been enhanced by integrating a gold plasmonic metasurface on the facet [94]. In contrast, Sagnac interferometers are less commonly used to induce Fano resonance in optical fibers, being more frequently reported in silicon waveguides [95].

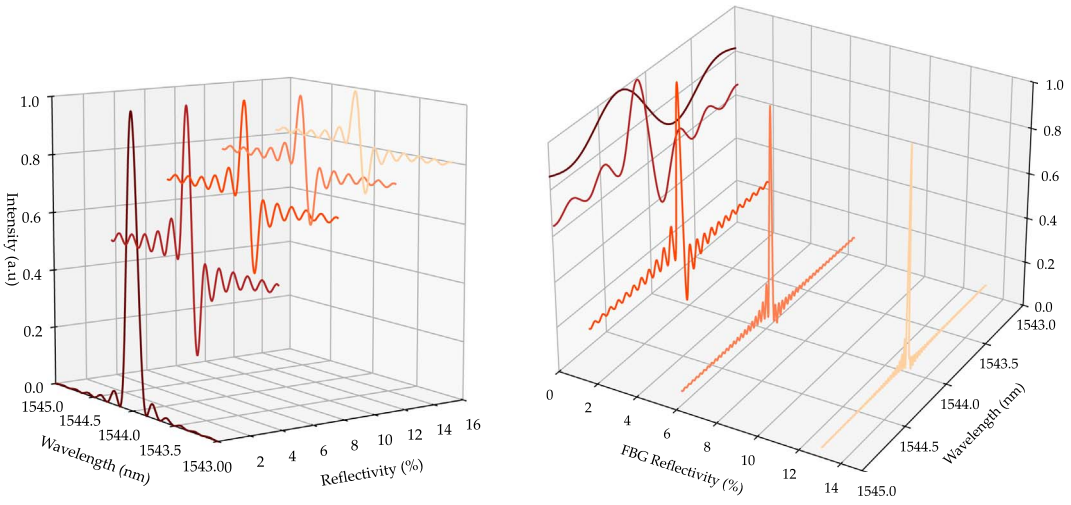
4.4 Bragg gratings structures

This review briefly addresses one of the most recent works in optical fiber sensors, published in 2024, which proposes a Fano-like FBG structure, as illustrated in Figure 6 [53]. This FBG structure is a phase-shift grating operated in reflection, in which the last period is modified to introduce the phase shift; that is, its periodicity is shorter than the nominal FBG period. This study investigated a Fano-FBG configuration in which the reflectivity at the fiber end was kept fixed while the FBG reflectivity was varied. The inverse case was then analyzed: the FBG reflectivity was held constant, and the fiber-tip reflectivity was varied externally. The objective was to identify the optimal parameters for obtaining a well-defined Fano spectrum on an optical spectrum analyzer (OSA). Based on these simulations, the structure's performance as a refractive-index sensor was demonstrated both theoretically and experimentally, highlighting its advantages as a sensing element and showcasing the potential of this new configuration to the scientific community.

4.4.1 An analytical approach to modeling the Fano-like FBG

The interferometric phenomenon that yields the FBG spectrum can be described using coupled-mode theory, as presented by Sakhabutdinov et al. [49]. Another possibility is to study the interaction between the incident electric field amplitude E and the grating structure. In this approach, the model is based on summing the contributions of the waves reflected at each refractive-index perturbation introduced by the grating modulation. The grating period between the successive reflective points Λ introduces a phase shift of $\varphi = \beta(2\Lambda)$, where $\beta = 2\pi n_{\text{eff}}/\lambda_0$ the propagation coefficient, n_{eff} the effective refractive index, and λ_0 the vacuum wavelength. To model the superposition of reflections, N reflective points are defined to simulate the FBG, each characterized by transmission and reflection coefficients of $t_\Lambda = 2n_1n_2/n_1 + n_2$ and $r_\Lambda = (n_1 - n_2)/(n_1 + n_2)$, respectively, where n_1 is the optical fiber core refractive index and n_2 is the grating refractive index.

The Fano-like resonance arises from interference between a discrete spectrum and a continuum spectrum. In this framework, a spacing is introduced between the grating's last pitch and the fiber tip, with a length ranging from 0 and Λ. The continuum mode originates itself from the Fresnel reflection at the fiber tip across this final spacing,, with reflection coefficient of $R_\Phi = [(n_{\text{eff}} - n_e)/(n_{\text{eff}} + n_e)]^2$, where n_e is the environment RI. The phase shift is determined by this specific length Φ, which in turn controls the tuning of the Fano asymmetry parameter q. Equations (2) and (3) provide the mathematical model obtained by summing the reflected electromagnetic waves, that reproduces the Fano-like asymmetrical spectrum generated by the Fano-FBG. In this numerical model, the FBG response is expressed as a geometric series, with the



(a) (b)

Fig. 7. Simulation of modal coupling efficiency: (a) influence of Fresnel reflectivity and (b) influence of the FBG reflected peak intensity on the resulting Fano profile.

Fresnel reflection at the fiber tip determined by the last term. The Fano parameter is controlled by the ratio between the length of the last pitch, Φ, and the Bragg period, Λ, of the FBG. For $\Phi \leq \Lambda/2$, the asymmetric line shapes correspond to Fano parameters between -1 and 0. When $\Phi = \Lambda/2$, the Fano parameter equals 0, whereas for $\Phi \geq \Lambda/2$, the Fano parameter lies between 0 and 1.

$$E_r = E\left[r_\Lambda \sum_{k=0}^{N} \left(t_\Lambda e^{i\beta\Lambda}\right)^{2k} + R_\Phi {t_\Lambda}^{2N+2} e^{i\beta(2\Phi+2N\Lambda)}\right], \quad (2)$$

$$E_r = E\left[\frac{\mathrm{r}_\Lambda\left(1-\left(t_\Lambda e^{i\beta\Lambda}\right)^{2N+2}\right)}{\left(1-\left(t_\Lambda e^{i\beta\Lambda}\right)^{2}\right)} + R_\Phi {t_\Lambda}^{2N+2} e^{i\beta(2\Phi+2N\Lambda)}\right]. \quad (3)$$

Based on the analytical model, the coupling efficiency of the Fano-FBG is investigated as a function of both the grating reflectivity and the Fresnel reflection coefficient at the fiber tip. In this analysis, the grating length is kept constant at 1 mm. In Figure 7a, the parameter R_Φ (0–3%) quantifies the strength of the secondary reflection that couples with the grating-reflected field. As R_Φ increases, the response evolves from a conventional, symmetric FBG spectrum (0%) to a clearly asymmetric Fano line shape (3%), evidencing progressively stronger interference. Figure 7b considers the Fresnel reflection associated with the silica–air interface at the fiber end, while the FBG reflectivity is swept from 0.1% to 13%. Increasing grating reflectivity reduces the relative contribution of the tip-reflected field, weakening the interference term; consequently, the grating-reflected component dominates and the Fano features progressively vanish. Conversely, higher tip reflectivity enhances the counter-propagating field, increasing the interference contrast and sharpening the Fano spectral asymmetry. Moreover, large grating reflectivity compresses the spectral bandwidth and drives the response toward the characteristic FBG peak, effectively masking the Fano profile. Therefore, achieving efficient modal coupling and inducing the Fano resonance requires a suitable trade-off between grating and tip reflectivities, ensuring comparable amplitudes of the two interfering reflected signals. Another relevant aspect is the tunability of the Fano asymmetric q - parameter via the phase shift introduced in the FBG. By adjusting the length of the final grating period, one can realize different interference regimes, yielding $q = \pm 1$ and $q = 0$. Notably, the case $q = 0$ corresponds to a symmetric FBG notch response or anti-reflection AR-FBG, as reported in [37]. For $q = \pm 1$, a representative example is provided by the 2024 study on the slice FBG approach [53], where this phase-shift produces a clearly asymmetric Fano profile.

4.4.2 Fano-like FBG applied as a sensing tool

In addition to the simulation work performed for the optical device, an experimental study was also carried out to evaluate its performance as a sensor for monitoring refractive-index variations in liquids. Different water–sugar solutions were prepared to obtain refractive indices ranging from 1.334 to 1.339. The experiments were performed using a Fano-like FBG with $q = -1$ (see Fig. 8a). The interrogation system was intensity-based, and self-referenced by measuring the intensity difference between the peak and the valley of the Fano profile; the experimental and numerical results are illustrated in Figure 8b. This approach was chosen because the Fresnel reflectivity strongly influences the signal intensity, while the peak wavelength exhibits low sensitivity to refractive-index variations. The experimental setup comprised broadband source with a spectral bandwidth of 100 nm centered at 1550 nm, a three-port optical fiber circulator (Thorlabs 6015-3) to route the reflected

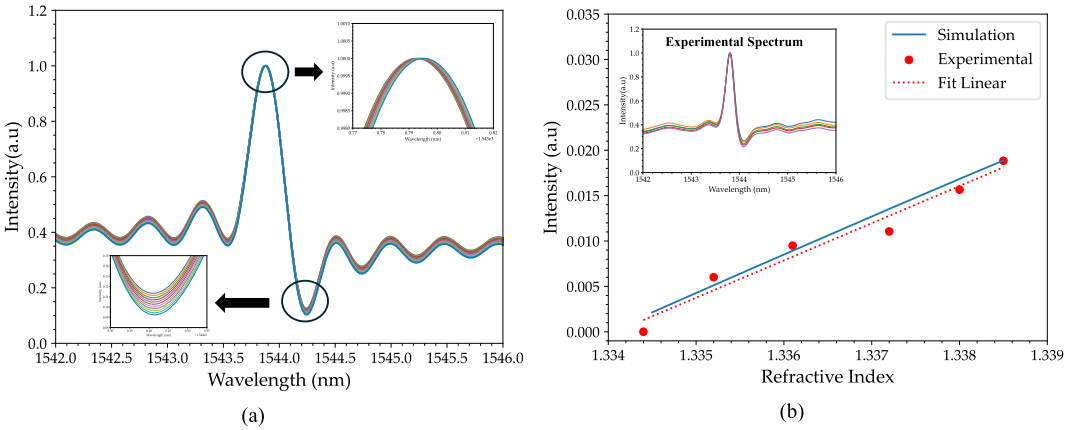


Fig. 8. Evaluation of the Fano-FBG, with $q = -1$, as an optical fiber sensor: a) simulated spectrum and b) comparison between the simulation predictions and the experimental results.

Table 1. RI sensors based on FBGs.

	Detection Limit (RIU)	Ref
FBG and etch-eroded fiber Fabry-Pérot interferometer	1.4×10^{-5}	[96]
Thinned fiber Bragg gratings	10^{-4}	[97]
Dual Fabry-Perot FBG	10^{-5}	[98]
End-face of a glass fiber connected to FBG	10^{-3}	[99]
In this Work (Fano-FBG)	6×10^{-4}	

signal, and an Optical Spectrum Analyzer (YOKOGAWA AQ6370C, Osaka, Japan) for spectral interrogation. The fiber-tip cleaving process was achieved using a polishing machine, with the support of the OSA, the fiber tip can be polished until the desire asymmetry is obtained (Table 1).

The validation of the mathematical model was performed by comparing the intensity variation within the defined refractive-index range. The simulations resulted in an FBG with 1% reflectivity, $\Delta n = 2.6 \times 10^{-4}$, and $N = 8452$. Experimentally, a linear sensitivity of 4.1 ± 0.4 RIU^{-1} was obtained ($r^2 = 0.980$), while the simulation yielded a linear sensitivity of 4.18 ± 0.01 RIU^{-1} ($r^2 = 0.99997$). The experimental results also indicated a detection limit of 6×10^{-4} RIU. A deviation of approximately 1% was observed between the real sensor performance and the simulation predictions, as demonstrated in Figure 8b. This demonstration shows the potential of a Fano-like FBG as a sensing element, both for wavelength-based measurements as temperature measurement and for self-referenced intensity measurements. In this case, it is possible to use two lasers with wavelengths set at each peak and apply filtering to separate the wavelength contributions. Thus, by using two photodetectors, we obtain a low-cost system. A comparison with other RI sensors reveals that this system provides measurement capabilities comparable to other intensity-based interrogation FBG systems. Also, it offers a simpler implementation for achieving higher sensitivity than many existing approaches, highlighting its potential as an effective optical sensing device.

5 Conclusion

The chronological evolution of studies on Fano resonance in optical fiber systems – from its early conceptual demonstrations in the mid-2000s to the highly integrated and tunable devices developed after 2020 – illustrates a clear trajectory toward functional maturity and application-oriented design. Early works focused on demonstrating the physical existence of Fano interference in guided-wave structures and microresonators, whereas more recent contributions have emphasized engineering control, spectral reconfigurability, and hybrid material integration. The demonstration of the Fano-like FBG highlights the potential of this structure as a simple device capable of exciting a Fano-like resonance. When demonstrated as a RI sensor with a resolution of 6×10^{-4} RIU, the sensitivity is not remarkable; however, the self-referencing method is useful for reducing the influence of external noise on the measurement. Looking to the future, research is expected to converge on several key directions. First, the integration of intelligent control mechanisms, such as thermo-optic, electro-optic, and opto-mechanical tuning, will likely enable adaptive Fano resonance devices capable of self-calibration and real-time environmental compensation. These developments may render Fano-based sensors and filters not only highly sensitive but also self-stabilizing and field-deployable for operation in harsh or dynamic environments. The ongoing miniaturization trend in photonic integration suggests that Fano-resonant fiber systems may evolve toward chip-scale architectures, maintaining the low-loss advantages of

optical fibers while achieving full compatibility with silicon photonics and planar integrated circuits. This evolution could lead to compact, mass-producible Fano-based devices directly integrated into telecommunication, biomedical, and quantum systems. In summary, the coming decade is poised to witness the emergence of adaptive, multifunctional, and miniaturized Fano-resonant fiber systems, bridging the gap between fundamental photonics and real-world applications.

Acknowledgments

This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the support UID/50014/2025 (https://doi.org/10.54499/UID/50014/2025. This work was also carried out under Project No. 15022 (ACLART – Prediction of clinical outcomes of anterior cruciate ligament reconstruction through in silico simulation combined with artificial intelligence), reference COMPETE2023-FEDER-00867200. The project was co-financed by the European Regional Development Fund (ERDF) under the PORTUGAL2030 program, and by the budget of the Foundation for Science and Technology, I.P. (FCT).

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

This article has no associated data generated and/or analyzed / Data associated with this article cannot be disclosed due to legal/ethical/other reason.

Author contribution statement

We expect that all authors will have reviewed, discussed, and agreed to their individual contributions ahead of this time. Contributions will be published before the references section, and they should accurately reflect contributions to the work. The Conceptualization, O. Frazão and V. Piaia; Software, V. Piaia. and P. Robalinho; Validation, S. Silva and O. Frazão.; Writing – Original Draft Preparation, O. Frazão and V. Piaia; Writing – Review & Editing, P. Robalinho and S. Silva.; Supervision, O. Frazão.

References

1 Cao G, Dong S, Zhou L-M, Zhang Q, Deng Y, Wang C, Zhang H, Chen Y, Qiu C-W, Liu X, Fano resonance in artificial photonic molecules, *Adv. Opt. Mater.* **8**, 1902153 (2020). https://doi.org/10.1002/adom.201902153.

2 Rasmussen TS, Yu Y, Mork J, Theory of self-pulsing in photonic crystal Fano lasers, *Laser Photonics Rev.* **11**, 1700089 (2017). https://doi.org/10.1002/lpor.201700089.

3 Lim WX, Manjappa M, Pitchappa P, Singh R, Shaping high-Q planar Fano resonant metamaterials toward futuristic technologies, *Adv. Opt. Mater.* **6**, 1800502 (2018). https://doi.org/10.1002/adom.201800502.

4 Sulabh, Singh L, Jain S, Kumar M, Nanophotonic device based on Fano resonance in engineered slot waveguide for optical detection of viral infections, *IEEE Sens. J.* **21**, 2805–2812 (2021). https://doi.org/10.1109/JSEN.2020.3023146.

5 Bekele D, Yu Y, Yvind K, Mork J, In-plane photonic crystal devices using Fano resonances, *Laser Photonics Rev.* **13**, 1900054 (2019). https://doi.org/10.1002/lpor.201900054.

6 Abujetas DR, Mandujano MAG, Méndez ER, Sánchez-Gil JA, High-contrast Fano resonances in single semiconductor nanorods, *ACS Photonics* **4**, 1814–1821 (2017). https://doi.org/10.1021/acsphotonics.7b00395.

7 Chen J, Gan F, Wang Y, Li G, Plasmonic sensing and modulation based on Fano resonances, *Adv. Opt. Mater.* **6**, 1701152 (2018). https://doi.org/10.1002/adom.201701152.

8 Dong G, Wang Y, Zhang X (2018) High-contrast and low-power all-optical switch using Fano resonance based on a silicon nanobeam cavity, *Opt. Lett.* **43**, 5977–5980. https://doi.org/10.1364/OL.43.005977.

9 Yu Y, Hu H, Oxenløwe LK, Yvind K, Mork J, Ultrafast all-optical modulation using a photonic-crystal Fano structure with broken symmetry, *Opt. Lett.* **40**, 2357–2360 (2015). https://doi.org/10.1364/OL.40.002357.

10 Solodovchenko N, Sidorenko M, Seidov T, Popov I, Nenasheva E, Samusev K, Limonov M, Cascades of Fano resonances in light scattering by dielectric particles, *Mater. Today* **60**, 69–78 (2022). https://doi.org/10.1016/j.mattod.2022.09.007.

11 Zangeneh-Nejad F, Fleury R, Topological Fano resonances, *Phys. Rev. Lett.* **122**, 014301 (2019). https://doi.org/10.1103/PhysRevLett.122.014301.

12 Fano U, Effects of configuration interaction on intensities and phase shifts, *Phys. Rev.* **124**, 1866–1878 (1961). https://doi.org/10.1103/PhysRev.124.1866.

13 Limonov MF, Rybin MV, Poddubny AN, Kivshar YS, Fano resonances in photonics, *Nature Photon.* **11**, 543–554 (2017). https://doi.org/10.1038/nphoton.2017.142.

14 Chiba A, Fujiwara H, Hotta J, Takeuchi S, Sasaki K, Fano resonance in a multimode tapered fiber coupled with a microspherical cavity, *Appl. Phys. Lett.* **86**, 261106 (2005). https://doi.org/10.1063/1.1951049.

15 Steinvurzel P, de Sterke CM, Steel MJ, Kuhlmey BT, Eggleton BJ, Single scatterer Fano resonances in solid core photonic band gap fibers, *Opt. Express* **14**, 8797–8811 (2006). https://doi.org/10.1364/OE.14.008797.

16 Totsuka K, Kobayashi N, Tomita M, Slow light in coupled-resonator-induced transparency, *Phys. Rev. Lett.* **98**, 213904 (2007). https://doi.org/10.1103/PhysRevLett.98.213904.

17 Vincetti L, Setti V, Fano resonances in polygonal tube fibers, *J. Lightwave Technol.* **30**, 31–37 (2012). https://doi.org/10.1109/JLT.2011.2177245.

18 Li B-B, Xiao Y-F, Zou C-L, Liu Y-C, Jiang X-F, Chen Y-L, Li Y, Gong Q, Experimental observation of Fano resonance in a single whispering-gallery microresonator, *Appl. Phys. Lett.* **98**, 021116 (2011). https://doi.org/10.1063/1.3541884.

19 Vincetti L, Setti V, Extra loss due to Fano resonances in inhibited coupling fibers based on a lattice of tubes, *Opt. Express* **20**, 14350–14361 (2012). https://doi.org/10.1364/OE.20.014350.

20 Chang C-M, Solgaard O, Fano resonances in integrated silicon Bragg reflectors for sensing applications, *Opt. Express* **21**, 27209–27218 (2013). https://doi.org/10.1364/OE.21. 027209.

21 Yu Y, Heuck M, Hu H, Xue W, Peucheret C, Chen Y, Oxenløwe LK, Yvind K, Mørk J, Fano resonance control in a photonic crystal structure and its application to ultrafast switching, *Appl. Phys. Lett.* **105**, 061117 (2014). https://doi.org/10.1063/1.4893451.

22 Yang Y, Saurabh S, Ward J, Chormaic SN, Coupled-mode-induced transparency in aerostatically tuned microbubble whispering-gallery resonators, *Opt. Lett.* **40**, 1834–1837 (2015). https://doi.org/10.1364/OL.40.001834.

23 Miao Y, Peng Y, Xiang Y, Li M, Lu Y, Song Y, Dynamic Fano resonance in thin fiber taper coupled cylindrical microcavity, *IEEE Photonics J.* **8**, 1–6 (2016). https://doi.org/10.1109/JPHOT.2016.2630316.

24 Zhao C, Gan X, Fang L, Han L, Chang K, Li D, Zhao J, Tunable Fano-like resonance enabled by coupling a microsphere with a fiber Mach–Zehnder interferometer, *Appl. Opt.* **55**, 5756–5760 (2016). https://doi.org/10.1364/AO.55.005756.

25 Gu B, Zhou Y, Yu X, Luan F, Fiber loop laser stabilized by Fano resonance in metallic grating coupled resonator, *IEEE Photonics Technol. Lett.* **28**, 1597–1600 (2016). https://doi.org/10.1109/LPT.2016.2560239.

26 Liao J, Wu X, Liu L, Xu L, Fano resonance and improved sensing performance in a spectral-simplified optofluidic micro-bubble resonator by introducing selective modal losses, *Opt. Express* **24**, 8574–8580 (2016). https://doi.org/10.1364/OE.24.008574.

27 Lin W, Zhang H, Chen S-C, Liu B, Liu Y-G, Microstructured optical fiber for multichannel sensing based on Fano resonance of the whispering gallery modes, *Opt. Express* **25**, 994–1004 (2017). https://doi.org/10.1364/OE.25.000994.

28 Miao Y, Li M, Song Y (2017) Ring fiber lasers based on EIT-like Fano resonances as a wavelength-selective element, *IEEE Photonics Technol. Lett.* **29**, 1900–1902. https://doi.org/10.1109/LPT.2017.2755696.

29 Li S, Wang Y, Jiao R, Wang L, Duan G, Yu L, Fano resonances based on multimode and degenerate mode interference in plasmonic resonator system, *Opt. Express* **25**, 3525–3533. https://doi.org/10.1364/OE.25.003525.

30 Zhang X, Yang Y, Shao H, Bai H, Pang F, Xiao H, Wang T, Fano resonances in cone-shaped inwall capillary based microsphere resonator, *Opt. Express* **25**, 615–621 (2017). https://doi.org/10.1364/OE.25.000615.

31 Zhang K, Wang Y, Wu Y-H, Enhanced Fano resonance in a non-adiabatic tapered fiber coupled with a microresonator, *Opt. Lett.* **42**, 2956–2959 (2017). https://doi.org/10.1364/OL.42.002956.

32 Pant R, Siva Shakthi A, Yelikar AB, Wideband excitation of Fano resonances and induced transparency by coherent interactions between Brillouin resonances, *Sci. Rep.* **8**, 9175 (2018). https://doi.org/10.1038/s41598-018-27444-8.

33 Zhang X, Bai H, Pan H, Wang J, Yan M, Xiao H, Wang T, In-line fiber michelson interferometer for enhancing the Q factor of cone-shaped inwall capillary coupled resonators, *IEEE Photonics J.* **10**, 1–8 (2018). https://doi.org/10.1109/JPHOT.2018.2818715.

34 Jiang B, Gan X, Gu L, Hao Z, Wang S, Bi Z, Zhang L, Zhou K, Zhao J, Fano-like resonance in an all-in-fiber structure, *IEEE Photonics J.* **11**, 1–7 (2019). https://doi.org/10.1109/JPHOT.2019.2936485.

35 Lu Y, Zhu X, Li J, Nie Y, Li M, Song Y, Tunable oscillating Fano spectra in a fiber taper coupled conical microresonator, *IEEE Photonics J.* **11**, 1–7 (2019). https://doi.org/10.1109/JPHOT.2019.2924257.

36 Wang H, Xu R, Zhang J, Zhou W, Shen D, Ultranarrow filter based on Fano resonance in a single cylindrical microresonator for single-longitudinal-mode fiber lasers, *Opt. Express* **27**, 22717–22726 (2019). https://doi.org/10.1364/OE.27.022717.

37 Tsai T-Y, Lee Z-C, Tsao H-X, Lin S-T, Minimization of Fresnel reflection by anti-reflection fiber Bragg grating inscribed at the fiber ends, *Opt. Express* **27**, 11510–11515 (2019). https://doi.org/10.1364/OE.27.011510.

38 Li J, Lu Y, Cao L, Li M, Song Y, Continuously tunable fiber laser based on Fano resonance filter of thin-fiber-taper-coupled conical microresonator, *Opt. Commun.* **466**, 125629 (2020). https://doi.org/10.1016/j.optcom.2020.125629.

39 Li A, Jiang B, Zhang P, Gan X, Hao Z, Hou Y, Zhang J, Li P, Zhao J, Realization and modulation of Fano-like lineshape in fiber Bragg gratings, *J. Lightwave Technol.* **39**, 4419–4423 (2021). https://opg.optica.org/jlt/abstract.cfm?uri=jlt-39-13-4419.

40 ElKabbash M, Letsou T, Jalil SA, Hoffman N, Zhang J, Rutledge J, Lininger AR, Fann C-H, Hinczewski M, Strangi G, Guo C, Fano-resonant ultrathin film optical coatings, *Nat. Nanotechnol.* **16**, 440–446 (2021). https://doi.org/10.1038/s41565-020-00841-9.

41 Hu X, Zhang H, Lin W, Wang Y, Liu B, Laser-controlled Fano resonance sensing based on WGM coupling in eccentric hole fibers integrated with azobenzene, *J. Lightwave Technol.* **39**, 320–327 (2021). https://opg.optica.org/jlt/abstract.cfm?uri=jlt-39-1-320.

42 Li M, Wang R, Fiber Fabry-Perot interferometer-based Fano resonance coupler for whispering-gallery-mode resonators, *IEEE Photonics Technol. Lett.* **33**, 585–588(2021). https://doi.org/10.1109/LPT.2021.3076999.

43 Sun X, Lei Z, Zhong H, He C, Liu S, Meng Q, Liu Q, Chen S, Kong X, Yang T, A quasi-3D Fano resonance cavity on optical fiber end-facet for high signal-to-noise ratio dip-and-read surface plasmon sensing, *Light Adv. Manuf.* **3**, 665–675 (2022). https://doi.org/10.37188/lam.2022.046.

44 Jiang B, Zhang X, Li A, Hou Y, Hao Z, Gan X, Zhao J, Electrically induced dynamic Fano-like resonance in a graphene-coated fiber grating, *Photon. Res.* **10**, 1238–1243 (2022). https://doi.org/10.1364/PRJ.453762.

45 Zhu J, Yin J, Optical-fibre characteristics based on Fano resonances and sensor application in blood glucose detection, *Opt. Express* **30**, 26749–26760 (2022). https://doi.org/10.1364/OE.463427.

46 Hairol Aman MA, Ahmad Noorden AF, Ahmad Fajri FA, Abdul Kadir MZ, Bahari I, Danial WH, Daud S, Bahadoran M, Integrating Vernier spectrum with Fano resonance for high sensitivity of an all-optical sensor, *J. Comput. Electron.* **22**, 276–287 (2023). https://doi.org/10.1007/s10825-022-01946-1.

47 Kudashkin DV, Khorev SV, Vatnik ID. Self-Stabilization by negative feedback of whispering gallery mode resonators based on optical fiber with active core. *J. Lightwave Technol.* **40**, 7351–7357 (2022). https://doi.org/10.1109/JLT.2022.3201655.

48 Gan X, Ban Z, Gao F, Bo F, Zhang G, Xu J, Fano-like spectrum with a quasi-independent tuning of slope ratio based on HE mode components coupled whispering gallery mode, *J. Lightwave Technol.* **41**, 2501–2505 (2023). https://doi.org/10.1109/JLT.2022.3231243.

49 Sakhabutdinov AZh, Agliullin TA, Hussein SMRH, Kuznetsov AA, Anfinogentov VI, Valeev BI, Fano-type resonance structures based on combination of fiber Bragg grating with Fabry-Perot interferometer, *Karbala International Journal of Modern Science* **9** (2023). https://doi.org/10.33640/2405-609X.3279.

50 Wang F, Lin S, Qu Y, Han X, Wei Y, Zhao Z, Zhang Y, All-optical modulation of Fano-like resonances in apodized fiber Bragg grating using Er/Yb doped fiber, *J. Phys. D: Appl.*

Phys. **57**, 135104 (2024). https://doi.org/10.1088/1361-6463/ad194e.
51 Chai M, Liu C, Wang M, Xue C, Xie C, Xu W, Shi J, Wu T, He X, Fano resonance in a microcylinder-taper coupling system for liquid refractive index sensing via axial separation method, *Opt. Lasers Eng.* **179**, 108253 (2024). https://doi.org/10.1016/j.optlaseng.2024.108253.
52 La TA, Ülgen O, Shnaiderman R, Ntziachristos V, Bragg grating etalon-based optical fiber for ultrasound and optoacoustic detection, *Nat. Commun.* **15**, 7521 (2024). https://doi.org/10.1038/s41467-024-51497-1.
53 Robalinho P, Piaia V, Soares L, Novais S, Ribeiro AL, Silva S, Frazão O, Phase-shifted fiber Bragg grating by selective pitch slicing, *Sensors* **24**, 6898 (2024). https://doi.org/10.3390/s24216898.
54 Feng D, Zhang P, Yang Q, Wu J, Zhao J, Jiang B, Generation of multiple and ultra-high slope Fano-like resonances via MZI and cascaded fiber gratings, *Opt. Lasers Eng.* **194**, 109223 (2025). https://doi.org/10.1016/j.optlaseng.2025.109223.
55 Wu J, Xuan X, Liu Z, Zhang X, Li Y, Gan X, Jiang B, Synchronous and controllable realization of Lorentzian, Fano, and Electromagnetically induced transparency lineshapes in an all-fiber system, *J. Lightwave Technol.* **43**, 4435–4441 (2025). https://doi.org/10.1109/JLT.2025.3532642.
56 Li Z, Cai L, Liu Y, Jing H, Sun H, Xia F, Zhao Y, Fano resonance in a whispering gallery mode microsphere resonator coupled with femtosecond laser-drilled tapered fiber, *Opt. Laser Technol.* **187**, 112809 (2025). https://doi.org/10.1016/j.optlastec.2025.112809.
57 Wang J, Li Y, Chen Y, He S, Liu Q, Modulation-free laser frequency stabilization via balanced detection of common-path Fano resonance in all-PM fiber cavity, *APL Photonics* **10**, 076106 (2025). https://doi.org/10.1063/5.0275633.
58 Li J, Rong J, Xing E, Jia T, Xing T, Xing G, Yue H, Yang X, Zhou Y, Liu W, Tang J, Liu J, Enhancing dynamic range in microcavity sensing via nonlinear harmonic of Fano resonance, *J. Lightwave Technol.* **43**, 3792–3800 (2025). https://doi.org/10.1109/JLT.2025.3528044.
59 Zou D, Liu R, Shi Y, Zhang A, Li J, Chen GJ, Dang H, Song Y, Hu M, Shum PP, Resonantly driven nonlinear dynamics of soliton molecules in ultrafast fiber lasers, *Adv. Photonics* **7**, 016005 (2025). https://doi.org/10.1117/1.AP.7.1.016005.
60 Rayleigh L, CXII. The problem of the whispering gallery, *Philosophical Mag. Ser. 1* **20**, 1001–1004 (1910). https://www.tandfonline.com/doi/pdf/10.1080/14786441008636993.
61 Garrett CGB, Kaiser W, Bond WL, Stimulated emission into optical whispering modes of spheres, *Phys. Rev.* **124**, 1807–1809 (1961). https://doi.org/10.1103/PhysRev.124.1807.
62 Yu H, Liu X, Sun W, Xu Y, Liu X, Liu Y, A brief review of Whispering Gallery Mode in sensing, *Opt. Laser Technol.* **177**, 111099 (2024). https://doi.org/10.1016/j.optlastec. 2024.111099.
63 Cai L, Li S, Xiang F, Liu J, Liu Q, Fano resonance in whispering gallery mode microcavities and its sensing applications, *Opt. Laser Technol.* **167**, 109679 (2023). https://doi.org/10.1016/j.optlastec.2023.109679.
64 Yariv A, Universal relations for coupling of optical power between microresonators and dielectric waveguides, *Electron. Lett.* **36**, 321–322 (2000). https://doi.org/10.1049/el:20000340.
65 Huang W-P, Coupled-mode theory for optical waveguides: an overview, *J. Opt. Soc. Am. A* **11**, 963–983 (1994). https://doi.org/10.1364/JOSAA.11.000963.
66 Li S, Cai L, Li Z, Liu Y, Liu Q, Fano-like line shape and its tuning effect in a double-handle whispering gallery mode microsphere resonator, *J. Lightwave Technol.* **42**, 6881–6890 (2024). https://doi.org/10.1109/JLT.2024.3412099.
67 Chenari Z, Latifi H, Ranjbar-Naeini OR, Zibaii MI, Behroodi E, Asadollahi A, Tunable Fano-like lineshape in an adiabatic tapered fiber coupled to a hollow bottle microresonator, *J. Lightwave Technol.* **36**, 735–741 (2018). https://doi.org/10.1109/JLT.2017.2772260.
68 Liu W, Wang R, Li W, Zhang B, Xing E, Zhou Y, Sun P, Chen J, Tang J, Liu J, Modulation of Fano-like resonance in spherical microbubble cavity for high sensitivity pressure sensing, *Appl. Phys. Express* **15**, 046504 (2022). https://doi.org/10.35848/1882-0786/ac5c94.
69 Ji P, Zhu M, Liao C, Zhao C, Yang K, Xiong C, Han J, Li C, Zhang L, Liu Y, Wang Y, In-fiber polymer microdisk resonator and its sensing applications of temperature and humidity, *ACS Appl. Mater. Interfaces* **13**, 48119–48126 (2021). https://doi.org/10.1021/acsami.1c14499.
70 Guo Y, Zhang Y, Jiang W, Wu Y, Lineshape modulation and sensing characteristics of resonator assembled with liquid-metal core edge-coupled with microfiber coupler, *J. Lightwave Technol.* **40**, 2516–2522 (2022). https://doi.org/10.1109/JLT.2022.3142011.
71 Jin X, Dong Y, Wang K, Stable controlling of electromagnetically induced transparency-like in a single quasi-cylindrical microresonator, *Opt. Express* **24**, 29773–29780 (2016). https://doi.org/10.1364/OE.24.029773.
72 Xiao Y-F, Jiang X-F, Yang Q-F, Wang L, Shi K, Li Y, Gong Q, Tunneling-induced transparency in a chaotic microcavity, *Laser Photonics Rev.* **7**, L51–L54 (2013).
73 Zhang S, Li J, Yu R, Wang W, Wu Y, Optical multistability and Fano line-shape control via mode coupling in whispering-gallery-mode microresonator optomechanics, *Sci. Rep.* **7**, 39781 (2017). https://doi.org/10.1038/srep39781.
74 Veluthandath AV, Bhattacharya S, Murugan GS, Bisht PB, Fano resonances and photoluminescence in self-assembled high-quality-factor microbottle resonators, *IEEE Photonics Technol. Lett.* **31**, 226–229 (2019). https://doi.org/10.1109/LPT.2018.2889433.
75 Meng-Yu W, Ling-Jun M, Yu Y, Hui-Kai Z, Tao W, Bin L, Lei Z, Yan-Jun F, Ke-Yi W, Selection of whispering-gallery modes and Fano resonance of prolate microbottle resonators, *Acta Phys. Sin.* **69**, 234203–234210 (2020). https://doi.org/10.7498/aps.69.20200817.
76 Wang Y, Zhao H, Li Y, Shu F, Chi M, Xu Y, Wu Y, Mode splitting revealed by Fano interference, *Photon. Res.* **7**, 647–651 (2019). https://doi.org/10.1364/PRJ.7.000647.
77 Wang J, Zhang X, Yan M, Yang L, Hou F, Sun W, Zhang X, Yuan L, Xiao H, Wang T, Embedded whispering-gallery mode microsphere resonator in a tapered hollow annular core fiber, *Photon. Res.* **6**, 1124–1129 (2018). https://doi.org/10.1364/PRJ.6.001124.
78 Bai XQ, Wang DN, An in-fiber coupler for whispering-gallery-mode excitation in a microsphere resonator, *IEEE Photonics Technol. Lett.* **32**, 188–191 (2020). https://doi.org/10.1109/LPT.2020.2965972.
79 Yang G, Cai L, Liu J, Zhao Y, Zhang Z, Xiang F, Zhao Y, In-fiber three-cavity coupling of whispering gallery mode resonators for narrow Fano resonance, *Opt. Laser Technol.* **188**, 112966 (2025). https://doi.org/10.1016/j.optlastec. 2025.112966.
80 Li B-B, Xiao Y-F, Zou C-L, Jiang X-F, Liu Y-C, Sun F-W, Li Y, Gong Q, Experimental controlling of Fano resonance in indirectly

coupled whispering-gallery microresonators, *Appl. Phys. Lett.* **100**, 021108 (2012). https://doi.org/10.1063/1.3675571.
81 Li J, Qu Y, Wu Y, Add-drop double bus microresonator array local oscillators for sharp multiple Fano resonance engineering, *J. Appl. Phys.* **123**, 104305 (2018). https://doi.org/10.1063/1.5022091.
82 Ding M, Qin H, Yin Y, Transmission characteristics and Fano-like lineshape in coupled-slotted microresonators, *J. Lightwave Technol.* **38**, 3687–3693 (2020). https://doi.org/10.1109/JLT.2020.2980845.
83 Yang Y, Wang Z, Zhang X, Zhang Q, Wang T, Recent progress of in-fiber WGM microsphere resonator, *Front. Optoelectron.* **16**, 10 (2023). https://doi.org/10.1007/s12200-023-00066-3.
84 Eggleton BJ, Kerbage C, Westbrook PS, Windeler RS, Hale A, Microstructured optical fiber devices, *Opt. Express* **9**, 698–713 (2001). https://doi.org/10.1364/OE.9.000698.
85 Couny F, Benabid F, Roberts PJ, Light PS, Raymer MG, Generation and photonic guidance of multi-octave optical-frequency combs, *Science* **318**, 1118–1121 (2007). https://doi.org/10.1126/science.1149091.
86 Amsanpally A, Debord B, Alharbi M, Ilinova E, Vincetti L, Gérôme F, Benabid F, *Fano resonance in inhibited coupling Kagome fiber* (CLEO, 2015). Paper STu4L.6 STu4L.6. https://doi.org/10.1364/CLEO_SI.2015.STu4L.6.
87 Miroshnichenko AE, Kivshar YS, Mach–Zehnder–Fano interferometer, *Appl. Phys. Lett.* **95**, 121109 (2009). https://doi.org/10.1063/1.3232224.
88 Ahsani V, Ahmed F, Jun MBG, Bradley C, Tapered fiber-optic Mach-Zehnder interferometer for ultra-high sensitivity measurement of refractive index, *Sensors* **19**, 1652 (2019). https://doi.org/10.3390/s19071652.
89 Lu Y, Yao J, Li X, Wang P, Tunable asymmetrical Fano resonance and bistability in a microcavity-resonator-coupled Mach–Zehnder interferometer, *Opt. Lett.* **30**, 3069–3071 (2005). https://doi.org/10.1364/OL.30.003069.
90 Li A, Jiang B, Zhang P, Gan X, Hao Z, Hou Y, Zhang J, Li P, Zhao J, Realization and modulation of Fano-like lineshape in fiber Bragg gratings, *J. Lightwave Technol.* **39**, 4419–4423 (2021). https://opg.optica.org/jlt/abstract.cfm?uri=jlt-39-13-4419.
91 Li L, Zhang X, Chen L, Optical bistability and Fano-like resonance transmission in a ring cavity-coupled Michelson interferometer, *J. Opt. A: Pure Appl. Opt.* **10**:075305 (2008). https://doi.org/10.1088/1464-4258/10/7/075305.
92 Agliullin TA, Sakhabutdinov AZ, Valeev BI, Kazakov NA, Zakamskiy VK, Combination of fiber optical interferometric structures as a universal approach to sensor development, *Opt. Technol. Telecommun.* **13738**, 58–64 (2025). https://doi.org/10.1117/12.3067423.
93 Piaia V, Alves MR, Robalinho P, Silva S, Frazão O, Fano resonance of fiber Bragg grating for liquid sensing, *J. Lightwave Technol.* 1–7 (2025). https://doi.org/10.1109/JLT.2025.3642982.
94 Zhang L, Shang X, Cao S, Jia Q, Wang J, Yan W, Qiu M, Optical steelyard: high-resolution and wide-range refractive index sensing by synergizing Fabry–Perot interferometer with metafibers, *PhotoniX* **5**, 24 (2024). https://doi.org/10.1186/s43074-024-00138-3.
95 Arianfard H, Weiss T, Yang Y, Bader J, Castelletto S, Peruzzo A, *Fano-like resonances in coupled Sagnac interferometers formed by a self-coupled waveguide* (2025). https://doi.org/10.48550/arXiv.2511.07048.
96 Liang W, Huang Y, Xu Y, Lee RK, Yariv A, Highly sensitive fiber Bragg grating refractive index sensors, *Appl. Phys. Lett.* **86**, 151122 (2005). https://doi.org/10.1063/1.1904716.
97 Iadicicco A, Cusano A, Campopiano S, Cutolo A, Giordano M, Thinned fiber Bragg gratings as refractive index sensors, *IEEE Sens. J.* **5**, 1288–1295 (2005). https://doi.org/10.1109/JSEN.2005.859288.
98 Liu Y, Liu X, Ma C, Zhou Y, Micro-structured optical fiber sensor for simultaneous measurement of temperature and refractive index, *Opt. Fiber Technol.* **41**, 168–172 (2018). https://doi.org/10.1016/j.yofte.2018.01.019.
99 Sohn KR, Liquid sensors using refractive intensity at the end-face of a glass fiber connected to fiber-Bragg grating, *Sens. Actuators A Phys.* **158**, 193–197 (2010). https://doi.org/10.1016/j.sna.2009.12.039.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 54
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026042
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE

OPEN ACCESS

L-band dual-wavelength fiber ring laser with automated polarization stabilization and switching for remote sensing applications

Santiago Tainta[1,2], Alvaro Salinas[1], Iñaki Janices[1], Uxue San-Miguel[1], Arturo Sanchez-Gonzalez[1,2], María José Erro[1,2], and Rosa Ana Perez-Herrera[1,2,*]

[1] Department of Electrical, Electronic and Communications Engineering, Public University of Navarra, 31006 Navarra, Spain
[2] Institute of Smart Cities (ISC), Public University of Navarra, 31006 Navarra, Spain

Received 30 January 2026 / Accepted 3 May 2026

Abstract. This work reports the experimental demonstration of a dual-wavelength L-band fiber ring laser for remote sensing applications. The system incorporates a polarization-sensitive semiconductor optical amplifier as the gain medium and two fiber Bragg gratings placed 25 km away from the laser cavity using standard single-mode fiber that serve both as wavelength-selective elements and sensing heads. Wavelength switching between single- and dual-channel lasing configurations is enabled by a simplified two-paddle motorized polarization controller. The system achieves optical signal-to-noise ratios exceeding 55 dB and power differences between lasing lines as low as 0.01 dB. To ensure long-term stability, an automatic control algorithm dynamically adjusts the polarization state in real time, compensating for environmentally induced polarization drift. The proposed setup provides a compact and robust solution for polarization-based wavelength switching in fiber lasers, with applications in the field of remote optical sensing.

Keywords: Booster semiconductor optical amplifier, Fiber Bragg grating, Fiber-optic ring cavity laser, Polarization-switchable laser, Remote sensing.

1 Introduction

The development of single and multiwavelength fiber-based lasers has been an active area of research for decades, not only as they can serve as optical sources in the C-band of optical fiber communication systems, but also because of their applicability in fields such as sensing or spectroscopy, among others. While mostly developed for the C-band, there have been several proposals also for the L-band (1565–1625 nm), since their use in this wavelength range allows to increase the capacity of fiber links [1], the simplification and improvement of lidar systems [2], the extension of sensing systems to other contexts [3] and the application in spectroscopy systems [4]. An especially relevant example is the detection of carbon monoxide, carbon dioxide, methane, and other potentially harmful gases with absorption bands in in the 1–2 μm spectral region, which has a clear impact in the protection of people and infrastructures in the oil and gas industries, water treatment plants, landfills, and commercial or domestic environments [5–7]. One of the most obvious advantages of using optical technologies in this context, the safety they offer by avoiding potentially flammable components, is reinforced if the sensor head can be placed at a distance from the rest of the components, especially if the wavelengths can be remotely switched in order to enable the detection of different gases simultaneously. Switchable L-band multiwavelength lasers have been previously proposed for such applications, primarily based on fiber ring schemes [8]. Erbium doped fiber, which is the usual gain medium at lower wavelengths, presents a reduced emission cross-section beyond 1.6 μm. As a result, achieving sufficient gain at these longer wavelengths requires either longer fiber or higher doping concentrations. Additionally, the strong homogeneous broadening gain that Erbium doped fibers present leads to strong mode competition and results in poor laser stability. Addressing this issue requires the use of more complex setups or additional stabilization techniques [9–11]. Semiconductor optical amplifiers (SOAs) are consequently an interesting option as gain media in this wavelength region [12, 13]. In order to enable wavelength selection within the laser cavity, different optical filters are usually included into the ring structure, such as Fiber Bragg Gratings (FBGs) [14], Sagnac filters [8, 15] or Mach-Zehnder interferometers [16], among other techniques. Wavelength switching is typically accomplished by modifying the polarization state within the cavity, altering the ring cavity losses and, consequently, modifying the gain spectral distribution.

* Corresponding author: rosa.perez@unavarra.es

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

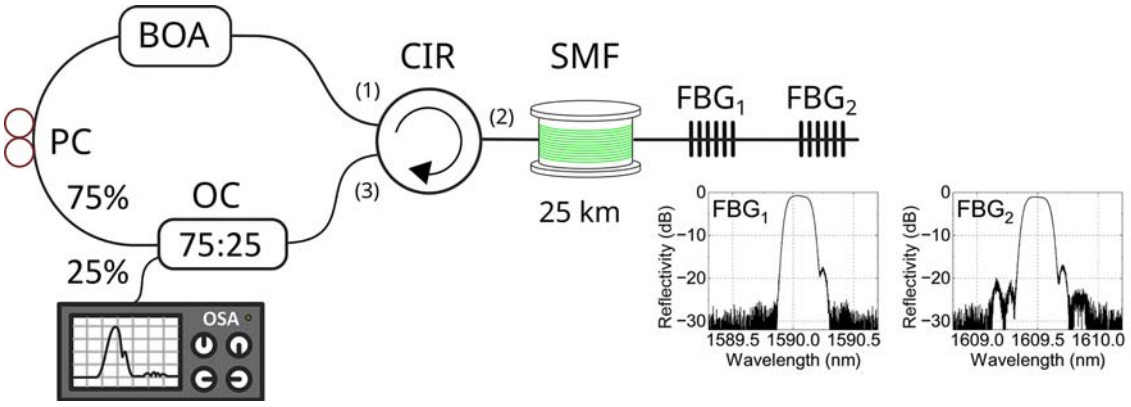


Fig. 1. Experimental setup of the proposed L-band multiwavelength remote sensor based on a motorized polarization-switchable fiber optic ring laser (BOA: Booster Optical Amplifier, PC: Polarization Controller, OC: Optical Coupler, CIR: Circulator, SMF: Single-Mode Fiber, FBG: Fiber Bragg Grating, OSA: Optical Spectrum Analyzer).

In [17] authors reported a system for remote sensing applications based on a switchable L-band fiber optic ring cavity laser incorporating a SOA and multiple FBGs. That study experimentally validated the capability of this architecture to detect temperature variations at remote locations. We demonstrated that by adjusting a motorized fiber polarization controller (PC) located 25 km away from the sensing head the system can be remotely switched between single-wavelength, dual-wavelength, and triple-wavelength lasing configurations. However, since the operation of the setup relies on the polarization properties of the components, the system exhibited limited stability, a common issue in polarization-sensitive architectures [10, 11, 18, 19]. In this work, a simplified version of the configuration described in [17] is presented, reducing the polarization control from three to just two paddles of a motorized PC to remotely achieve the desired wavelength selection. Furthermore, we implement an automatic control mechanism for the PC adjustment, resulting in significantly improved system stability over long periods of time. This approach effectively decouples the sensing information from power fluctuations: while the PC-tuning algorithm compensates for amplitude drifts in the power domain, the sensing data remain encoded in the wavelength shift of the spectral lines, ensuring that the measurement integrity is preserved.

2 Experimental setup

The experimental configuration of the laser system, depicted in Figure 1, is based on a fiber-optic laser operating at two distinct wavelengths and implemented using a ring cavity architecture. Notably, the system is implemented using non-polarization-maintaining elements, which represents a worst-case scenario as polarization fluctuations occur throughout the setup. While the use of polarization-maintaining components would likely result in a higher stability and enable a more deterministic output for a given PC configuration, the current setup allows for testing the robustness of the stabilization mechanism. A three-port optical circulator (CIR) is employed to control light propagation within the ring cavity. Optical gain for lasing is provided by an L-band booster-type semiconductor optical amplifier (BOA, model S9FC1080P, from Thorlabs). The circulator directs the amplified signal from the BOA toward a reflective filtering module, enabling the filtered light to be re-coupled into the cavity and ensuring unidirectional propagation of the optical field.

Figure 2a presents the amplified spontaneous emission (ASE) spectra of the BOA at various driving current levels. At a driving current of 500 mA, the BOA exhibits a peak power of −29 dBm at 1583 nm with a 120 nm bandwidth. As observed, increasing the driving current result in both higher output power and a broader emission spectrum while the peak wavelength shifts toward shorter wavelengths. By design the BOA is inherently polarization-sensitive, amplifying only a specific linear polarization state. This property effectively makes the BOA act as a polarization-selective element within the laser cavity. To experimentally demonstrate this behavior, a PC followed by a polarization beam splitter (PBS) were employed to analyze the output polarization dependence as illustrated in Figure 2b. A Thorlabs MTC320 motorized controller was employed to alter polarization via stress-induced birefringence. By coiling single-mode fiber around independent spools to create fractional wave plates, the system, controlled by Thorlabs software, allows for precise polarization adjustment. This PC was used to align the amplification axis of the BOA with one of the PBS outputs. Under these conditions, at a bias current of 100 mA, the BOA delivers a peak output power of −44 dBm at a central wavelength of 1583 nm, consistent with the results shown in Figure 2a. However, nearly all the power is delivered to the aligned PBS output while the other output port, that corresponds to the orthogonal linear polarization state, is close to the noise level, as depicted in Figure 2c.

As shown in Figure 1, a PC included at the BOA input port allows precise adjustment of the cavity polarization-dependent loss spectrum and, thus, switching between different wavelength emission configurations. Most lasers that incorporate a PC to ensure emission at specific wavelengths rely on the fact that only a particular state of polarization

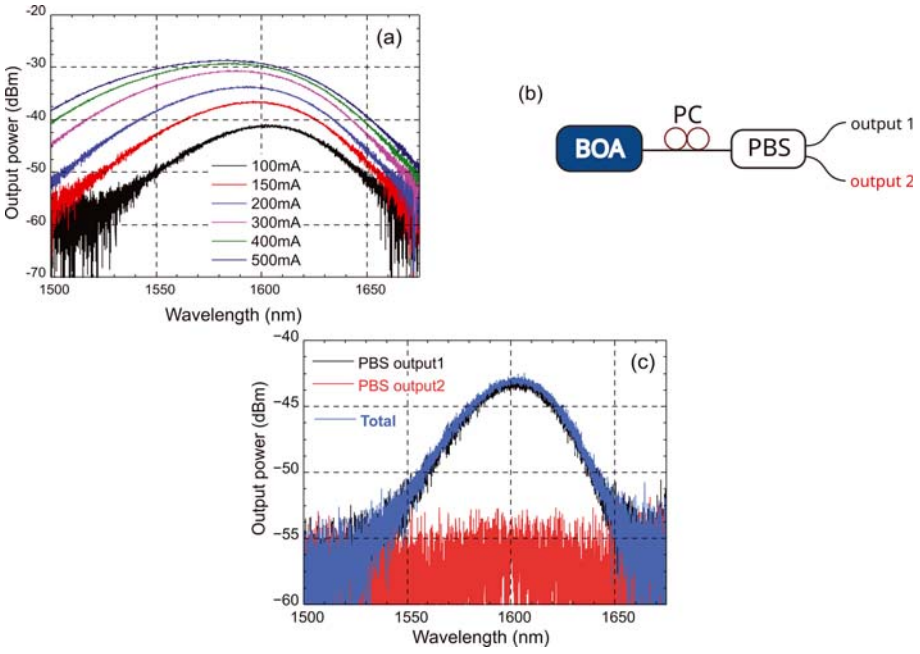


Fig. 2. (a) ASE spectra at BOA's output for different driving currents; (b) Experimental setup employed for evaluating BOA's polarization dependence; (c) Output power spectra at the PBS output ports.

(SOP) can resonate within the cavity [20–22]. In conventional three-paddle PCs, formed by a cascade of three retarder plates in the form QWP–HWP–QWP (QWP: quarter-wave plate, HWP: half-wave plate), precise adjustment of the orientations of the three paddles allows an arbitrary input polarization to be transformed into a well-defined output SOP [23], thereby enabling resonance at the wavelength associated with that polarization state. In contrast, in the present work, a reduced two-paddle PC, consisting of a QWP followed by an HWP, is employed. Although this configuration cannot generate an arbitrary SOP independently of the input, Jones-matrix analysis shows that it is sufficient to convert a general elliptical polarization into a linearly polarized state with a controllable orientation by proper adjustment of the plates' angles [24]. When combined with the strong polarization-dependent gain of the BOA, which is shown in Figure 2c, this linear polarization rotation translates into distinct cavity gains for different SOPs, enabling polarization-selective amplification and wavelength switching without the need for a full three-paddle PC. Hence, with only two paddles to be adjusted instead of the usual three, the tuning process for selecting the desired lasing configuration becomes simpler and faster. The ring is closed with a 75:25 coupler that extracts part of the generated light for its measurement with an optical spectrum analyzer with a resolution of 0.03 nm, (OSA, model MS9740B-009, from Anritsu).

The filter used to define the emitted lines wavelengths is constructed by concatenation of two FBGs. Their central wavelengths, 1590 nm (FBG_1) and 1610 nm (FBG_2), are positioned within the BOA's L-band ASE spectrum. As illustrated in Figure 1, both FBGs exhibit a 0.22 nm bandwidth with reflectivities of −0.75 dB and −1.05 dB, respectively. Despite a minor mismatch in reflectivities, the variation is insufficient to inhibit dual-wavelength operation. Besides their filtering function, these FBGs are separated from the rest of the setup by a 25 km spool of single mode fiber (SMF), serving also as the remotely controlled sensing mechanism, as demonstrated in [17], and enabling long distance sensing applications. All experimental measurements were conducted at room temperature, and no vibration isolation or temperature compensation techniques were employed.

3 Results

The switchable operation of the system is validated by driving the BOA at bias currents of 250, 300, and 350 mA, with all measurements performed at room temperature. The laser emission configuration is controlled by adjusting the rotation of both PC paddles. To explore the full range of possible operating modes under different biasing conditions, all configurations – namely no-lasing, single-wavelength, and dual-wavelength states – are systematically tuned using a search algorithm. This algorithm evaluates all possible paddle configurations of the motorized PC and optimizes different parameters depending on the target configuration. For single-wavelength and no-lasing states, the paddle positions are optimized to maximize suppression of undesired emission lines. For dual-wavelength operation, the optimization minimizes the optical power difference between the two lasing peaks, ensuring balanced output across both channels.

An example of the tunability of the system can be seen in Figure 3. In this example, the QWP is held fixed while the HWP is rotated to different positions with a 5-degree resolution. As shown, different combinations of optical

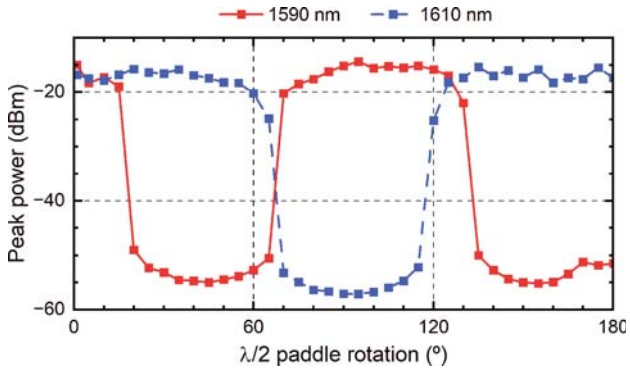


Fig. 3. Optical peak power for each emission wavelength as a function of the HWP adjustment, with the QWP held fixed.

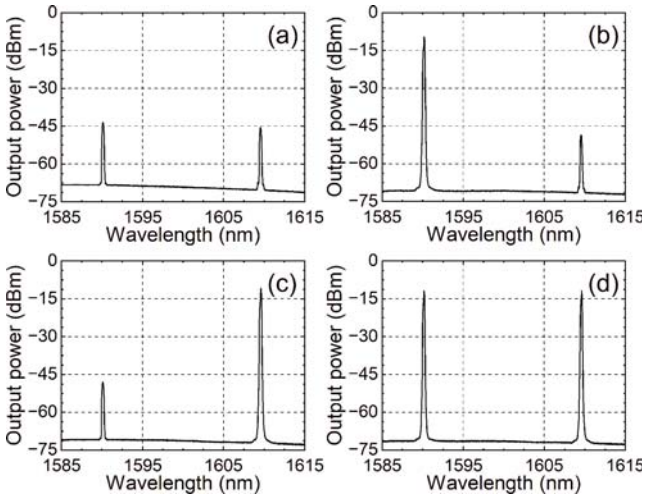


Fig. 4. Output spectra measured in the optical domain for the: (a) no-lasing, (b) single-wavelength at 1590 nm (c) single-wavelength at 1610 nm and (d) dual wavelength lasing configurations, for a biasing current of 300 mA at the BOA.

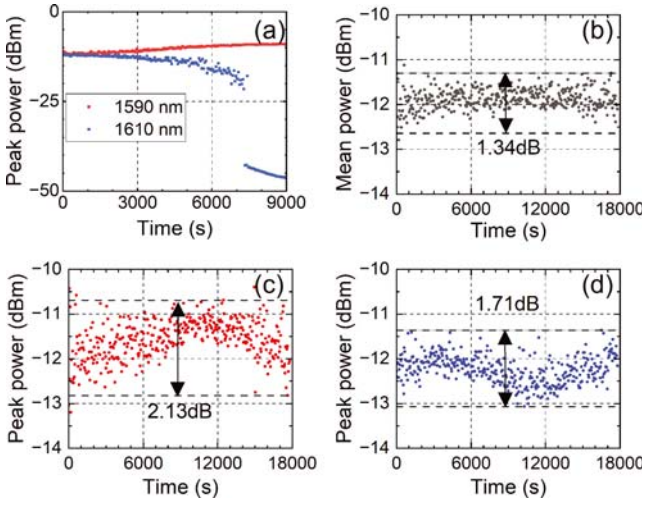


Fig. 5. (a) Output peak powers for a dual-wavelength configuration in free-running mode; (b) mean peak power; (c) 1590 nm and (d) 1610 nm peak for the dual-wavelength configuration with the stabilization loop activated, indicating the CL at 95% (dashed lines).

power for the emission wavelengths can be achieved depending on the paddle positions. Therefore, by simultaneously adjusting both paddles, an optimization process can be performed to obtain the desired output combinations.

The results of the tuning process for all lasing configurations at a bias current of 300 mA are shown in Figure 4. Figure 4a corresponds to the no-lasing configuration, where both emission wavelengths exhibit a power level below −45 dBm. In contrast, the single-wavelength emission configurations presented in Figures 4b and 4c show that both emission lines exceed a peak optical power of −15 dBm, resulting in an optical signal-to-noise ratio (OSNR) greater than 55 dB in both cases, as well as a main lobe-to-sidelobe suppression ratio (MSSR) exceeding 40 dB. Finally, Figure 4d illustrates the dual-wavelength configuration, where the peak output power of each emission wavelength is −11.84 dBm. The power difference measured in this configuration is less than 0.01 dB, and the OSNR for both wavelengths is higher than 60 dB. These parameters align with established standards for high-performance laser-sensing in the C [25], O [26], and L-bands [3].

Table 1 provides a comprehensive summary of the results obtained for all configurations across the three bias current levels. The data include the power OSNR of each emission line, along with the power difference between both lines for each configuration. The results are consistent across all cases, with only minor variations attributed to increased available power at higher bias currents. In all non-lasing cases, line suppression exceeds 42 dB, while in all single-line emission configurations, the MSSR is at least of 37 dB and the OSNR is greater than 55 dB. For the dual-wavelength emission configuration, the OSNR surpasses 58 dB in all cases, with the maximum power difference between emission lines remaining below 0.5 dB and reaching as low as 0.01 dB in one instance, demonstrating an exceptional level of output equalization. It is worth noting that these results could be further improved with enhanced polarization control, which was limited in this study by the motorized controller's minimum rotation angle of 0.2°. Nonetheless, the results validate the proposed system and demonstrate its ability to maintain high performance under varying operational conditions.

The dual-wavelength configuration with a 300 mA bias current was chosen to evaluate the stability of the system. First, output spectra were captured every 45 s for a total of 2.5 h, without any adjustments of the PC during all the experiment duration. The registered evolution of the two wavelengths peak powers is plotted in Figure 5a, showing that the system is not able to preserve the dual-wavelength emission, since the 1610 nm line power progressively decreases and eventually falls drastically after approximately 2 h, becoming then the laser a single line device. This behavior prevents the use of this free-operating

Table 1. Switching results for different configurations and biasing currents.

		Peak power (dBm)		OSNR (dB)		Peak power diff. (dB)
		λ_1	λ_2	λ_1	λ_2	$\lambda_1 - \lambda_2$
250 mA	No-lasing	−52.2	−52.4	18.8	18.6	0.2
	λ_1 lasing	−10.1	−51.9	61.8	20.1	41.8
	λ_2 lasing	−48.8	−11.9	23.2	60.0	36.9
	$\lambda_1 + \lambda_2$ lasing	−12.7	−13.1	59.4	58.8	0.5
300 mA	No-lasing	−43.5	−45.6	25.5	23.4	2.0
	λ_1 lasing	−9.7	−48.3	61.0	22.0	38.9
	λ_2 lasing	−48.1	−10.8	22.7	60.0	37.3
	$\lambda_1 + \lambda_2$ lasing	−11.8	−11.8	59.5	59.5	0.0
350 mA	No-lasing	−49.4	−50.0	21.1	20.5	0.6
	λ_1 lasing	−8.3	−50.7	62.2	19.8	42.4
	λ_2 lasing	−54.0	−14.4	16.9	56.6	39.6
	$\lambda_1 + \lambda_2$ lasing	−10.7	−10.8	60.0	60.0	0.1

setup in optical sensing applications requiring stable long-term dual wavelength operation.

Since the cause of the undesired variations of the generated spectrum are the unavoidable random fluctuations of the polarization state of the light propagating within the ring, an automatic control loop was included to respond and recover from these oscillations. The operational procedure of the control loop begins with an initialization phase to identify a reference angular configuration that achieves the desired spectral configuration. Once this configuration is achieved, the system continuously monitors the optical power spectrum and triggers a local search algorithm if the difference between the peak powers of the desired spectrum and the measured spectrum exceeds a 1 dB threshold. This search prioritizes the HWP due to its critical impact on equalization, testing positive and negative angular variations of increasing magnitude until an error below 0.5 dB is achieved again. Should no valid equalization be found by rotating only the HWP, the algorithm selects the best-performing angle and subsequently adjusts the QWP rotation to identify the optimal combination within the search interval by following a similar procedure. While this mechanism significantly improves the management of slow drifts, the search process itself introduces transient fluctuations, such as momentary drops in emission lines, indicating that the current implementation is not yet fully optimized for all operating conditions. However, as undesired polarization fluctuations typically occur at slower rates than the time required for the control loop to reoptimize the PC positions, this scheme correctly anticipates and prevents significant deviations in the produced spectrum, resulting in an important improvement on the stability over long periods of continuous operation of the laser. When the control loop is enabled, the output spectrum remains stable, as demonstrated in Figures 5b–5d, over an extended 5 h recording period. In Figure 5b the mean value of the peak powers for both wavelengths is presented, where the maximum fluctuation of the signal is 1.34 dB for a 95% confidence level (CL). Additionally, Figures 5c–5d show the peak powers of each wavelength, which present a maximum fluctuation of 2.13 and 1.71 dB respectively for a 95% CL. These fluctuations could potentially be further reduced, since the algorithm currently adjusts the PC paddles only when the power difference between the two lines exceeds a predefined threshold, without accounting for the individual variation trends. Additionally, as previously noted, the resolution of the control is constrained by the minimum rotation step of the paddles in our device, which is 0.2°.

4 Conclusion

A multiwavelength switchable fiber ring laser operating in the L-band for sensing applications has been experimentally demonstrated. The system is based on a polarization-sensitive SOA as the gain medium and two FBGs that define the emitted wavelengths and that can simultaneously function as remote sensing elements. The FBGs were deployed 25 km from the system header via standard single-mode fiber, thereby enabling long-range remote sensing applications. The laser can be electronically switched at the header between all emission configurations – no lasing, single-wavelength, or dual-wavelength operation – with OSNR consistently exceeding 55 dB, MSSR of at least 37 dB, and power differences between lasing lines in the dual-wavelength configuration below 0.5 dB. Wavelength switching is enabled by a simplified motorized polarization controller composed of only two paddles (a QWP followed by a HWP), without the need for additional polarization-dependent elements. This configuration supports automatic control, enabling stable long-term operation without the need for environmental isolation or polarization-maintaining fiber thanks to a closed loop algorithm that self-stabilizes the output by fine-tuning the paddles rotations.

Funding

This work was supported in part by projects PID2022-137269OB, funded by MCIN/AEI/10.13039/501100011033 and FEDER "A way to make Europe" and PID2019-104426GB-100 funded by MCIN/AEI, and the Public University of Navarre collaboration grants.

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

The datasets generated and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

Author contribution statement

Conceptualization, S.T., M.J.E. and R.A.P-H.; Methodology, S.T., M.J.E. and R.A.P-H; Software, A.S., A.S-G., S.T., M.J.E. and R. A.P-H; Validation, A.S., I.J., U.S-M, A.S-G.; Formal Analysis, S. T, M.J.E. and R.A.P-H.; Investigation, S.T, M.J.E., A. S-G. and R.A.P-h.; Resources, S.T., M.J.E. and R.A.P-H.; Data Curation, A.S., I.J., U.S-M, A.S-G.; Writing – Original Draft Preparation, S.T, M.J.E. and R.A.P-H.; Writing – Review & Editing, S.T., A.S., I.J., U.S-M, A.S-G., M.J.E. and R.A.P-H.; Visualization, A.S., I.J., U.S-M, A.S-G.; Supervision, S.T., M.J.E. and R.A.P-H.; Project Administration, S.T., M.J.E. and R.A.P-H.; Funding Acquisition, S.T., M.J.E. and R.A.P-H.

References

1 Lopez V, Zhu B, Moniz D, Costa N, Pedro J, Xu X, Kumpera A, Dardis L, Rahn J, Sanders S, Optimized design and challenges for C&L band optical line systems, *J. Lightwave Technol.*, **38**, 5 (2020). https://doi.org/10.1109/JLT.2020.2968225.

2 Wang KX, Gao CQ, Lin ZF, Wang Q, Gao MW, Huang S, Chen CY, 1645 nm coherent Doppler wind lidar with a single-frequency Er:YAG laser, *Opt. Expr.*, **28**, 10 (2020). https://doi.org/10.1364/OE.392092.

3 Perez RA, Ullan A, Leandro D, Fernandez-Vallejo M, Quintela MA, Loayssa A, Lopez-Higuera JM, Lopez-Amo M, L-Band multiwavelength single-longitudinal mode fiber laser for sensing applications, *J. Lightwave Technol.*, **30**, 8 (2012). https://doi.org/10.1109/JLT.2011.2174138.

4 Sato H, Saito N, Akagawa K, Wada S, Tashiro H, Electronically tunable-laser light sources for near infrared spectroscopy, *J. Near Infrared Spectrosc.*, **11**, 4 (2003). https://doi.org/10.1255/jnirs.375.

5 Hu M, Hu M, Wang W, Wang Q, Wavelength-scanned all-fiber cavity ring-down gas sensing using an L-band active fiber loop, *Appl. Phys. B*, **128**, 30 (2022). https://doi.org/10.1007/s00340-022-07757-3.

6 Zhang H, Bi S, Zhang Q, Tian C, Wang Z, The fiber ring laser intra-cavity gas sensor for C2H2 and CO2 detection based on photoacoustic spectroscopy, *Infrared Phys. Technol.*, **131**, 104623 (2023). https://doi.org/10.1016/j.infrared.2023.104623.

7 Marshall J, Stewart G, Whitenett G, Design of a tunable L-band multi-wavelength laser system for application to gas spectroscopy, *Measure. Sci. Technol.*, **17**, 5 (2006). https://doi.org/10.1088/0957-0233/17/5/S15.

8 Zhai Y, Lu B., Meng L, Chen HI, Bai J, 1.6 μm tunable and switchable single-frequency erbium-doped fiber laser, *Opt. Expr.*, **32**, 22 (2024). https://doi.org/10.1364/OE.540981.

9 Lin Q, Wang Y, Lv C, Zhang H, Gang T, Zhang C, Yang X, Lu B, Chen H, Conversion of bound states to noise-like pulses in an L-band mode-locked fiber laser, *Opt. Fiber Technol.*, **94**, 104291 (2025). https://doi.org/10.1016/j.yofte.2025.104291.

10 Zhang J, Lv X, Ren W, Zhang L, Fu S, Ren G, Zheng X, Multi-wavelength switchable erbium-doped fiber laser based on single mode- tapered multicore- single mode fiber interferometer, *Opt. Fiber Technol.*, **93**, 104215 (2025). https://doi.org/10.1016/j.yofte.2025.104215.

11 Zhao Q, Pei L, Zheng J, Tang M, Xie Y, Li J, Switchable, widely tunable and interval-adjustable multi-wavelength Erbium-doped fiber laser based on cascaded filters, *J. Lightwave Technol.*, **38**, 8 (2020). https://doi.org/10.1109/JLT.2020.2973454.

12 Ummy MA, Madamopoulos N, Joyo A, Kouar M, Dorsinville R, Tunable multi-wavelength SOA based linear cavity dual-output port fiber laser using Lyot-Sagnac loop mirror, *Opt. Exp.*, **19**, 4 (2011). https://doi.org/10.1364/OE.19.003202.

13 Liu T, Jia D, Yang T, Wang Z, Liu Y, Stable L-band multi-wavelength SOA fiber laser based on polarization rotation, *App. Opt.*, **56**, 10 (2017). https://doi.org/10.1364/AO.56.002787.

14 Yan M, Luo S, Zhang L, Zhang Z, Xia Y, Triple-wavelength switchable Erbium doped fiber laser with cascaded asymmetric exposure long-period fiber gratings, *Opt. Exp.*, **15, 7** (2007). https://doi.org/10.1364/OE.15.003685.

15 Sulaiman AH, David AP, Ismail A, Abu Hassan NN, Abdullah F, Jamaludin MZ, Yusoff NM, Sagnac-interferometer-based multiwavelength SOA fiber laser assisted by an intensity-dependent loss mechanism, *Appl. Opt.*, **63**, 5 (2024). https://doi.org/10.1364/AO.506145.

16 Geng X, Jiang Y, Gu H, Luo S, Sun M, Li L, Switchable multi-wavelength fiber lasers based on asymmetric biconical fiber tapers, *Opt. Commun.*, **548**, 129837 (2023). https://doi.org/10.1016/j.optcom.2023.129837.

17 Perez-Herrera RA, Janices I, San-Miguel U, Salinas A, Sanchez-Gonzalez A, Erro MJ, Tainta S, Polarization-switched fiber optic ring laser for L-band multiwavelength remote sensing applications, in *Proc. of 29th International Conference on Optical Fiber Sensors, Porto, Portugal, 136397V* (2025), **Volume 13639**, https://doi.org/10.1117/12.3062987.

18 Ye G, Liu B, Dai M, Ma Y, Shirahata T, Yamashita S, Set SY, Pump-power-controlled L-band wavelength-tunable mode-locked fiber laser utilizing nonlinear polarization evolution in all-polarization-maintaining fibers, *Opt. Lett.*, **49**, 9 (2024). https://doi.org/10.1364/OL.518882.

19 Aliza HEM, Sulaiman AH, Ismail A, Abdullah F, Yusoff NM, Ibrahim SA, Jamaludin MZ, Enhanced multiwavelength random fiber laser based on hybrid optical amplifier incorporating Sagnac loop mirror interferometer, *Result. Phys.*, **72**, 108212 (2025). https://doi.org/10.1016/j.rinp.2025. 108212.

20 Wei L, Xu X, Khatak A, Henley B, Continuously tunable comb filter based on a high-birefringence fiber loop mirror with a polarization controller, *J. Lightwave Technol.*, **39**, 14 (2021). https://doi.org/10.1109/JLT.2021.3077252.

21 Feng S, Mao Q, Shan L, Lit, Reflectivity characteristics of the fiber loop mirror with a polarization controller, *Opt. Commun.*, **277** (2007). https://doi.org/10.1016/j.optcom.2007.05.016.

22 Hao Y, Miao M, Liao W, Yang K, Polarization influence on Er3+−doped multi-wavelength Brillouin fiber laser based on fiber loop mirror, *Photonics*, **11**, 659 (2024). https://doi.org/10.3390/photonics11070659.

23 Heismann F, Analysis of a reset-free polarization controller for fast automatic polarization stabilization in fiber-optic transmission systems, *J. Lightwave Technol.*, **12**, 4 (1994). https://doi.org/10.1109/50.285366

24 Heidrich H, Hoffmann D, Helmot CHv, Ahlers H, Lithium niobate guided-wave network for a coherent receiver, *Opt. Lett.*, **42**, 1 (1989). https://doi.org/10.1364/OL.14.000099

25 Pevec S, Donlagic D, Multiparameter fiber-optic sensors: a review, *Opt. Eng.*, **58**, 7, 464 (2019), Art. no. 072009. https://doi.org/10.1117/1.OE.58.7.072009.

26 Ahmad H, Nizamani B, Melkumov MA, Lobanov AS, Tunable dual-wavelength bismuth-doped fiber laser in O-band using in-line six-mode fiber filter. *Opt. Fiber Technol.*, **95**, 104393 (2025). https://doi.org/10.1016/j.yofte.2025.104393.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 5
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2025056
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

Anomalous reflection coefficient underlying experimentally validated outstanding transmission through metal–dielectric–metal microcavities

Alejandro Doval, Lucía Suárez-Fernández, Raúl de la Fuente, and Yago Arosa*

Instituto de materiais da USC - iMATUS, Grupo de Nanomateriais, Fotónica e Materia Branda, Departamento de Física Aplicada, Universidade de Santiago de Compostela, E-15782, Santiago de Compostela, Spain

Received 27 October 2025 / Accepted 29 December 2025

Abstract. Analytical modelling of light transmission through a metal-insulator-metal geometry embedded in a coupling glass surrounding medium is possible through an extended Fabry-Pérot formula. Two distinct coupled surface plasmon resonance branches are allowed inside such microcavity, where two thin metallic layers act as mirrors delimiting an inner dielectric material. In agreement with transfer-matrix method simulations, the resulting theoretical expressions predict a large and almost constant transmittance even for intracavity thicknesses greater than light's penetration depth. Results at $\lambda = 800$ nm have been validated experimentally and show optical transmittance over 10% until nearly 3 μm. This high transmittance under such unexpected conditions, related to an anomalously high mirror reflection coefficient, sheds light on new possibilities for the design of optical devices. The experimental setup successfully used to corroborate the validity of the transmittance formula over different angular, spectral and geometrical conditions is also presented.

Keywords: Plasmonics, Coupled surface plasmons, Microcavities, Transmittance, Resonance.

1 Introduction

Coupled Surface Plasmons (CSP) in metal-dielectric-metal (MDM) and dielectric-metal-dielectric (DMD) geometries started attracting theoretical interest since the late 20th century [1]. Initial experimental validations [2] opened the path for posterior applications of both arrangements. Specifically, referring to the case of MDM geometry that we will be addressing in this work, applications range from guiding systems [3] to spectral filtering [4], including sensors [5] or spectroscopy [6] as well.

Our recent studies [7] have focused on developing and validating an analytical model for transverse magnetic (TM)-polarized light transmission through an optical plasmonic microcavity (MC), involving an MDM structure. The MC is formed by two thin plane metallic mirrors (M) separated by a low-index dielectric gap (L) and surrounded by two semi-infinite higher-index dielectric media (H). An analytical expression for T – see equation (1) – was obtained by application of Fresnel coefficients at the interfaces, and the subsequent results were calculated by implementing Fresnel formalism for coherent optical scattering at an interface and validated by comparison with simulations using in-house developed software based on transfer-matrix-method [8]. Experimental validation was also carried out, using two identical symmetrically arranged prisms playing the role of the external semi-infinite high-index media, which couple and decouple light to the possible plasmonic resonances of the inner structure.

$$T = \frac{|t_{\mathrm{LMH}}|^2 |t_{\mathrm{HML}}|^2}{4|r_{\mathrm{LMH}}|^2 \left[\sinh^2\left(k''_{L,\perp} d - \ln|r_{\mathrm{LMH}}|\right) + \sin^2\left(k'_{L,\perp} d + \varphi_{\mathrm{LMH}}\right)\right]} \quad (1)$$

where r_{LMH} and t_{LMH} are reflection and transmission field amplitude coefficients for each of the two three-medium mirror structures; φ_{LMH} represents the phase of r_{LMH}; d denotes the intracavity thickness; and $k_{L,\perp} = k'_{L,\perp} + ik''_{L,\perp}$ is the component of the intracavity wavevector in the direction perpendicular to the interfaces. This equation can be seen as a generalization of the common formula for transmittance in Fabry-Pérot (FP) interferometers to all incidence angles, from which the different resonant modes in the system can be obtained. On the one hand, $k_{L,\perp}$ is real for incidence angles θ below the critical angle for incidence from H to L, θ_{cr}, for which volume resonances are allowed. That case corresponds to the well-known FP regime, in which harmonic waves can

* Corresponding author. yago.arosa.lobato@usc.es

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

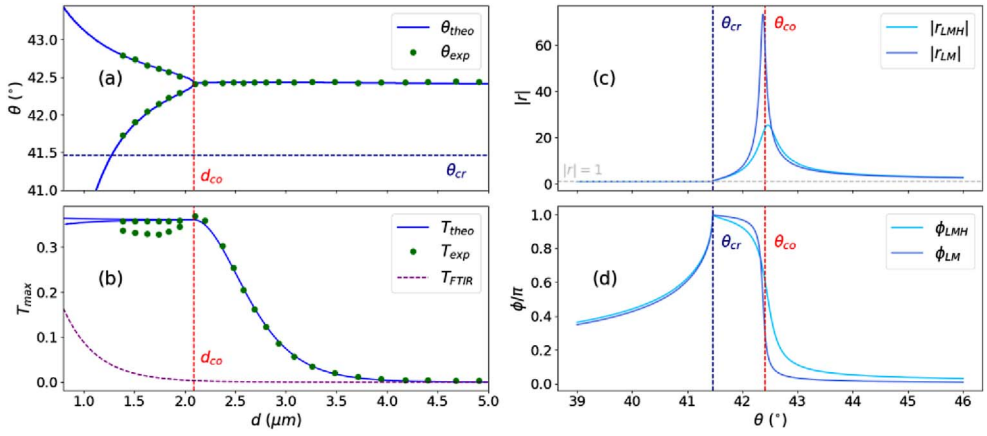


Fig. 1. (a) Angle of incidence at which the transmittance of each CSP is maximum as a function of cavity thickness. (b) Maximum transmittance at each CSP resonance peak as a function of cavity thickness, FTIR transmission at θ_{co} without Ag for comparison. (c) Modulus and (d) phase of r_{LMH} (lighter) and r_{LM} (darker) as functions of the incident angle. All results correspond to a setup using silver as metal ($d_M = 39$ nm), BK7 as prism and air filling the gap between the mirrors, at $\lambda = 800$ nm.

propagate between the two mirrors at photonic resonance conditions. On the other hand, for $\theta > \theta_{cr}$, $k_{L,\perp}$ becomes imaginary, and only CSP resonances are allowed. In both cases, the dependence on the separation between the mirrors, d, in equation (1) becomes straightforward, since it only appears inside the argument of either the sine or the hyperbolic sine, respectively. As a result, CSP resonances correspond to zeros of the hyperbolic sine, leading to two distinct solutions for each d at two different θ, as shown in Figure 1a. It is also remarkable that one of these two curves is the continuation of the first FP resonance, forming a hybrid branch. For thick enough cavities these two plasmonic resonances become degenerated, merging at the same angular position – named the coalescence angle θ_{co} – for $d > d_{co} = \max[(\ln|r_{LMH}|)/k''_{L,\perp}]$.

In this work, we emphasize the high values of transmittance in the system even for over-wavelength-thick cavities. The presented theoretical and experimental studies of maximum transmittance as a function of d reveal that it remains not only noticeable but also large for d exceeding by far the predicted penetration depth of light in the intracavity L medium. Thus, transmittance of CSP resonances in this MC can be said to resemble some kind of enhanced light tunneling. Analysis of equation (1) reveals that r_{LMH} is the key parameter determining this phenomenon, especially in the plasmonic regime. A brief theoretical analysis on the behavior of that reflection coefficient is also presented. Its modulus becomes much larger than unity near the coalescence. This may seem surprising if one thinks of Fresnel coefficients as the square root of reflectance (R). However, there is no issue with energy conservation here because the waves involved are evanescent and $|r| = \sqrt{R}$ does not hold.

2 Material and methods

The experimental setup used to test our theoretical predictions is shown in Figure 2. A supercontinuum light source (SuperK COMPACT, NKT Photonics, 450–2400 nm) illuminates the MC, with power adjusted by a variable neutral optical density. A cube polarizer provides a transmitted TM beam, which is a requirement for the excitation of surface plasmons. Simultaneously, part of the original beam is deflected towards a photodiode, used to monitor the beam power. The MC consists of two identical rectangular BK7 glass coupling prisms, whose largest faces are coated with thin silver films, facing each other and creating a micron-scale air gap in between. That cavity thickness is determined by iteratively comparing with subsequent theoretical configurations, varying d with the desired precision. It is modified in ~20 nm steps using a piezoelectric actuator (Thorlabs PIAK25), which is connected to the prisms through custom 3D-printed mounts. These are mounted on a rotation platform (URS75BCC from Newport Optics), that allows us to control the angle of incidence (±6 mdeg accuracy). Finally, the transmitted signal is collected with a photodiode preceded by a spectral filter, enabling monochromatic mapping of T across a 2D $\theta - d$ space, using incident light as a reference.

The thin silver layers on the prisms, used as MC mirrors, were deposited by physical vapor deposition (PVD), evaporating the metal through heating. The prisms were placed inside the vacuum chamber of a Baltec BAE250 Coating System by Balzers, and the pressure was reduced to ~10^{-3} Pa. Afterwards, the silver was heated until it started to evaporate (~140 °C), and an approximately constant deposition rate was used to obtain silver films with a nominal thickness around 39 nm. Later, this value was found to be correct within an uncertainty range of ±5 nm with a Dektak3 Surface Profilometer, as an extra check.

3 Results and discussion

The analytical model developed for transmittance of the MC was experimentally validated. Angular positions of

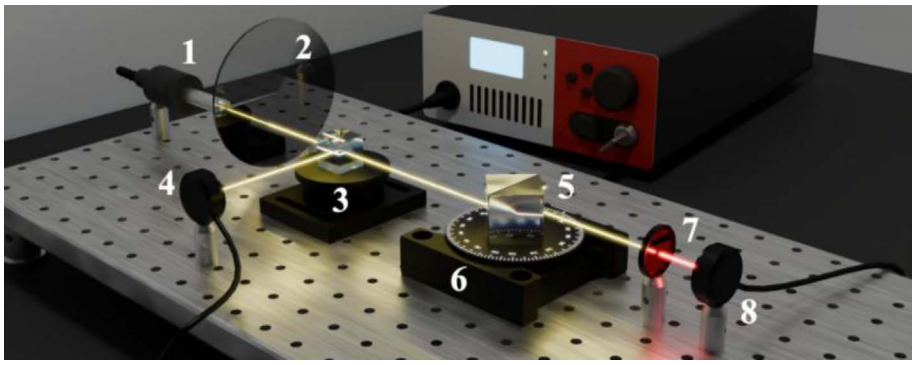


Fig. 2. 3D schematic drawing of the experimental setup. (1) Output of the supercontinuum light soruce, (2) variable optical density filter, (3) cube acting as polarizer and beam splitter, (4) power monitoring photodiode, (5) couple of prisms conforming the MC, (6) rotation platform, (7) narrrowband spectral filter, (8) photodiode detector for measuring transmittance.

the transmittance maxima for the two CSP resonances as a function of d at a fixed wavelength $\lambda = 800$ nm are displayed in Figure 1a, while their peak values are presented in Figure 1b. Both graphs show the agreement between theoretical predictions and experimental measurements. Those results correspond to a BK7-Silver-Air MC, with metallic mirror thickness $d_{\mathrm{M}} = 39$ nm. Peak values for transmittance remain high and almost constant until the coalescence thickness, which is well beyond the wavelength (in this case, $d_{\mathrm{co}} = 2.1$ μm). Besides, the maximum observed transmittance decreases slowly for $d > d_{\mathrm{co}}$, staying above 10% until $d = 2.9$ μm, exceeding by far the theoretical penetration depth of light into the air gap, calculated to be $\delta = [k''_{L,\perp}(\theta_{\mathrm{co}})]^{-1} \approx 652\,\mathrm{nm}$ in this situation, with $\theta_{\mathrm{co}} \approx 42.4°$.

The resonance conditions derived from equation (1) are clearly highly dependent on r_{LMH}, whose modulus and phase appear in the arguments of the hyperbolic sine and sine, respectively (with the former determining CSP resonances and the latter defining FP resonances). This three-layer reflection coefficient can be expressed in terms of Fresnel coefficients r_{ML}, r_{MH}:

$$r_{\mathrm{LMH}} = \frac{-r_{ML} + r_{MH}\, e^{i\phi_M}}{1 - r_{ML}\, r_{MH}\, e^{i\phi_M}} \quad \text{with}$$
$$\phi_M = 2\, d_M \cdot k_{M\perp} = 2\, d_M \cdot k_0 \sqrt{\varepsilon_M - \varepsilon_H \sin^2\theta}, \tag{2}$$

where ε_i stands for the permittivity of medium i and $k_{\mathrm{M}\perp}$ denotes the component of the wavevector inside the metal normal to the interfaces. The modulus and phase of r_{LMH} are plotted in Figure 2c and 2d as functions of the incident angle θ, together with those for Fresnel coefficient r_{LM}, corresponding to the limiting case where the delimiting metallic mirrors are semi-infinite ($d_M \to \infty$). The modulus $|r_{\mathrm{LMH}}|$ is seen to behave differently below and above θ_{cr}. When harmonic propagation inside the cavity is allowed, it stays slightly below 1. In turn, it grows far greater than 1 above θ_{cr}, reaching its maximum value around θ_{co}. Separation from θ_{co} is due to the role played by the rest of the argument of the hyperbolic sine in equation (1). The thicker the metallic mirrors, the higher the peak in $|r_{\mathrm{LMH}}|$, converging to $|r_{\mathrm{LM}}|$ for large d_{M}. For $\theta > \theta_{\mathrm{cr}}$, $|r_{\mathrm{LMH}}|$ and $|r_{\mathrm{LM}}|$ being greater than 1 means no problem in terms of energy conservation, since the amplitudes related by them correspond to the electric fields of evanescent waves [9]. Focusing on the phase, it grows gradually for $\theta < \theta_{\mathrm{cr}}$, reaching π at $\theta = \theta_{\mathrm{cr}}$. For larger angles, it starts decreasing slowly until θ approaches θ_{co}, when it suddenly drops to nearly zero, undergoing a slow decline again at the end. Finally, comparison with r_{LM} reveals that the drop is steeper for larger d_{M}. This anomalous behavior of r_{LMH} above θ_{cr} – large modulus and step-like phase – happens especially at the same angles $\theta \approx \theta_{\mathrm{co}}$ at which outstanding transmittance occurs.

4 Conclusions

The experimentally validated analytical formula for T through an MDM structure predicts high and constant transmittance for cavity thicknesses far above the penetration depth of light. This outstanding transmittance is key for the use of CSP-based arrangement in devices for different applications. These include sensing thin media, for which measurements based on light transmission instead of reflection can simplify experimental setups; highly selective optical filtering, spectroscopy or light control at the nanoscale.

Funding

This research was funded through projects by the Spanish Ministry of Science, Innovation and Universities (PID2024-156552OA-I00), Universidade de Santiago de Compostela (USC 2024-PU031) and Xunta de Galicia (GRC ED431C 2024/06), respectively. Finally, AD thanks the Spanish Ministry of Science, Innovation and Universities for the financial support through FPU21/01302, as well as YA acknowledges Xunta de Galicia for the postdoctoral fellowship ED481D-2024-001.

Conflicts of interest

The authors have nothing to disclose. They certify that they have no financial conflicts of interest (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) in connection with this article.

Data availability statement

Data associated with this article is available under request.

Author contribution statement

Conceptualization, R.F.; Methodology, Y.A.; Software, A.D., Y.A.; Validation, A.D.; Formal Analysis, R.F.; Investigation, A.D., Y.A., L.S., R.F.; Data Curation, Y.A., L.S.; Writing – Original Draft Preparation, A.D.; Writing – Review & Editing, R.F, Y.A.; Visualization, Y.A., A.D.; Supervision, R.F.; Project Administration & Funding Acquisition, Y.A., R.F.

References

1 Economou EN, Surface plasmons in thing films, Phys. Rev. 182, 2 (1969). https://doi.org/10.1103/PhysRev.182.539.
2 Welford KR, Sambles JR, Coupled surface plasmons in a symmetric system, J. Mod. Opt. **35**, 9 (1988). https://doi.org/10.1080/09500348814551611.
3 Marquis CD et al., Excitation of "forbidden" guided-wave plasmon polariton modes via direct reflectance using a low refractive index polymer coupling layer, PLoS One **17**, 10 (2022). https://doi.org/10.1371/journal.pone.0276522.
4 Yoon YT, Lee SS, Transmission type color filter incorporating a silver film based etalon, Opt. Exp. 18, 5 (2010). https://doi.org/10.1364/OE.18.005344.
5 Huang BR et al., Reduction of angular dip width of surface plasmon resonance sensor by coupling surface plasma waves on sensing surface and inside metal–dielectric–metal structure, J. Vacuum Sci. Technol. A. 31, 6 (2013). https://doi.org/10.1116/1.4821505.
6 Zhang H et al., Snapshot computational spectroscopy enabled by deep learining, Nanophotonics. 13, 22 (2024). https://doi.org/10.1515/nanoph-2024-0328.
7 Doval A, Arosa Y, de la Fuente R, Coupled surface plasmons and resonant optical tunnelling in symmetric optical microcavities, Opt. Laser Technol. 192, B (2025). https://doi.org/10.1016/j.optlastec.2025.113602.
8 Balili RB, Transfer matrix method in nanophotonics, Int. J. Mod. Phys. Conf. Ser. 17 (2012). https://doi.org/10.1142/S2010194512008057.
9 Kalkal Y, Kumar V, Understanding energy propagation during reflection of an evanescent electromagnetic wave, Am. J. Phys. 89, 9 (2021). https://doi.org/10.1119/10.0004834.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 39
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026038
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Numerical modeling of InGaAs/AlGaAs heterostructure for short-wave infrared nBn photodetectors

Çağrı Tok[1,2,*], Mert Satılmış[1], Habibe Keleş[1,5], Fikri Oğuz[1], Hüseyin Sarı[2], and Ekmel Özbay[1,3,4]

[1] Nanotechnology Research Center, Bilkent University, 06800 Ankara, Turkey
[2] Department of Physics Engineering, Ankara University, 06560 Ankara, Turkey
[3] Department of Physics, Bilkent University, 06800 Ankara, Turkey
[4] Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Turkey
[5] Department of Materials Science and Nanotechnology Engineering, TOBB University of Economics and Technology, Ankara 06560, Turkey

Received 29 December 2025 / Accepted 8 April 2026

Abstract. nBn infrared photodetectors have emerged as a strong alternative to conventional pn and pin-based structures due to their low dark current, fast response, and suppression of Shockley–Read–Hall (SRH) generation pathways. In this study, an InP-based InGaAs/AlGaAs/InGaAs nBn heterostructure was designed and numerically analyzed using SILVACO TCAD at 300 K to investigate its electro-optical behavior under various bias conditions. The effects of barrier thickness and band alignment engineering were systematically evaluated with a particular focus on minimizing the effective valence band offset and blocking majority-carrier leakage without hindering photocarrier transport. The simulated band diagrams confirm that the AlGaAs barrier layer enables efficient majority-carrier suppression while maintaining low-resistance conduction paths for photogenerated carriers. As a result, the proposed device exhibits low dark current and high responsivity performance comparable to planar InGaAs structures even under low bias operation. These results highlight the significance of optimized barrier design in achieving high detectivity without relying on complex fabrication routes. Overall, the presented findings demonstrate the potential of tailored nBn architectures for next-generation short-wave infrared imaging, low-photon-flux sensing, and high-dynamic-range optoelectronic systems.

Keywords: nBn InGaAs, nBn SWIR, Low noise, SILVACO TCAD, Low dark-current, InGaAs/AlGaAs heterostructure.

1 Introduction

Short-wavelength infrared (SWIR) photodetectors have become pivotal components in modern optoelectronic systems, finding applications in defense imaging, medical diagnostics, environmental monitoring, and space-based instrumentation due to their high quantum efficiency and broad spectral coverage [1–3]. Among these, InGaAs-based detectors stand out for their low noise, lattice compatibility with InP, and high detectivity across the 0.9–1.7 μm range. However, conventional pn InGaAs photodiodes increasingly fall short in meeting the demands of next-generation imaging systems that require ultra-low dark current and stable operation at room temperature. In mesa-structured geometries, surface leakage and edge recombination effects dominate, while in quadrant-type devices with large active areas, dark current rises sharply, limiting the achievable signal-to-noise ratio [4, 5].

To overcome these limitations, barrier-type architectures especially the nBn configuration have attracted significant research attention [6–8]. The nBn concept, consisting of an n-type absorber, a wide-bandgap barrier, and an n-type contact layer, aims to block majority carrier (electron) flow while allowing minority carrier (hole) transport with minimal resistance. This configuration effectively suppresses surface and generation–recombination (G-R) dark currents without introducing a depletion region in the absorber [9]. A key design challenge, however, lies in the selection of an appropriate barrier material. Recent studies have explored a variety of barrier layers, including AlGaAs, InAlGaAs AlGaAsSb, and GaAsSb compounds, yet conflicting results have been reported regarding their valence-band alignment and hole transport efficiency [10–13]. These discrepancies underline the need for further investigation into band alignment optimization for nBn structures operating in the SWIR range.

In this study, we design and simulate an nBn InGaAs SWIR photodetector incorporating a lightly n-doped

* Corresponding author: cagri.tok@bilkent.edu.tr

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Physical models used in SILVACO ATLAS simulation for nBn InGaAs photodetector design.

Model	Syntax	Description
Mobility	CVT	To enable transverse field, doping, and temperature-dependent parts of mobility
Optical	OPTR	To enable band-to-band recombination for direct band semiconductors
Shockley–Read–Hall	SRH	To enable recombination. It uses fixed minority carrier lifetimes
AUGER recombination	Auger	To enable direct transition of three carriers
Fermi distribution	Fermi	To enable carrier statistics. Suitable for highly doped regions
Quantum tunneling	QTUNN	To enable quantum tunneling through the conduction and valence band barrier due to a semiconductor
Quantum tunneling Band to Band	QTUN.BBT	To enables the band-to-band mode of the direct quantum tunneling model
Non-local trap assisted tunneling	TAT. NONLOCAL	To enables the non-local tunneling model in the calculation of the field effect enhancement factors.
Impact ionization	IMPACT SELB	To enables the impact ionization. Recommended for most case, includes temperature dependent parameters.

$Al_{0.57}Ga_{0.43}As$ barrier layer, which is hypothesized to provide near-flat valence-band alignment while maintaining sufficient conduction-band offset to block electron transport. Numerical simulations are performed using SILVACO ATLAS TCAD at 300 K, enabling accurate modeling of band profiles, carrier distributions, and current transport mechanisms. The optimized device exhibits a significant reduction in dark current compared to conventional pn InGaAs structures while preserving responsivity, indicating that proper band alignment through barrier engineering can lead to improved detectivity and thermal stability in SWIR photodetectors.

2 Simulation and methodology

A detailed computational framework was developed to investigate the electrical transport and band-alignment characteristics of the proposed nBn-InGaAs photodetector. Simulations were performed using the SILVACO ATLAS, employing the physical models listed in Table 1 and the material parameters given in Table 2 [14–17].

In the simulation, the CVT mobility model was employed to account for doping, temperature, and transverse field effects on carrier mobility. Optical recombination was included via the OPTR model, while SRH and Auger models were used to describe trap-assisted and high-injection recombination processes, respectively. Fermi statistics were considered to ensure accurate carrier distribution in highly doped regions. Additionally, QTUNN and QTUN.BBT models were activated to capture quantum tunneling effects, including band-to-band tunneling across the barrier layer. To further improve the accuracy of dark current analysis, the TAT.NONLOCAL model was incorporated to account for non-local trap-assisted tunneling mechanisms, and the IMPACT SELB model was included to consider impact ionization effects with temperature dependence.

Vertical device stack (from substrate up) was modeled as InP (n^+ contact) / InGaAs (absorber) / $Al_xGa_{1-x}As$ (barrier) / InGaAs (n^- contact) shown in Figure 1. Owing to its low lattice mismatch with the InGaAs and the advancements in epitaxial growth techniques, the AlGaAs barrier has emerged as the most suitable barrier material for InGaAs-based nBn photodetectors [18]. The reduced lattice strain at the heterointerface ensures high crystalline quality and stable electronic properties, facilitating efficient carrier transport while effectively suppressing dark current.

During the design of the nBn structure, systematic simulations were conducted to assess how different Al–Ga molar fractions ($Al_xGa_{1-x}As$, with x ranging from 0.50 to 0.60) influence the electronic band alignment within the barrier layer. This compositional tuning was particularly crucial, as the barrier must balance two competing requirements: minimizing the valence band offset (VBO) to ensure unhindered hole transport, while simultaneously maintaining a sufficiently large conduction band offset (CBO) to suppress the flow of photogenerated electrons. Achieving this balance is essential for reducing dark current without compromising photocurrent extraction. The simulation results indicate that an Al composition of $x = 0.57$ provides the most favorable trade-off among the examined cases. At this composition, the VBO reaches its minimum value ($\Delta E_v \approx 35$ meV), effectively eliminating the barrier for holes, whereas the CBO ($\Delta E_c \approx 1.26$ eV) remains sufficiently high to block majority-carrier leakage. Figure 2a illustrates the equilibrium band diagrams for the tested molar fractions, clearly demonstrating how incremental changes in Al content reshape both conduction and valence band edges. This optimized band configuration is particularly advantageous for nBn photodetectors, as it ensures that photogenerated minority carriers contribute efficiently to the photocurrent while majority carriers are strongly suppressed, ultimately leading to enhanced detectivity and improved bias stability.

In Figure 2b, the 150 mV reverse bias confines Fermi level band bending within the large bandgap AlGaAs barrier, leaving the InGaAs absorption layer largely undepleted and thus effectively mitigating G-R dark current, in line

Table 2. Material parameters used in nBn InGaAs photodetector design.

Parameter	InP	InGaAs	AlGaAs
Bandgap	1.35 eV	0.734 eV	1.99 eV
Electron affinity	4.40 eV	4.67 eV	3.41 eV
Permittivity	12.5	13.9	11.5
Electron SRH lifetime		1 ns	
Hole SRH lifetime		20 ns	
Trap density		1×10^{14} cm^{-3}	
Electron Auger coefficient	3.7×10^{-31} cm^6/s	1.8×10^{-28} cm^6/s	5×10^{-30} cm^6/s
Hole Auger coefficient	8.7×10^{-30} cm^6/s	1.8×10^{-28} cm^6/s	1×10^{-31} cm^6/s
Electron effective mass	0.0795 m_0	0.0332 m_0	0.77 m_0
Hole effective mass	0.6 m_0	0.471 m_0	0.734 m_0

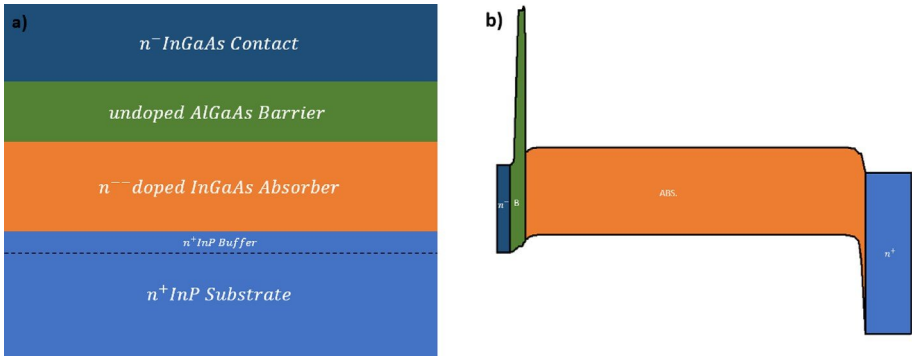


Figure 1. (a) Planar design, and (b) band diagram of nBn structure.

with the core design principles of nBn photodetectors [19]. Therefore, an Al composition of 0.57 was adopted for subsequent electrical and optical simulations to ensure optimal device performance.

Following the determination of the optimal Al composition ratio for the nBn InGaAs structure, the influence of varying N-type doping concentrations within the AlGaAs barrier layer was examined Figure 3 presents the corresponding energy band diagrams under a 150 mV reverse bias for different doping levels. As simulated, increasing the N-type doping concentration leads to a higher VBO, whereas the undoped AlGaAs barrier exhibits the most favorable band alignment. Nevertheless, to ensure that the simulation results accurately reflect realistic device behavior and considering that background doping cannot be completely avoided during epitaxial growth the electrical simulations were performed assuming a maximum donor concentration of 1×10^{16} cm^{-3} in the AlGaAs barrier. This consideration ensures a closer correspondence between the simulated and experimentally achievable device performance [20].

Based on the optimized band alignment parameters, the device structure was defined as follows: 0.2 µm n$^-$-$In_{0.53}Ga_{0.47}As$ top n-contact layer (5×10^{17} cm^{-3}), 0.1 µm n$^-$-$Al_{0.57}Ga_{0.43}As$ barrier layer (1×10^{16} cm^{-3}), 5 µm n$^-$-$In_{0.53}Ga_{0.47}As$ absorber layer (5×10^{15} cm^{-3}), 0.5 µm n$^+$-InP bottom buffer n-contact layer (5×10^{18} cm^{-3}) and 500 µm n$^+$-InP substrate (5×10^{18} cm^{-3}). The barrier thickness was determined according to the photocurrent versus potential curves presented in Figure 4 of Section 3. In addition, the absorber layer was designed to be relatively thick to achieve low capacitance, and lightly n-doped to further reduce the dark current [21].

3 Results and discussion

In this section, all simulations were conducted at 300 K, with the lateral device active area set to 300 µm to reflect the geometry of commercial detectors, and all current values were normalized accordingly. Figure 4a shows the simulated current as a function of applied bias for different barrier thicknesses. Increasing the barrier thickness clearly reduces the dark current, but at the same time causes a relative decrease in photocurrent. Detailed analysis revealed that a barrier thickness of 100 nm provides the optimal balance: it maintains photocurrent levels comparable to those of thinner barriers while yielding nearly an order-of-magnitude reduction in dark current. In contrast, barrier layers thicker than 100 nm introduce a more pronounced photocurrent degradation, indicating that additional thickness imposes unnecessary transport resistance without a

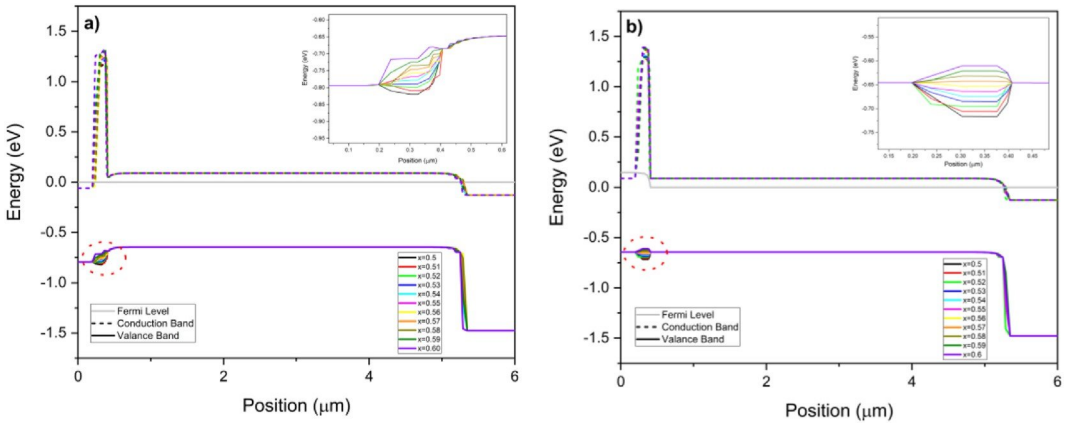


Figure 2. Simulated energy band diagrams of the nBn InGaAs structure for different Al compositions: (a) under equilibrium conditions and (b) under −150 mV reverse bias voltage.

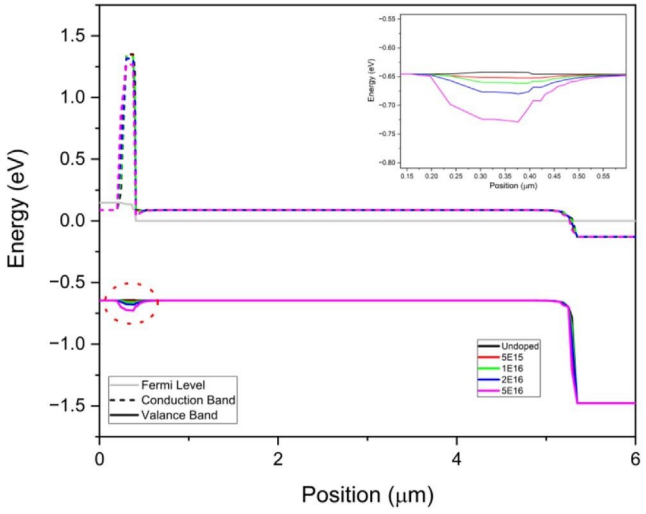


Figure 3. Energy band diagrams of the nBn InGaAs structure under a 150 mV reverse bias for different N-type doping concentrations in the AlGaAs barrier.

commensurate benefit in dark-current suppression. Thus, to optimize the trade-off between dark-current reduction and photo-response, the barrier thickness in the nBn-InGaAs structure was selected as 100 nm [6]. In Figure 4b, a sensitivity analysis was performed to evaluate the impact of band alignment uncertainties by varying the Al composition in the $Al_xGa_{1-x}As$ (x = 0.55–0.59) 100 nm barrier. The results show that variations in Al content do not produce significant changes in dark current characteristics, indicating a weak dependence of device performance on moderate fluctuations in CBO and VBO. These findings confirm the robustness of the proposed nBn-InGaAs design against variations in material composition and the associated band offset uncertainties.

As illustrated in Figure 4a, the nBn-InGaAs structure exhibits a significantly lower dark current compared to conventional mesa and planar InGaAs photodiodes [22, 23], while maintaining nearly the same photocurrent magnitude. Furthermore, the photocurrent characteristics presented in Figure 4a were obtained under back-illuminated operation with an incident optical power density of 0.1 W/cm^2.

Figure 5 presents the decomposition of the dark current components as a function of reverse bias, providing insight into the dominant transport mechanisms in the proposed device. The total dark current (SUM) increases from approximately 3×10^{-14} A at 0 V to about 1×10^{-10} A at 1 V, exhibiting a gradual bias-dependent increase. Among the individual components, the SRH generation current dominates in the low-to-moderate bias range, indicating that generation–recombination processes are the primary contributor to the dark current [24, 25].

In contrast, the tunneling current increases with reverse bias and becomes comparable to the SRH current at higher voltages, while remaining slightly lower overall, confirming that field-assisted leakage is effectively controlled by the barrier design. The diffusion current is nearly bias-independent and negligible in magnitude, suggesting minimal thermally driven carrier injection. Overall, these results demonstrate that the AlGaAs barrier effectively suppresses diffusion-related leakage, while the dark current is primarily governed by SRH generation at low bias and increasingly influenced by tunneling mechanisms at higher bias [26, 27]. These results confirm that the optimized barrier design effectively minimizes leakage pathways without compromising carrier collection efficiency, demonstrating the superior dark-current suppression capability of the nBn structure.

Following the analysis of the photocurrent characteristics, the simulated nBn-InGaAs structure was further evaluated in terms of its breakdown voltage and capacitance behavior, as presented in Figure 6. The results in Figure 6a show that the diode enters breakdown at approximately −82.5 V, primarily due to the combined effect of the thick absorber and barrier layers. This relatively high breakdown voltage highlights the structural robustness of the nBn

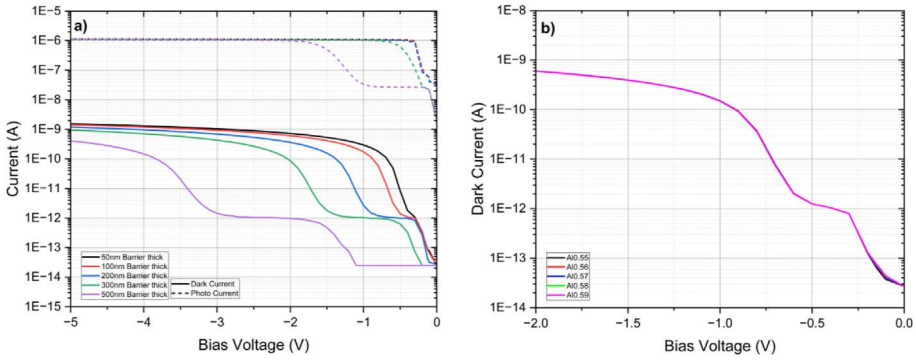


Figure 4. (a) Dark and photocurrent of the nBn-InGaAs structure with various barrier thickness, (b) dark current graph with different Al ratio.

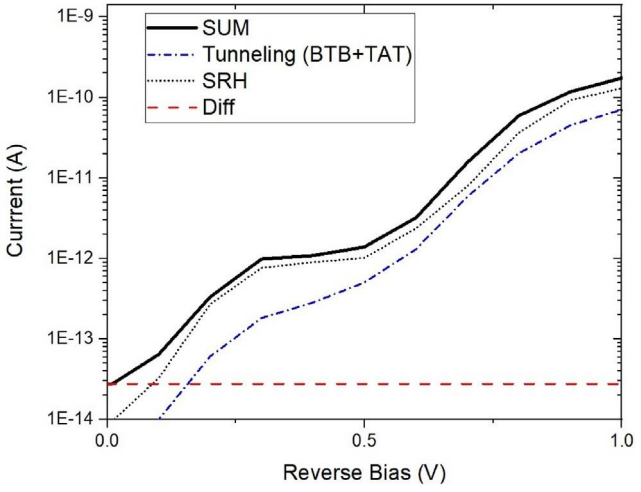


Figure 5. Dark current components as a function of reverse bias voltage at 300 K.

configuration under strong reverse bias conditions and underscores its superior reliability compared to conventional pin InGaAs photodiodes [28], particularly in high-voltage and wide-dynamic-range applications. For comparison, conventional InGaAs pn photodiodes typically exhibit relatively low breakdown voltages due to the strong electric field concentration within the depletion region, which promotes early onset of avalanche multiplication through Impact Ionization. In such structures, the high doping gradients result in a narrow depletion width and consequently a large peak electric field, leading to breakdown at relatively low reverse biases. In contrast, nBn architectures eliminate the pn junction and instead rely on a unipolar barrier to control carrier transport. As a result, the electric field is more uniformly distributed across the device, significantly reducing the peak field intensity [29, 30].

Figure 6b shows that the device exhibits a capacitance of approximately 9.2 pF at −5 V bias and 1 kHz frequency, which can be attributed to the presence of the thick absorber layer. Such a low capacitance not only minimizes noise but also enhances temporal response and dynamic range, resulting in improved signal fidelity for high-speed and low-noise infrared detection [31]. These findings demonstrate that, beyond its dark-current suppression capability, the nBn-InGaAs architecture combines electrical stability with excellent frequency response, confirming its potential for next-generation infrared photodetector technologies.

After evaluating the electrical characteristics, the photoresponsivity performance of the device was examined. Prior to responsivity analysis, the optical reflection behavior of the InP substrate was characterized, revealing that approximately 30% of the incident short-wavelength radiation is reflected from its surface. To enhance photon-to-electron conversion efficiency and improve overall responsivity, an anti-reflection (AR) coating was designed based on the reflectance formulation [32] equation (1). Figure 7a compares the reflectance spectra of a bare InP surface with structures incorporating Si_3N_4 (140 nm) and SiO_2 (150 nm) AR layers in the wavelength range of 0.8–1.7 μm. As illustrated, the AR-coated configuration significantly suppresses surface reflection, reducing it from nearly 30% to approximately 0.02% at 1.55 μm, which is expected to yield improved photocarrier generation and enhanced detector responsivity. Si_3N_4 and SiO_2 were selected as AR materials due to their low extinction coefficients in the sub-2 μm band, making them well-suited for near-infrared photodetector applications [33].

$$r_1 = \frac{n_0 - n_1}{n_0 + n_1} \tag{1}$$

$$r_2 = \frac{n_1 - n_2}{n_1 + n_2} \tag{2}$$

$$r_3 = \frac{n_2 - n_3}{n_2 + n_3} \tag{3}$$

$$\theta_1 = \frac{2\pi n_1 t_1}{\lambda} \tag{4}$$

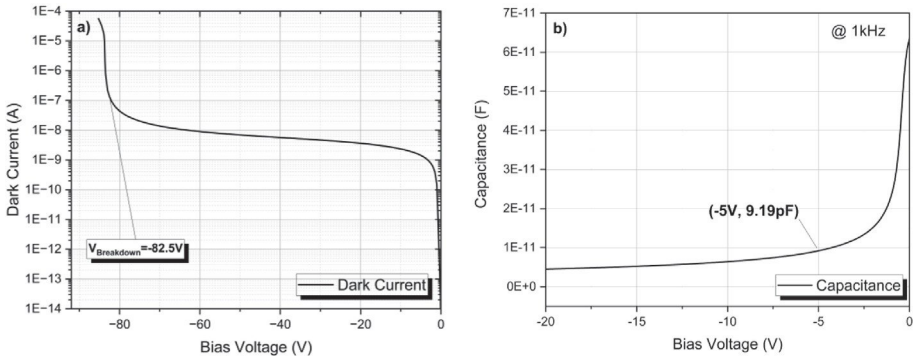


Figure 6. (a) Simulated I–V characteristics of the nBn-InGaAs structure and (b) C–V characteristics at 1 kHz.

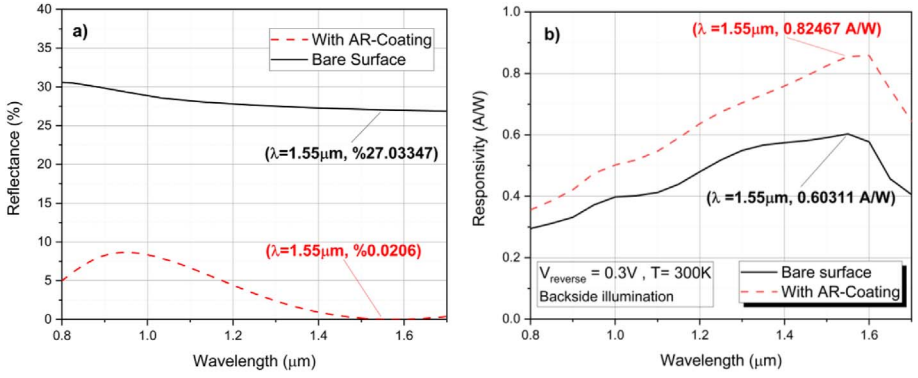


Figure 7. (a) Reflection and (b) responsivity graphs of the bare and AR-coated surfaces as a function of wavelength.

$$\theta_2 = \frac{2\pi n_2 t_2}{\lambda} \tag{5}$$

See equation (6) at the bottom of the page.

To facilitate the definition of the reflectance expression in equation (6), the intermediate relations provided in equations (1)–(5) were utilized. Where n_0 is denotes the refractive index of the incident medium, n_1 the refractive index of the Si_3N_4 layer, n_2 the refractive index of the SiO_2 layer, and n_3 the refractive index of the InP substrate. Similarly, t_1 and t_2 represent the physical thicknesses of the Si_3N_4 and SiO_2 layers, respectively, while λ corresponds to the wavelength of the incident radiation.

In Figure 7b, the wavelength-dependent responsivity characteristics of the nBn-InGaAs photodetector are presented. To clearly demonstrate the influence of the anti-reflection coating on photo-response, the responsivity curves of both the AR-coated device and the bare surface are shown together. As observed, the application of the AR layer yields a substantial enhancement in photocarrier

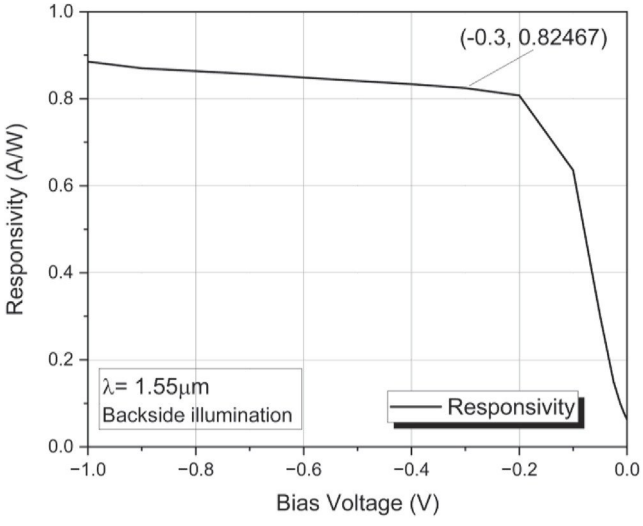


Figure 8. Bias voltage-dependent responsivity graph of nBn-InGaAs structure.

$$R = \frac{\begin{array}{c} r_{12} + r_{22} + r_{32} + r_{12}r_{22}r_{32} + 2r_1r_2\left(1 + r_{32}\right)cos(2\theta_1) + 2r_2r_3\left(1 + r_{12}\right)cos(2\theta_2) + 2r_1r_2r_3cos(2(\theta_1 + \theta_2)) \\ +2r_1r_2r_3cos(2(\theta_1 - \theta_2)) \end{array}}{\begin{array}{c} 1 + r_{12}r_{22} + r_{22}r_{32} + r_{12}r_{22}r_{32} + 2r_1r_2\left(1 + r_{32}\right)cos(2\theta_1) + 2r_2r_3\left(1 + r_{12}\right)cos(2\theta_2) + 2r_1r_2r_3cos(2(\theta_1 + \theta_2)) \\ +2r_1r_2r_3cos(2(\theta_1 - \theta_2)) \end{array}} \tag{6}$$

Table 3. Performance comparison of the proposed nBn photodetector with reported InGaAs-based photodetector.

Device	Dark current	Responsivity	EQE	Ref
nBn-InGaAs	~1 pA ($V_r = 0.3$ V, $\phi = 300$ μm)	0.824 A/W ($\lambda = 1.55$ μm, $V_r = 0.3$ V)	–	This work
nBn-InGaAs	~8 nA ($V_r = 0.3$ V, $\phi = 300$ μm)	–	% 45 ($\lambda = 1.55$ μm, $V_r = 0.25$ V)	[15]
nBn-InGaAs	~0.9 nA ($V_r = 0.3$ V, $\phi = 300$ μm)		% 80 ($\lambda = 1.55$ μm, $V_r = 0.2$ V)	[6]
pBin-InGaAs	~70 pA ($V_r = 0.3$ V, $\phi = 500$ μm)	0.62 A/W ($\lambda = 1.55$ μm, $V_r = 0.2$ V)	–	[36]
pn-InGaAs	~10 pA ($V_r = 0.3$ V, $\phi = 300$ μm)	1.08 A/W ($\lambda = 1.55$ μm)	% 86 ($\lambda = 1.55$ μm)	[37]
pin-InGaAs	~41 pA ($V_r = 5$ V, $\phi = 73$ μm)	0.99 A/W ($\lambda = 1.55$ μm, $V_r = 5$ V)	% 79 ($\lambda = 1.55$ μm)	[38]

extraction efficiency: at the telecommunication wavelength of 1.55 μm [34], the responsivity increases by approximately 36%, improving from 0.60 to 0.82 A/W. This improvement is consistent with the reflectance reduction achieved through AR design and confirms that minimizing front-surface optical losses directly enhances the detector's photon-to-electron conversion capability. In the numerical simulations performed in this study, the TCAD software provides external quantum efficiency (EQE) rather than responsivity as a direct output. Therefore, the EQE values were converted to responsivity using equation (7), enabling a direct comparison between the simulated and AR-enhanced device performance [35]. It should also be noted that the reported responsivity values correspond to conditions in which the nBn-InGaAs structure is illuminated from the backside under an incident optical power density of 0.01 W/cm^2. Backside illumination ensures that the incident photons efficiently reach the absorbing InGaAs layer without encountering additional front-side metallization losses, thereby allowing a more accurate assessment of the intrinsic photo-response. Collectively, these results verify that integrating an appropriately engineered AR coating is an effective approach for maximizing optical coupling and elevating responsivity in nBn-type infrared photodetectors.

$$R = \eta \frac{q}{hf} \tag{7}$$

where η is the EQE of the detector for a given wavelength, q is the electron charge, f is the frequency of the incident photons, and h is the Planck constant.

Figure 8 shows the responsivity variation with applied bias voltage, indicating that the proposed nBn design maintains high responsivity even at very low operating bias due to the minimized VBO and optimized barrier engineering. Notably, its performance approaches that of cutting-edge planar InGaAs detectors (Table 3), confirming the viability of the structure as a strong candidate for high-performance, low-power infrared photodetection.

To benchmark the proposed device, a comparison with previously reported InGaAs photodetectors is summarized in Table 3. The proposed nBn structure exhibits an ultra-low dark current of 1 pA at −0.3 V, which is significantly lower than reported nBn-InGaAs devices (0.9–8 nA) under similar bias conditions, corresponding to an improvement of several orders of magnitude. Compared to conventional pn junction, PIN diode, and pBin structures, which typically show dark currents in the range of 10–70 pA, the proposed device still demonstrates superior suppression.

The responsivity of 0.824 A/W at 1.55 μm is comparable to literature values, although slightly lower than the 1.08 A/W reported for pn-based devices. Overall, the results indicate that the proposed nBn design achieves an excellent trade-off between ultra-low dark current and competitive optical performance.

4 Conclusion

In this work, an InGaAs-based nBn photodetector incorporating an $Al_{0.57}Ga_{0.43}As$ barrier layer was systematically designed and numerically analyzed using SILVACO ATLAS. Through detailed band alignment engineering, the optimal barrier composition was identified to achieve a near-zero valence band offset ($\Delta E_v \approx 35$ meV) while maintaining a sufficiently large conduction band offset ($\Delta E_? \approx 1.26$ eV). This configuration effectively enables barrier-free transport of minority carriers while strongly suppressing majority-carrier leakage, addressing one of the fundamental limitations of conventional InGaAs photodiodes. The optimized device demonstrates ultra-low dark current on the order of ~1 pA at low reverse bias (−0.3 V), representing orders-of-magnitude improvement compared to previously reported nBn-InGaAs structures. This significant reduction is primarily attributed to the suppression of Shockley–Read–Hall generation and diffusion-related leakage mechanisms, as confirmed by current component analysis. At higher biases, the controlled contribution of tunneling current further verifies the effectiveness of the barrier design in limiting field-assisted leakage. In addition to its low-noise performance, the proposed structure maintains competitive optical characteristics. By inte-

grating a Si_3N_4/SiO_2 anti-reflection coating, surface reflectance was reduced to nearly zero at 1.55 μm, resulting in a responsivity enhancement of approximately 36% (from 0.60 to 0.82 A/W). The device also exhibits low capacitance (∼9.2 pF at −5 V) and a high breakdown voltage (−82.5 V), indicating its suitability for high-speed, low-noise, and wide dynamic range applications. Overall, this study demonstrates that precise control of barrier composition, doping, and layer thickness enables simultaneous optimization of electrical and optical performance in nBn architectures [39]. The presented design achieves an excellent trade-off between ultra-low dark current and high responsivity under low-bias operation, making it a strong candidate for next-generation short-wave infrared photodetectors operating under low-photon-flux and high-sensitivity conditions.

Funding

This research received no external funding.

Conflicts of interest

The authors declare that there are no conflicts of interest related to this article.

Data availability statement

The data is available from the corresponding author under reasonable request.

Author contribution statement

Ç. Tok was responsible for the conceptualization, design, simulation, formal analysis, investigation, methodology, resources, data curation, visualization, and writing of the original draft, as well as the review and editing of the manuscript.

M. Satılmış focused on the visualization, resources, methodology, investigation, and critical review of the manuscript.

H. Keleş handled the data curation, methodology, investigation, and re-editing of the manuscript.

F. Oğuz provided conceptualization, supervision, and contributed to the re-editing and design.

H. Sarı provided supervision and contributed to the conceptual framework.

E. Özbay oversaw the methodology, conceptual framework, and final manuscript approval.

References

1 Rogalski A, Infrared detectors: an overview, *Infrared Phys. Technol.* **43**.3–5, 187–210 (2002). https://doi.org/10.1016/S1350-4495(02)00140-8.
2 Thimsen E, Sadtler B, Berezin MY, Shortwave-infrared (SWIR) emitters for biological imaging: a review of challenges and opportunities, *Nanophotonics* **6**(5), 1043–1054 (2017). https://doi.org/10.1515/nanoph-2017-0039.
3 Schindler K, Wolf J, Krabbe A, Characterization of InGaAs-based cameras for astronomical applications using a new VIS-NIR-SWIR detector test bench, in *Ground-based and Airborne Telescopes V*, **Vol. 9145** (SPIE, 2014). https://doi.org/10.1117/12.2057052.
4 Satılmış M, et al. Investigation of sin x: H surface passivation impact on inasp/ingaas e-swir photodiodes, *IEEE Sensors J* **24**(19), 29927–29936 (2024). https://doi.org/10.1109/JSEN.2024.3443747.
5 Datta S, Joshi A, Rue J. Large-area InGaAs quad photoreceiver for laser interferometry space antenna, in *Nanophotonics and Macrophotonics for Space Environments IV*, **Vol. 7817** (SPIE, 2010). https://doi.org/10.1117/12.861400.
6 Uzgur F, Kocaman S, InGaAs nBn SWIR detector design with lattice-matched InAlGaAs barrier, *Turk. J. Elect. Eng. Comput. Sci.* **27**(1), 1–10 (2019). https://doi.org/10.3906/elk-1802-197.
7 Kopytko M., et al., Engineering the bandgap of unipolar HgCdTe-based nBn infrared photodetectors, *J. Elect. Mater.* **44** (1), 158–166 (2015). https://doi.org/10.1007/s11664-014-3511-9.
8 Shi Q, et al., Progress on nBn infrared detectors, *J. Infrared Millim. Waves* **41**, 1 (2022). https://doi.org/10.11972/j.issn.1001-9014.2022.01.010.
9 Walker AW, Denhoff MW, Minority carrier diffusion lengths and mobilities in low-doped n-InGaAs for focal plane array applications, in *Infrared Technology and Applications XLIII*, **Vol. 10177** (SPIE, 2017). https://doi.org/10.1117/12.2258616.
10 Tok Ç, et al., Design and simulation of SWIR nBn-InGaAs photodetector with AlGaAs barrier, in *EPJ Web of Conferences*, **Vol. 335** (EDP Sciences, 2025). https://doi.org/10.1051/epjconf/202533511002.
11 Şahin A, et al., A numerical design for SWIR/eSWIR dual-band operation with InGaAs nBn structures, in *Quantum Sensing and Nano Electronics and Photonics XVIII*, **Vol. 12009** (SPIE, 2022). https://doi.org/10.1117/12.2610186.
12 Liang Y, et al., InP-based GaAsSb/AlGaAsSb/T2SL barrier-type low-bias tunable dual-band NIR/eSWIR photodetectors, *Opt. Express* **32**(13), 23822–23830 (2024). https://doi.org/10.1364/oe.528762.
13 Xie R, et al., Extended short-wave infrared barrier structure focal plane array based on InGaAs/GaAsSb type-ii superlattices, *J. Lightwave Technol.* **43**(5), 2271–2276 (2025). https://doi.org/10.1109/JLT.2024.3491901.
14 Silvaco I, *ATLAS User's Manual: A 2D Numerical Device Simulator* (2016).
15 Şahin A, et al., Dual-band InGaAs nBn photodetectors at 2 μm, *Appl. Phys. Lett.*, **120**, 9 (2022). https://doi.org/10.1063/5.0080950.
16 Khamis MA, et al., Effect of multiplication and absorption layers width on avalanche multiplication gain in InGaAs/InP avalanche photodiode, *Int. J. Eng. Technol.* **7.4.35**, 559–563 (2018).
17 Salem MS, et al., Performance optimization of the InGaP/GaAs dual-junction solar cell using SILVACO TCAD, *Int. J. Photoenergy* **(2021)** (1), 8842975 2021. https://doi.org/10.1155/2021/8842975.
18 Liu S, et al., Investigation of the growth temperature of AlGaAs barrier layer on optical and crystal quality of InGaAs/AlGaAs multi-quantum wells and AlGaAs single layer grown by molecular beam epitaxy (MBE), *Mater. Sci. Semicond. Process.* **187**, 109140 (2025). https://doi.org/10.1016/j.mssp.2024.109140.
19 Maimon S, Wicks GW, nBn detector, an infrared detector with reduced dark current and higher operating temperature, *Appl. Phys. Lett.* **89**, 15 (2006). https://doi.org/10.1063/1.2360235.
20 He Z, et al., The effect of unintentional carbon incorporation on the electrical properties of AlGaAs grown by MOCVD, *Opt. Mater.* **108**, 110227 (2020). https://doi.org/10.1016/j.optmat.2020.110227.

21 Liu H, et al., Design and fabrication of high performance InGaAs near infrared photodetector, *Nanomaterials* **13**(21), 2895 (2023). https://doi.org/10.3390/nano13212895

22 Koronnov AA, et al., InGaAs/InP PIN photodiode for optical receivers in pulsed-laser range finding systems, *Bull. Lebedev Phys. Inst.* **52**(Suppl 1), S106–S112 (2025). https://doi.org/10.3103/S1068335625600263.

23 Andreev DS, et al., Investigation of planar photodiodes of a focal plane array based on a heteroepitaxial InGaAs/InP structure, *J. Commun. Technol. Electr.* **61**(10), 1220–1225 (2016). https://doi.org/10.1134/S1064226916100028.

24 Savich GR, et al., Dark current filtering in unipolar barrier infrared detectors, *Appl. Phys. Lett.* **99**, 12 (2011). https://doi.org/10.1063/1.3643515.

25 Savich GR, et al., Use of unipolar barriers to block dark currents in infrared detectors, in *Infrared Technology and Applications XXXVII*, **Vol. 8012** (SPIE, 2011). https://doi.org/10.1117/12.884075.

26 Gravrand O, Boulard F, Ferron A, Ballet P, Hassis W, A new nBn IR detection concept using HgCdTe material. *J. Elect. Mater.* **44**(9), 3069–3075 (2015). https://doi.org/10.1007/s11664-015-3821-6.

27 Thi L, Yen YK, Mori N, A comparison of mechanisms for improving dark current characteristics in barrier infrared photodetectors, *Jpn. J. Appl. Phys.* **59**(4), 044005 (2020). https://doi.org/10.35848/1347-4065/ab7e77.

28 Bauer JG, Trommer R, Long-term operation of planar InGaAs/InP pin photodiodes, *IEEE Trans. Electron Devices* **35**(12), 2349–2354 (2002). https://doi.org/10.1109/16.8813.

29 Zhang Q, et al., Tuning Ga_2O_3-based avalanche photodetectors performance through barrier layer thickness optimization, *Mater. Design* **239**, 112823 (2024). https://doi.org/10.1016/j.matdes.2024.112823.

30 Zhang Q, et al., Enhanced gain and detectivity of unipolar barrier solar blind avalanche photodetector via lattice and band engineering, *Nat. Commun.* **14**(1), 418 (2023). https://doi.org/10.1038/s41467-023-36117-8.

31 Bucsa I, et al., Large area, low capacitance InGaAs PINs, in *Physics and Simulation of Optoelectronic Devices XXIX*, **Vol. 11680** (SPIE, 2021). https://doi.org/10.1117/12.2578824.

32 Moys BA The theory of double-layer antireflection coatings, *Thin Solid Films* **21**(1), 145–157 (1974). https://doi.org/10.1016/0040-6090(74)90097-2.

33 Oh G, et al., Broadband antireflective coatings for high efficiency InGaP/GaAs/InGaAsP/InGaAs multi-junction solar cells, *Solar Energy Mater. Solar Cells* **207**, 110359 (2020). https://doi.org/10.1016/j.solmat.2019.110359.

34 Milostnaya I, et al., Superconducting single-photon detectors designed for operation at 1.55-μm telecommunication wavelength, *J. Phys. Conf. Ser.* **43**(1) IOP Publishing (2006). https://doi.org/10.1088/1742-6596/43/1/326.

35 Sze SM, Li Y, Kwok KNg, *Physics of Semiconductor Devices* (John Wiley & Sons, 2021).

36 Klem JF, et al., Mesa-isolated InGaAs photodetectors with low dark current, *Appl. Phys. Lett.* **95**(3), (2009) https://doi.org/10.1063/1.3184807.

37 Dolas MH, et al.. Low dark current diffusion limited planar type InGaAs photodetectors, in *Infrared Sensors, Devices, and Applications IX*, **Vol. 11129** (SPIE, 2019). https://doi.org/10.1117/12.2528666.

38 Chang S-H, et al., Ultra high performance planar InGaAs PIN photodiodes for high speed optical fiber communication, *Sens. Actuat. A Phys.* **133**(1), 9–12 (2007). https://doi.org/10.1016/j.sna.2006.04.023.

39 Martyniuk P, Rogalski A, HOT infrared photodetectors, *Opto-Elect. Rev.* **21**(2), 239–257 (2013). https://doi.org/10.2478/s11772-013-0090-x.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 18
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026010
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

Luminescent thermal history sensing potential in Pr^{3+}-activated YAG matrix

Oskar Bogucki[1,2,*], Marcin Kaczkan[1], Anna Kozłowska[2], Aleksandra Pantoł-Boczon[2,3], and Agnieszka Malinowska[2]

[1] Warsaw University of Technology, Institute of Microelectronics and Optoelectronics, Koszykowa 75, Warsaw 00-662, Poland
[2] Łukasiewicz Research Network Institute of Microelectronics and Photonics, al. Lotników 32/46, Warsaw 02-668, Poland
[3] Warsaw University of Technology, Faculty of Chemistry, al. Noakowskiego 3, Warsaw 00-664, Poland

Received 30 December 2025 / Accepted 4 February 2026

Abstract. In this study, it was demonstrated that Pr^{3+}-doped YAG powder can record thermal history in the temperature range of the template provided by 1100–1600 °C. Upon heating, the material undergoes irreversible structural phase transformations (YAM → YAP → single-phase YAG) induced by 2 h of thermal annealing, leading to permanent changes in both the emission intensity and the excited-state lifetime of Pr^{3+} ion. Temperature can be determined using two approaches: by analyzing the intensity ratio of the $^3P_0 \rightarrow {}^3H_4$ and $^1D_2 \rightarrow {}^3H_4$ emission bands, which increases linearly over 1300–1600 °C with a sensitivity of 0.001 $°C^{-1}$, and by measuring the luminescence decay time of the $^1D_2 \rightarrow {}^3H_4$ transition, which decreases linearly over 1100–1600 °C with a sensitivity of 0.38 µs $°C^{-1}$. The combination of both methods enables durable and quantitative tracking of the material's thermal history, confirming the suitability of YAG:Pr^{3+} powders for high-temperature diagnostics, thermal mapping, and monitoring of industrial processes.

Keywords: YAG:Pr^{3+}, Thermal history sensor, Luminescence, Optical thermometry.

1 Introduction

Temperature is a key thermodynamic parameter in industrial and research applications [1]. At high temperatures, conventional measurement techniques such as thermocouples, resistance sensors, or pyrometers suffer from limitations related to physical contact and restricted material durability [2–4]. Moreover, rare-earth-doped optical fibers are typically unstable above ~1100 °C, limiting their use under extreme conditions [5]. Here, luminescence-based techniques offer a viable alternative. These methods are based on the emission of RE^{3+} ions embedded in crystalline hosts such as YAG, which undergo predictable and irreversible structural transformations after 2 h of thermal annealing. These transformations permanently modify the luminescence properties, enabling the material to function as a fluorescent thermal history sensor (THS) [6]. Pr^{3+} ions are particularly attractive activators due to their specific energy structure, enabling a broad excitation range and intense visible emission [7, 8].

In the Y_2O_3–Al_2O_3 system, increasing temperature induces a sequence of phase transitions–from YAM to YAP and finally to single-phase YAG above ~1600 °C [9]. Each phase provides a distinct local crystal field for Pr^{3+} ions, modifying electronic levels, phonon interactions, and thus luminescence properties [10]. Luminescence intensity [11] and lifetime [12] are strongly determined by the local crystal structure, which evolves irreversibly depending on both the temperature and duration of thermal exposure. Therefore, these changes can be exploited to construct a fluorescent THS. Crucially, these changes originate from irreversible phase transitions, enabling permanent recording of past thermal exposure and forming the basis of THSs, which allow reconstruction of the maximum temperature experienced by a material [10, 13]. Compared to temperature-indicating paints, phosphor powders offer superior chemical stability, quantitative readout, and applicability to complex geometries [13].

In the present work, YAG:1%Pr powder was investigated as a potential THS material. The analysis comprised two detection approaches based on the fluorescence intensity ratio (FIR) as well as an analysis of luminescence decay times. The FIR can be defined either as a comparison of the same emission peak measured at different temperatures [14, 15] or, more commonly in optical thermometry, as the ratio of two distinct emission peaks measured at the same

* Corresponding author: oskar.bogucki.dokt@pw.edu.pl

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

temperature. The latter, self-referencing method, provides higher accuracy and greater robustness against fluctuations in excitation power, optical losses, and environmental conditions [11], and therefore constitutes the primary focus of the present study. In the second detection approach, the luminescence decay times of the $^1D_2 \rightarrow {}^3H_4$ transition were analyzed. This transition is characterized by a significantly longer lifetime and a stronger temperature dependence than the $^3P_0 \rightarrow {}^3H_4$ transition, which makes lifetime-based analysis a sensitive complementary tool for determining the maximum temperature reached during thermal exposure.

2 Materials and methods

The samples of yttrium-aluminum garnet doped with Pr^{3+} ions (general formula $Y_{2.97}Pr_{0.03}Al_5O_{12}$) were synthesized using a solid-state reaction assisted by high-energy milling, using high-purity commercial powders of Al_2O_3 (Krahn Ceramics, > 99.99%), Y_2O_3 (NYC YT3WP, 99.9%), and Pr_6O_{11} (Thermo Scientific Chemicals, 99.996%). Stoichiometric amounts of the starting materials were mixed and milled with anhydrous ethanol (Chempur, 99.8%) and tetraethyl orthosilicate (Sigma-Aldrich), added at a concentration of 4.66 µL per gram of total mass. Planetary milling was carried out in a Fritsch Pulverisette 7 Premium Line system using 5 mm Si_3N_4 balls. The milling process consisted of ten cycles of 10 min at 300 rpm, each followed by a 10 min cooling break. An 80 mL silicon nitride jar and 250 milling balls were used. The slurry was dried on a hot plate at 80 °C for 24 h to remove ethanol. The powder was then subjected to two stages of thermal treatment in air muffle furnaces. In the first stage, the powder was heated to 600 °C at 5 °C/min (30 min dwell), then to 900 °C at 7 °C/min (30 min dwell). In the second stage, it was heated to 600 °C at 10 °C/min (15 min dwell), then to the target temperature of 1100–1600 °C at 10 °C/min (120 min dwell), followed by controlled cooling to 400 °C at 10 °C/min.

After annealing, the obtained powders were manually ground in an alumina mortar to obtain a homogeneous powder fraction suitable for subsequent characterization. The XRD patterns were collected in the 2θ range of 5–100° with a step size of 0.01° and a scanning rate of 2.0156° per minute. Cu $K\alpha_1$ radiation (λ = 1.5406 Å) with a nickel filter was used. The quantitative phase composition and lattice parameters were evaluated by Rietveld refinement.

Emission spectra and excited-state lifetimes were measured at room temperature using an Edinburgh Instruments FS5 spectrometer with a 150 W xenon lamp, Czerny–Turner monochromators, an optical system optimized for powder measurements, and a temperature-stabilized single-photon-counting PMT (Hamamatsu R928). Spectra were corrected for instrumental response using FLUORACLE. Lifetime measurements were performed in time-resolved mode, and decay curves were fitted using dedicated exponential-fit routines in FLUORACLE to obtain Pr^{3+} emission lifetimes.

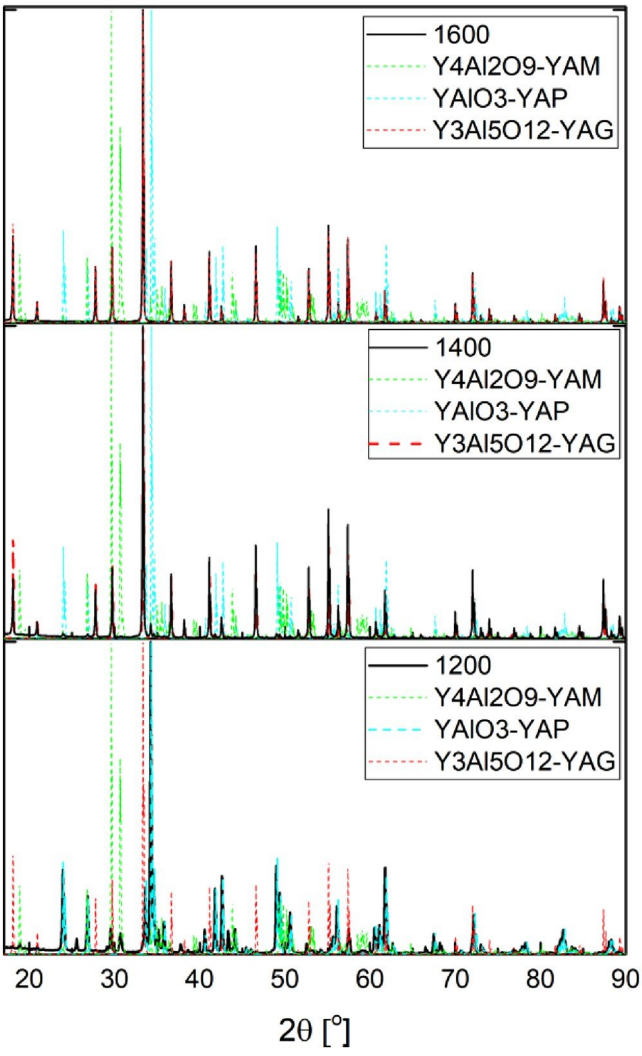


Figure 1. XRD pattern of YAG:1%Pr^{3+} annealed at 1200 °C, 1400 °C, 1600 °C, for 2 h, with overlaid reference lines of the best-matching phase standards from crystallographic databases.

3 Results and discussion

3.1 Structural characterization

YAG:1%Pr^{3+} samples were synthesized and thermally treated at temperatures ranging from 1100 to 1600 °C for 2 h. Subsequently, XRD measurements were performed to analyse the crystalline structure formed at each annealing temperature. The XRD pattern of the sample annealed at 1600 °C, 1400 °C, and 1200 °C is presented in Figure 1.

Phase identification analysis indicates that the sample annealed at 1200 °C for 2 h is multiphase, with a predominance of the intermediate YAP crystalline phase (YAP = 71.4 wt.%, YAM = 6.8 wt.%, YAG = 1.79 wt.%, Al_2O_3 = 19.7 wt.%, Y_2O_3 = 0.33 wt.%). With increasing annealing temperature, the content of the target YAG crystalline phase increases. At an annealing temperature of 1400 °C, the YAG phase becomes dominant (YAG = 92.1 wt.%, YAP = 5.01 wt.%, Al_2O_3 = 2.9 wt.%). Upon further

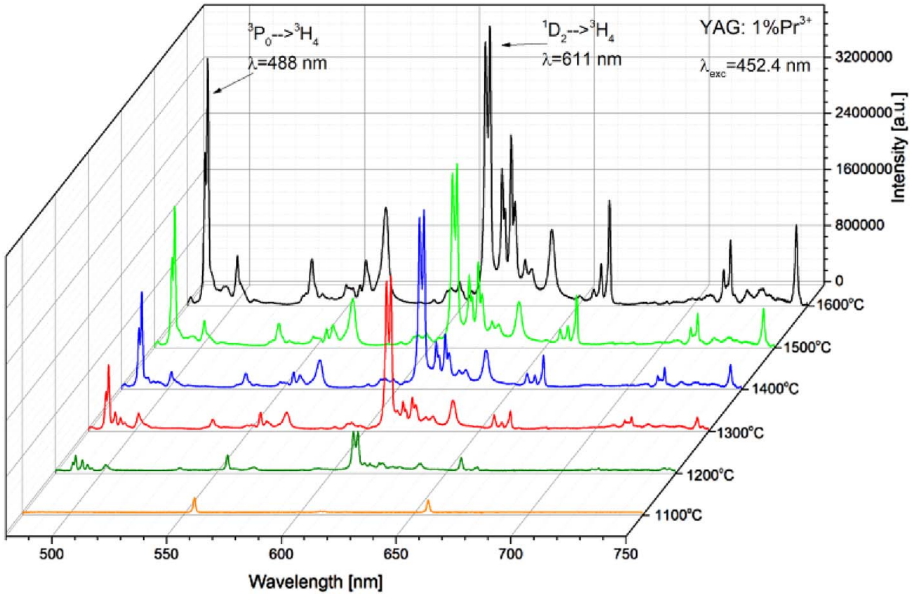


Figure 2. Emission spectra of YAG:1% Pr^{3+} samples annealed to different temperatures.

increasing the annealing temperature to 1600 °C, a nearly single-phase YAG sample is obtained (YAG = 99.1 wt.%, YAP = 0.3 wt.%, YAM = 0.6 wt.%).

3.2 Emission spectra

The measured emission spectra of YAG:1%Pr^{3+} powders are shown in Figure 2.

As can be seen in Figure 2, the intensity of the emission lines varies with the temperature at which the crystalline phases were formed, showing an overall increase in peak intensities with increasing formation temperature. The intensities of the two strongest peaks exhibit distinct trends with annealing temperature: the 488 nm line increases more markedly, while the 611 nm line shows a comparatively weaker dependence. After thermal annealing at 1100 °C for 2 h, the luminescence is too weak to reliably determine the intensities of the 488 nm and 611 nm peaks.

Accordingly, the intensity ratio of the $^3P_0 \rightarrow {}^3H_4$ (488 nm) and $^1D_2 \rightarrow {}^3H_4$ (611 nm) transitions was measured as a function of annealing temperature (Fig. 3).

A linear increase in the $^3P_0 \rightarrow {}^3H_4/{}^1D_2 \rightarrow {}^3H_4$ peak intensity ratio with increasing annealing temperature allowed calibration from 1300 to 1600 °C. The slope, determined by least-squares fitting, was $0.001 \pm 7.49 \times 10^{-5} \circ C^{-1}$, defining the powder's temperature sensitivity.

3.3 Fluorescence decay time

An analysis of the change in excited-state decay times under the influence of temperature in the range of 1100 °C to 1600 °C was carried out. Figure 4 presents the examination of the $^1D_2 \rightarrow {}^3H_4$ transition.

Changes in the annealing temperature of the luminescent powder also affect the lifetimes of the excited states. As the annealing temperature increases, the luminescence decay time becomes shorter. At 1600 °C, a single-phase

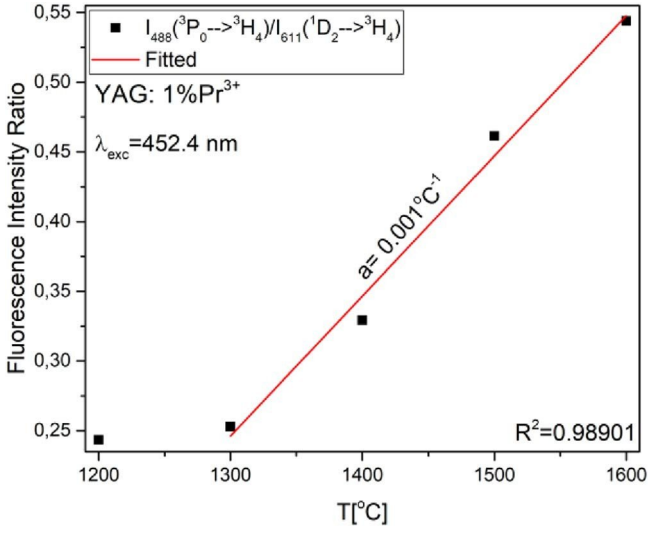


Figure 3. Change of peak intensity ratios as a function of temperature in YAG:1%Pr^{3+} powder.

YAG sample with high symmetry and a regular crystal structure is obtained. Such a homogeneous environment promotes radiative transitions and stronger electron–photon interactions, resulting in shorter decay times.

At lower annealing temperatures, the powder is multiphase, initially composed mainly of YAM and YAP phases. Upon heating, these phases irreversibly transform into YAG, creating a more homogeneous environment as the temperature rises. The evolving crystal structure modifies the crystal field symmetry and the local environment of the active Pr^{3+} ions, thereby altering the intensity parameters and probabilities of electronic transitions. Consequently, the lifetime of the $^1D_2$ level decreases as the annealing temperature increases.

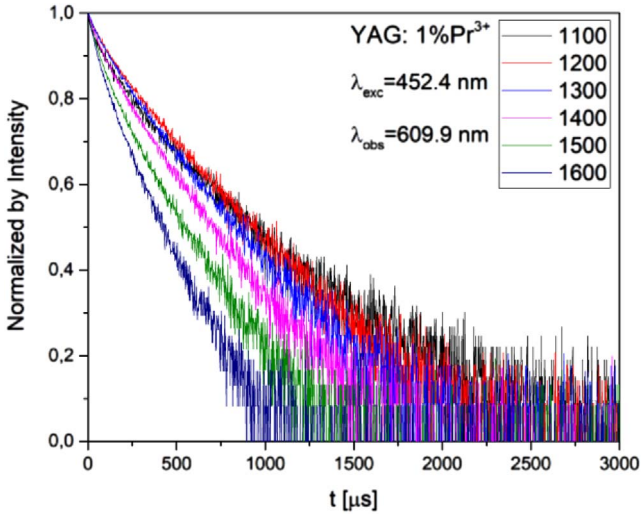


Figure 4. Luminescence decay curves resulting from the $^1D_2 \rightarrow {}^3H_4$ transition in YAG:1%Pr powder, recorded after annealing at different temperatures ranging from 1100 to 1600 °C.

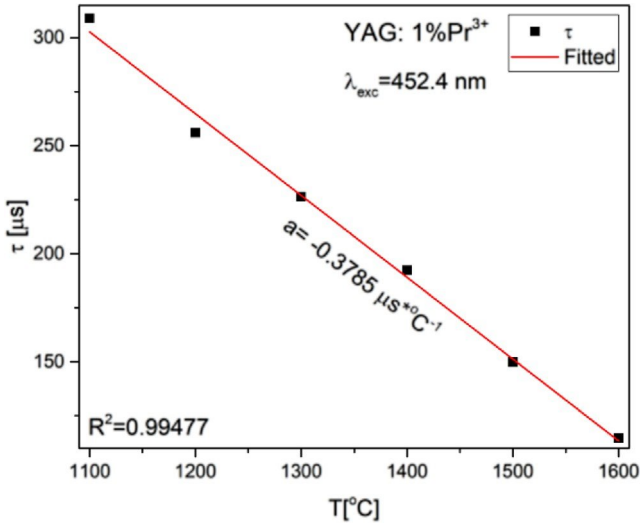


Figure 5. Measured lifetime of the $^1D_2$ excited state of Pr^{3+} ions in YAG:1%Pr powder as a function of annealing temperature.

The determined luminescence decay constants as a function of temperature are presented (Fig. 5).

A linear decrease in decay time with temperature allowed calibration from 1100 to 1600 °C. Least-squares fitting gave a slope of −0.3785 ± 0.0137 μs/°C, defining the powder's temperature sensitivity.

4 Conclusions

In this article, the potential of YAG:1%Pr^{3+} powder as a THS is presented. It was demonstrated that the maximum temperature experienced by the investigated material during an 2-hour treatment can be determined by analysis of specific emission intensity ratios and luminescence decay times.The results show that by examining the intensity ratio of the $^3P_0 \rightarrow {}^3H_4$ (488 nm) and $^1D_2 \rightarrow {}^3H_4$ (611 nm) transitions, temperature detection is possible in the range of 1300–1600 °C with a sensitivity of 0.001 ± 7.49 × 10^{-5} $°C^{-1}$. Furthermore, it was shown that thermal history can also be determined based on the excited-state lifetimes of Pr^{3+} ions. Analysis of the luminescence decay time for the $^1D_2 \rightarrow {}^3H_4$ transition (λ_{exc} = 454.2 nm, λ_{obs} = 609.9 nm) revealed that temperature detection in the range of 1100–1600 °C is achievable with a sensitivity of −0.3785 ± 0.0137 μs. $°C^{-1}$. Although the lifetime-based approach enables temperature sensing over a broader temperature range, (1100–1600 °C), the FIR method provides higher sensitivity, as the intensity ratio of the $^3P_0 \rightarrow {}^3H_4$ (488 nm) and $^1D_2 \rightarrow {}^3H_4$ (611 nm) transitions changes more strongly with temperature than the $^1D_2 \rightarrow {}^3H_4$ luminescence decay time. It was also observed that as the temperature at which the powder had been held decreases, the material transforms into a multiphase structure (YAM → YAP → YAG). This structural evolution changes the symmetry of the local environment of Pr^{3+} ions, resulting in reduced luminescence intensity and prolonged excited-state lifetimes. These correlations confirm that YAG:1%Pr^{3+} is a promising material for luminescence-based recording and readout of thermal history in high-temperature processes.

Funding

Presented research results were carried out as part of the 7th edition of the Implementation Doctorate programme funded by the Polish Ministry of Science and Higher Education.

Conflicts of interest

The authors declare that they have no competing interests to report.

Data availability statement

Data will be available after request

Author contribution statement

Conceptualization, O. Bogucki, M. Kaczkan and A.Kozłowska.; Methodology, O. Bogucki.; Validation, O.Bogucki., M. Kaczkan. and A. Kozłowska.; Formal Analysis, O. Bogucki.; sample synthesis A. Pantoł-Boczon.; Investigation, O. Bogucki.; Resources, O. Bogucki.; Data Curation, O. Bogucki.; Writing – Review & Editing, O. Bogucki.; Supervision, M. Kaczkan. and A. Kozłowska

References

1 Feng J, Xiong L, Wang S, Li S, Li Y, Yang G, Fluorescent temperature sensing using triarylboron compounds and microcapsules for detection of a wide temperature range on the micro- and macroscale, *Adv. Funct. Mater.* **23**, 340–345 (2013). https://doi.org/10.1002/adfm.201201712.

2 Baker B, *Temperature Sensing Technologies, AN673* (1998). Avalable at http://www.microchip.com.

3 Douglas J, Smith C, Taylor S, An integrated approach to the application of high bandwidth optical pyrometry to turbine blade surface temperature mapping, in: *Proceedings of the 18th International Congress on Instrumentation in Aerospace Simulation Facilites* (IEEE, 1999). https://doi.org/10.1109/ICIASF.1999.827143.

4 Kerr C, Ivey P, An overview of the measurement errors associated with gas turbine aeroengine pyrometer systems, *Measure. Sci. Technol.* **13**, 873 (2002).

5 Sun T, Zhang ZY, Grattan KTV, Erbium/Ytterbium fluorescence based fiber optic temperature sensor system, *Rev. Sci. Instrum.* **71**, 4017 (2000). https://doi.org/10.1063/1.1289682.

6 Yen WM, Shionoya S, Yamamoto H, *Phosphor Handbook*, 2nd ed (CRC Press/Taylor and Francis, Boca Raton, 2007).

7 Sun H, Peng D, Wang X, Tang M, Zhang Q, Yao X, Green and red emission for $(K_{0.5}Na_{0.5})NbO_3$:Pr ceramics, *J. Appl. Phys.* **111**, 046102 (2012). https://doi.org/10.1063/1.3686193.

8 Wang B, Lin H, Yu Y, Chen D, Zhang R, Xu J, Wang Y, Ce^{3+}/Pr^{3+}: YAGG: a long persistent phosphor activated by blue light, *J. Am. Ceram. Soc.* **97**, 2539 (2014). https://doi.org/10.1111/jace.12967.

9 Ikesue A, Kinoshita T, Kamata K, Yoshida K, Fabrication and optical properties of high-performance polycrystalline Nd: YAG ceramics for solid-state lasers, *J. Am. Ceram. Soc.* **78**, 4 (1995). https://doi.org/10.1111/j.1151-2916.1995.tb08433.x.

10 Heyes A L, Rabhiou A, Feist JP, Kempf A, Thermal history sensing with thermographic phosphors, *AIP Conf. Proc.* **1552**, 891 (2013). https://doi.org/10.1063/1.4821412.

11 Tian Y, Tian Y, Huang P, Wang L, Shi Q, Cui C (2016) Effect of Yb^{3+} concentration on upconversion luminescence and temperature sensing behavior in Yb^{3+}/Er^{3+} co-doped YNbO4 nanoparticles prepared via molten salt route, *Chem. Eng. J.* **297**, 26. https://doi.org/10.1016/j.cej.2016.03.149.

12 Shakirova JR, Shevchenko NN, Baigildin VA, Chelushkin PS, Khlebnikov AF, Tomashenko OA, Solomatina AI, Starova GL, Tunik SP, Eu-based phosphorescence lifetime polymer nanothermometer: a nanoemulsion polymerization approach to eliminate quenching of Eu emission in aqueous media, *ACS Appl. Polym. Mater.* **2**, 537–547 (2020). https://doi.org/10.1021/acsapm.9b00952.

13 Amiel S, Copin E, Sentenac T, Lours P, Le Maoult Y, On the thermal sensitivity and resolution of a YSZ:Er^{3+}/YSZ:Eu^{3+} fluorescent thermal history sensor, *Sens. Actuators A Phys.* **272**, 42–52 (2018). https://doi.org/10.1016/j.sna.2018.01.040.

14 Chi F, Jiang B, Zhao Z, Chen Y, Wei X, Duan C, Yin M, Xu W, Multimodal temperature sensing using Zn_2GeO_4:Mn^{2+} phosphor as highly sensitive luminescent thermometer, *Sens. Actuators B Chem.* **296**, 126640 (2019). https://doi.org/10.1016/j.snb.2019.126640.

15 Zhao Z, Hu F, Cao Z, Chi F, Wei X, Chen Y, Duan C, Yin M, Highly uniform and monodisperse β-$NaYF_4$:Sm^{3+} nanoparticles for a nanoscale optical thermometer, *Opt. Lett.* **43**, 835–838 (2018). https://doi.org/10.1364/OL.43.000835.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 28
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026022
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Enhancing the fluorescence of low quantum yield fluorophores using plasmonic metasurfaces

Veronica Anăstăsoaie, Iuliana Mihalache, Oana Brincoveanu, Adrian Dinescu, Cristian Kusko, Dana Cristea, and Roxana Tomescu*

National Institute for Research and Development in Microtechnology – IMT Bucharest, 126A, Erou Iancu Nicolae Street, 077190, Voluntari, Ilfov, Romania

Received 30 January 2026 / Accepted 4 March 2026

Abstract. Metasurfaces represent an innovative category of optical surfaces that can be engineered to exhibit distinctive properties that are not observed in conventional materials. It has been demonstrated that these structures exhibit robust light-matter interactions at the nanoscale, which renders them a promising solution for enhancing the sensitivity of various biomedical sensing technologies. This work presents a versatile approach for enhancing the broadband spectral emission of various fluorophores. The approach involves developing large-area, lithography-free plasmonic platforms without tailoring the geometry of the metasurface to a specific chromophore. The investigation encompasses the fluorescence behaviour of low quantum yields fluorophores such as Crystal Violet, Nile Red, and Rose Bengal, dispersed in ethanol solutions and coated onto the metallic arrays with random configuration. The results indicate that metallic nanostructures efficiently support plasmon-mediated fluorescence enhancement, and the degree of this enhancement depends strongly on both the fluorophore type and the metal layer thickness. The highest fluorescent enhancement factor was obtained for Crystal Violet on silver metasurfaces.

Keywords: Plasmonic metasurface, Lithography-free platforms, Fluorescence enhancement, Nanophononics, Low quantum yield fluorophore.

1 Introduction

Recent advances and technological innovations in the field of plasmonics have led to a variety of exciting applications that address important sensing challenges. In this context, extensive studies are conducted on fluorescence for biosensing purposes. The fluorescence analysis method is known for properties like easiness, high sensitivity and selectivity, and short response time [1, 2]. A particularly important topic in this research area is the enhancement of fluorescence for low-quantum-yield fluorophores which have applications in biomedical imaging [3], sensing [4], and optoelectronics [5]. These materials exhibit strong absorption, environmental sensitivity, or useful photochemistry; however, their practical use is frequently limited by weak emission and rapid photodegradation under high excitation irradiance. Therefore, various methods have emerged to amplify fluorescence signals for increasing detection [6–8]. Plasmonic metamaterials have been shown to be a promising solution that utilizes the intense interaction between light and matter at the nanoscale to increase the sensitivity of various biomedical spectral sensing technologies. Moreover, metal-enhanced fluorescence (MEF) is a method of enhancing fluorescence signals. Specifically, when a chromophore molecule is in close proximity to metal nanostructures, its fluorescence is enhanced due to the intensification of the localized electromagnetic field specific to plasmonic metasurfaces [9, 10]. However, this method has an important disadvantage: the fluorescence signal is quenched if the fluorophore is in direct contact with the metallic nanostructure. To overcome this problem, an ultra-thin film is required to act as a spacer.

Current research emphasizes the development of metasurface arrays that are thin, easy to fabricate, and capable of manipulating wave propagation both at the interface and in free space. These properties demonstrate a significant potential for enhancing fluorescence detection, making them invaluable tools in biosensing applications [11, 12]. Although, the devices tailored for specific resonance achieve sharp field maxima [13], the technological processes required for their fabrication are complex and expensive. Meanwhile, the random morphology metasurfaces offer a wide range emission due to disordered dispersers supporting various localized plasmon modes that overlap. This improves

* Corresponding author: roxana.tomescu@imt.ro

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

coupling bandwidth and field redistribution, enhancing fluorophore excitation and emission over a broad spectrum [14–16]. Also, broadband responses can result from constructive interference of multiple multipolar moments in disordered assemblies, broadening the spectral response by suppressing narrow reflectivity features [17]. Here, we propose lithography-free metasurfaces as a scalable, cost-effective alternative to traditional nanofabrication for fluorescence biosensing. By using random metal nanoparticle aggregates, we achieved a 423-fold fluorescence enhancement while avoiding the high costs of precision-engineered structures [18]. Large-scale plasmonic metasurfaces can be fabricated also via laser-induced periodic surface structures or chemical synthesis. Although these devices offer efficient manufacturing of ordered arrays, it prioritizes polarization control over broadband fluorescence enhancement [19, 20]. Similarly, high-performance chemical systems like gold nanotubes face large-area integration challenges [21, 22]. Random aggregates overcome these issues through superior substrate compatibility and simple technological processes.

For fluorescence-based sensors Crystal Violet (CV), Nile Red (NR), and Rose Bengal (RB) fluorophores have been found to be highly useful, all exhibiting low quantum yields – 0.7, 0.11, 0.019, respectively [23–25]. Studies on these fluorophores investigated their interactions with nanostructures to understand the mechanisms and limitations of enhancement platforms. CV, an achiral weak emitter, has found extensive use in MEF intensification studies, particularly for the detection of single molecules [21] or in structures with localized intense fields [26]. NR was utilised in the study of enhanced emission on bimetallic plasmonic platforms (silver nanoparticles and gold nanolayers) derived from sol-gel [27]. RB aggregates have been extensively studied in the context of interaction with metal nanoparticles – AgNPs and GNPs [28–30]. Most high-performance MEF platforms reported in the literature rely on nanoparticle colloids, spacer-engineered films, or lithographically defined arrays to control distance and resonance conditions. While effective, these methods present fabrication complexity, limited throughput, and constrained active area, which complicate optimization and translation to large-area coatings or device-scale substrates. In contrast, thermally annealed Ag thin films demonstrate that simple post- processing can tune plasmonic response via morphology changes (particle reshaping and spacing), and Al plasmonics offers UV-VIS response with strong sensitivity to oxide and morphology [18, 31].

Previously, we demonstrated that influence of plasmonic metasurfaces on fluorescence intensity enhancement for rhodamine R6G, a hight quantum yield fluorophore, on silicon or glass substrate [18]. This work presents a simple plasmonic metasurface approach for intensifying the fluorescence emission of low-quantum-yield dyes. This method uses thin-film-derived nanostructures that do not require a specific configuration of the meta-atom geometry or lithography. These large area platforms can also provide scalable, morphology-driven, plasmon-enhanced fluorescence. We explore thermally annealed Ag nanostructured films (15–25 nm) which develop from isolated islands to interconnected networks, and the ultrathin Al nanostructured films (2–6 nm), which exhibit intrinsic nanoscale dielectric separation, owing to native oxide. Native oxide provides nanoscale separation that minimise fluorescence quenching while enabling coupling to localized plasmonic fields. We used Crystal Violet, Nile Red, and Rose Bengal as low-emissivity and environmentally sensitive fluorophores to link fluorescence enhancement to morphological changes, which were confirmed by SEM and particular excitation settings for each dye.

2 Material and methods

In this investigation, silicon was utilized as the substrate, while the metals under analysis were aluminium (Al) and SILVER (Ag). The metasurfaces were manufactured on a large area of 1.5 cm^2 by electron beam evaporation using a TEMESCAL FC-2000 (Temescal, Livermore, CA, USA) for Al and Elettrorava system (Elettrorava S.r.l., Torino, Italy) for Ag. This method was used to deposit either discontinuous layers (extremely thin – e.g. 2 nm, 4 nm and 6 nm – thick mass equivalent – for Al) or thin continuous layers (15 nm, 20 nm, and 25 nm for Ag) subjected to specific annealing conditions to achieve percolation of nanoaggregates. Table 1 show the process parameters employed for the fabrication of the proposed metasurfaces.

Stock solutions of Crystal Violet (CV), Nile Red (NR), and Rose Bengal (RB) fluorophores (from Carl ROTH) were prepared in ethanol at a concentration of 4 mM. Thin films were fabricated via drop-casting 50 µL aliquots of each dye solution onto the nanostructured surface containing metallic meta-atoms. The morphological characterisation of the samples obtained was performed using scanning electron microscopy (SEM) with a Field Emission Scanning Electron Microscope (FEG-SEM) – Nova NanoSEM 630 (FEI Company, USA). Also, the meta-atoms size distributions were determined from high-resolution SEM images by measuring approximately 200 individual nanoaggregates for each sample. The resulting size histograms were fitted using a Gaussian function to extract the mean particle size, standard deviation (SD), and full width at half maximum (FWHM).

Steady-state fluorescence emission and optical absorption measurements were performed using an FLS920 fluorescence spectrometer (Edinburgh Instruments Ltd., UK) featuring a 450 W xenon arc lamp as the excitation source. Excitation wavelengths were selected according to the absorption maxima of each fluorophore: 550 nm – CV, 540 nm – NR, and 500 nm – RB. The obtained emission spectra revealed bands centred at wavelengths characteristic of thin-film fluorescence: 636 nm – CV, 635 nm – NR, and 571 nm RB.

3 Results and discussions

SEM micrographs of the 2 nm, 4 nm, and 6 nm (equivalent mass) Al thin layers (Fig. 1) reveal that ultrathin

Table 1. Process parameters for metal layer deposition and thermal annealing.

Material	Thickness (nm)	Deposition rate (Å/s)	Pressure (mbar)	Thermal annealing conditions
Ag	15	0.3	2.5×10^{-6}	250 °C, 30 min
	20, 25			250 °C, 1 h
Al	2, 4, 6	0.1	$>10^{-6}$	–

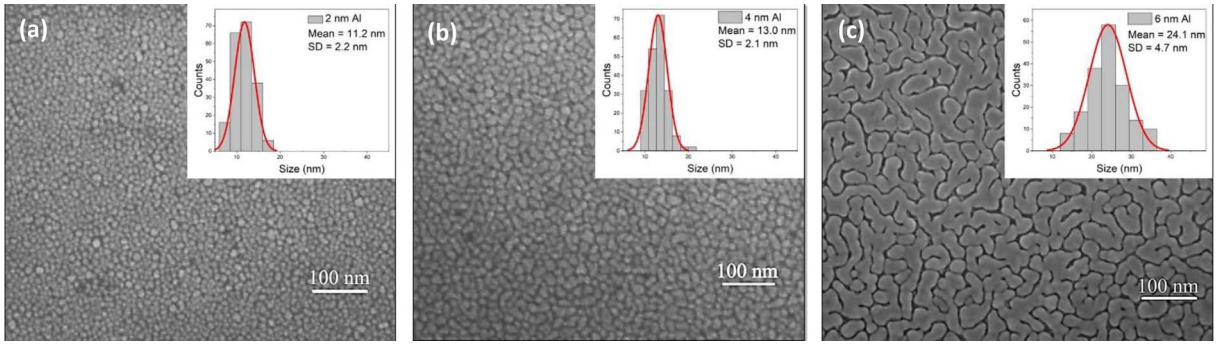


Fig. 1. SEM micrographs showing the morphology after the Al layer deposition process: (a) 2 nm, (b) 4 nm, and (c) 6 nm.

aluminium undergoes a morphological transition from isolated islands to a gradually interconnected percolated network of nanostructures as thickness increases from 2 to 6 nm. The statistical analysis (inset to Fig. 1) reveals a systematic increase of the mean particle size with increasing aluminium thickness. Specifically, the mean size increases from 11.2 nm for the 2 nm Al layer to 24.1 nm for the 6 nm Al layer, while the corresponding standard deviation increases from 2.2 nm to 4.7 nm, indicating enhanced particle coalescence and growth at higher nominal thicknesses. The resulting size histograms also show the FWHM values for each layer (9–14, 10–15, and 18–29 nm).

The enhancement factor varies depending on the shape and distance of the metallic nanoaggregates [32]. Taking this into account, we collected the fluorescence emission across the entire surface of the nanostructured platform. Due to the resulting surface filling factor and the consistent shapes and sizes of the nanoaggregates achieved under the same deposition or annealing conditions, the MEF remains relatively unchanged when analysing new samples, ensuring high reproducibility. The steady-state fluorescent emission spectra of CV, NR, and RB on nanostructured Al films were compared to a reference consisting of a silicon substrate without metasurfaces coated with fluorophore. Figure 2 shows the obtained fluorescence spectra, with an inset graph on a logarithmic scale illustrating the magnitude orders of the best enhancement result. The fluorescence enhancement depends on spectral overlap between excitation/emission bands and plasmonic scattering/near-field response. For all three fluorophores, nanostructured Al induced changes in fluorescence intensity without significantly affecting the characteristic spectral band profiles. This indicates that the dominant interaction is photonic (near-field/plasmonic) and not a chemical modification of the fluorophore. Such behaviour is largely associated with MEF, in which local electromagnetic fields and modified radiative decay rates can increase emission intensity without necessarily changing spectral maxima.

The EF are calculated relative to the respective emission maxima of each fluorophore. The values vary depending on the type of fluorophore – specifically, their quantum efficiency – rather than the layer thickness. For NR, we obtain a maximum EF of approximately 3.2 when we use a thin film of 4 nm thickness. In the meantime, for CV and RB, we obtain maximum EF values of 2.8 and 7.7, respectively, when we use a thickness of 6 nm.

Silver thin films with thicknesses of 15 nm, 20 nm, and 25 nm were deposited on silicon and subsequently subjected to thermal annealing. This treatment induced the formation of metal aggregates, resulting in randomly distributed meta-atom structures. Figure 3 shows SEM images of the annealed Ag layers at all three thicknesses, revealing the morphology of the formed nanostructures. SEM analysis demonstrates a systematic thickness-dependent morphological transition in the annealed Ag films. The 15 nm film exhibits isolated, quasi-ellipsoidal nanoparticles with minimal inter-particle connectivity. At 20 nm, partial aggregation occurs, forming bigger anisotropic islands with increased contact between adjacent structures. The 25 nm film displays complete percolation, characterized by an interconnected spiderweb-like metallic network.

Similar to Al films, the particle size statistics for Ag nanoaggregates (see inset in Fig. 3) show an increase in mean particle size with layer thickness: 124 nm for the 15 nm layer, 189 nm for the 20 nm layer, and 234 nm for the 25 nm layer. The corresponding standard deviation increases as well, from 34 nm to 57 nm. Furthermore, the surface filling factor, which is defined as the fraction of the substrate surface covered by Ag nanoparticles, increases with layer thickness. Specifically, it increases from 32% for the 15 nm layer, to 37% for the 20 nm layer, and to 76% for the 25 nm layer. This increase in surface coverage reflects

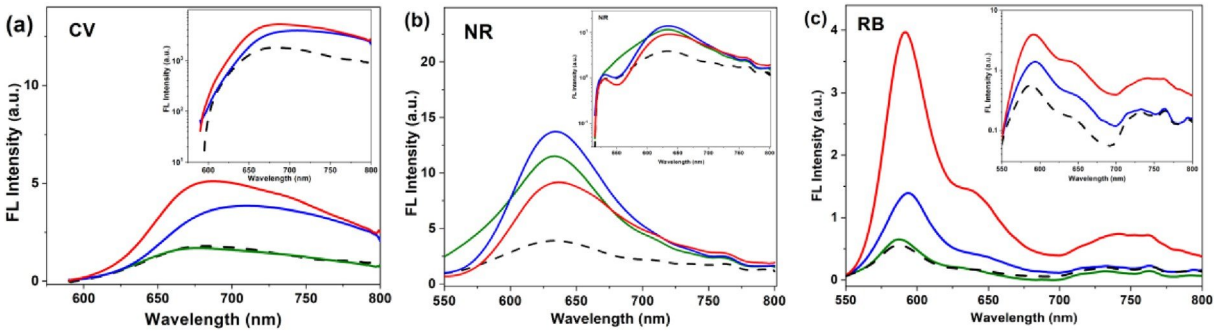


Fig. 2. Fluorescence intensity spectra of a) CV, b) NR and c) RB on nanostructured Al films; dashed black: reference, green/blue/red: 2/4/6 nm Al. Inset are the MEF with logarithmic y scale.

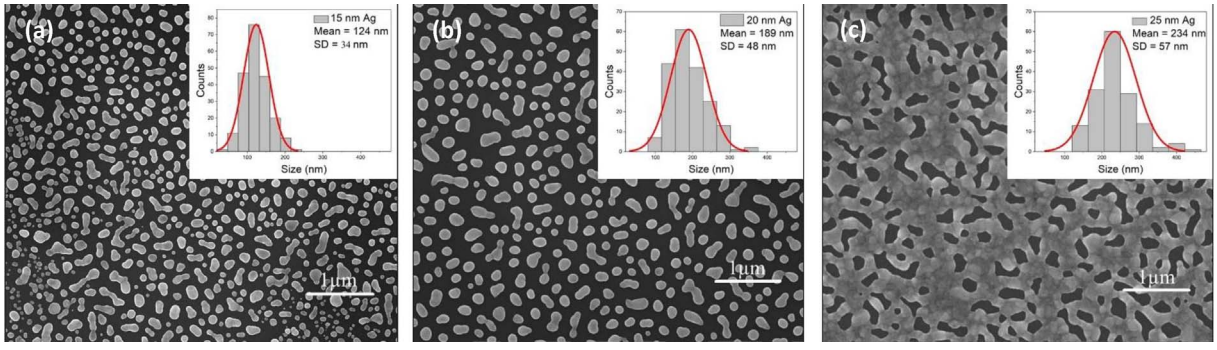


Fig. 3. SEM micrographs showing the morphology obtained following the deposition of Ag layers structured by annealing: (a) 15 nm, (b) 20 nm, (c) 25 nm.

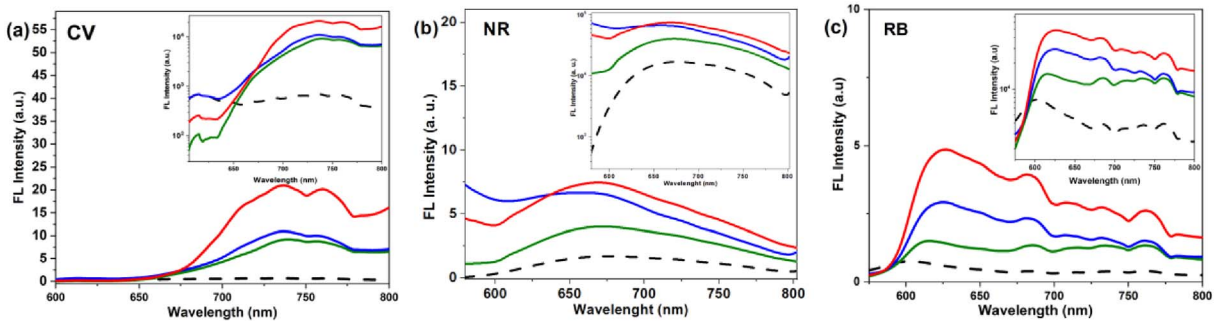


Fig. 4. Fluorescence intensity spectra of a) CV, b) NR and c) RB on nanostructured Ag films; dashed black: reference, green/blue/red: 15/20/25 nm Ag. Inset are the MEF with logarithmic y scale.

the transition from isolated nanoparticle regimes toward a densely packed morphology, which is consistent with enhanced surface diffusion, particle aggregation, and lateral growth during thermal treatment.

Figure 4 presents the fluorescence emission spectra of CV, NR, and RB fluorophores films deposited on thermally annealed Ag nanostructures, in comparison with a reference substrate, with an inset graph on a logarithmic scale. The fluorescence measurements reveal substantial intensity increases relative to the reference for all fluorophores, with EF dependent on Ag thickness and fluorophore: for CV the EF increases from ≈6.8 (15 nm) to ≈10 (20 nm) and ≈19 (25 nm), for NR from ≈2.4 (15 nm) to ≈4 (20 nm) and ≈4.4 (25 nm), and for RB from ≈2 (15 nm) to ≈4 (20 nm) and ≈6 (25 nm).

Table 2 summarizes the fluorescence EF for CV, NR, and RB deposited on nanostructured Al and Ag films of varying thicknesses. For Al substrates, moderate enhancement is observed, with EF values increasing with thickness for CV and RB and showing an optimum at 4 nm for NR.

Table 2. Fluorescence enhancement factors for CV, NR, RB on nanostructured Al films (2, 4, 6 nm) and Ag films (15, 20, 25 nm).

Fluorophore	Enhancement factor (EF) on specific layer					
	Aluminium (mass equivalent thickness)			Silver		
	2 nm	4 nm	6 nm	15 nm	20 nm	25 nm
Crystal Violet	1	2	3	8.6	10	19
Nile Red	3	3.2	2.3	2.4	4	4.4
Rose Bengal	1.2	3	7.7	2	4	6

In contrast, Ag films exhibit substantially stronger enhancement for all fluorophores, particularly at larger Ag thicknesses. CV shows the highest enhancement overall, reaching EF = 19 on the 25 nm Ag film, while RB and NR reach maxima of EF = 6 and EF = 4.4, respectively. These results indicate that Ag nanostructures support more efficient plasmon-mediated fluorescence enhancement than Al under the studied conditions, with the degree of enhancement strongly dependent on both the fluorophore and metal layer thickness. Furthermore, the highest EF are obtained when using the 25 nm Ag film. The high filling factor observed for this metasurface platform suggests the onset of particle percolation, which influences significantly the plasmonic coupling. Additionally, the spiderweb pattern that forms after annealing – consisting of a high number of sharp angles – significantly intensifies the localized electromagnetic field.

4 Conclusion

In this study we presented a simple plasmonic metasurface approach for enhancing fluorescence of low quantum yield dyes using thin film derived nanostructures. We performed thermal annealing on Ag films (15–25 nm), and deposited mass equivalent ultrathin Al films (2–6 nm) to develop large plasmonic metasurface platforms comprising of randomly configurate metallic nano-aggregates. The native oxide of studied metals provides the separation required to partially mitigate the extreme near-field quenching while enabling coupling to localized plasmonic fields from nanoscale features. We investigated the behaviour of the fluorescence intensity for Crystal Violet, Nile Red, and Rose Bengal in the presence of metal nanostructures. The EF are calculated relative to the respective emission maxima of each fluorophore. The values vary depending on the type of fluorophore – specifically, their quantum efficiency – rather than the layer thickness. For Al, promising results concerning fluorescence intensity enhancement were obtained using 6 nm mass equivalent layers, with the best EF for Rose Bengal. At low thicknesses (e.g., 6 nm), Al forms monodisperse, nearly periodic nanostructures with very low heights. This leads to a uniform distribution of localized field enhancements. Since the structures are uniformly spaced, a large portion of the fluorophores can interact with the plasmonic resonances, resulting in consistent enhancement.

For Ag the fluorescence measurements revealed substantial intensity increases, with EF dependent on Ag thickness and fluorophore, the best result was obtained for Crystal Violet for 25 nm with an EF of 19. Silver forms large aggregates, evolving from hemispherical shapes to percolated spiderweb at a thickness of 25 nm. The enhancement occurs via two mechanisms: i) height matching ensures most molecules are within the plasmonic interaction zone of the 25 nm structures, and ii) natural spacing, where the native oxide layer prevents quenching by maintaining distance between the dyes and metal surface. Although aluminium is suitable for thin layers, the taller percolated structures offered by annealed silver thicker films exhibit significant enhancement of fluorescence emission.

The proposed low-cost, lithography-free platforms offer a versatile approach for enhancing the broadband spectral emission of a variety of fluorophores without specifically tailoring the geometry for each chromophore, providing a scalable, morphology-driven, plasmon-enhanced fluorescence. Furthermore, the fabrication method used in this study allows for nanostructuring large wafers of various substrate – silicon, glass, or ceramics, for a multitude of hands-on applications requiring a rapid response (e.g. refractive index-based sensors, fluorescence sensors, photocatalysis, etc.).

Acknowledgments

The authors acknowledge Gabriel Craciun for electron beam deposition using Elettrorava system.

Funding

This research was funded by the Core Program of IMT-Bucharest within the National Research Development and Innovation Plan 2022-2026, carried out with the support of Romanian Ministry of Research, Innovation and Digitization, project no. 2307.

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

This article has no associated data generated and/or analyzed.

Author contribution statement

Conceptualization, R.T. and V.A.; Methodology, D.C.; Validation, V.A, C.K. and R.T; Formal Analysis, V.A. and R.T.;

Investigation, V.A., I.M., O.B., A.D.; Data Curation, V.A., O.B., and I.M.; Writing – Original Draft Preparation, V.A. and R.T.; Writing – Review & Editing, V.A., R.T., D.C. and C.K.; Supervision, R.T., D.C. and C.K.; Project Administration, R.T. and D.C.; Funding Acquisition, R.T. and D.C.

References

1 Huang Y, Zhai J, Liu L, Shang Z, Zhang X, Huang H, Shen B, Chen G, Recent developments on nanomaterial probes for detection of pesticide residues: a review, *Anal Chim Acta.* **1215**, 339974 (2022). https://doi.org/10.1016/j.aca.2022.339974.

2 Bai L et al., Molecular imprinted polymer capped luminescent metal-organic framework for the determination of pyrethroids and its metabolite, *Microchem J.* **193**, 109047 (2023). https://doi.org/10.1016/j.microc.2023.109047.

3 Narasimhan V, Siddique RH, Hoffmann M, Kumar S, Choo H, Enhanced broadband fluorescence detection of nucleic acids using multipolar gap-plasmons on biomimetic Au metasurfaces, *Nanoscale* **11**, 3750 (2019). https://doi.org/10.1039/C9NR03178B.

4 Xu J, Morton W, Jones D, Tabish TA, Ryan MP, Xie F, Significant quantum yield enhancement for near infrared fluorescence dyes by silica templated silver nanorods, *Appl. Phys. Rev.* **9**, 031406 (2022). https://doi.org/10.1063/5.0082187.

5 Traverso AJ et al., Low-loss, centimeter-scale plasmonic metasurface for ultrafast optoelectronics, *Optica* **8**(2), 202 (2021). https://doi.org/10.1364/OPTICA.400731.

6 Bardhan R, Grady NK, Cole JR, Joshi A, Halas NJ, Fluorescence enhancement by Au nanostructures: nanoshells and nanorods, *ACS Nano.* **3**(3), 744 (2009). https://doi.org/10.1021/nn900001q.

7 Yu L, Yang Y, Jiang X, Li Y, He X, Chen L, et al. A self-calibrating ratiometric fluorescence sensor with photonic crystal-based signal amplification for the detection of tetracycline in food, *Food Chem.* **451**, 139418 (2024). https://doi.org/10.1016/j.foodchem.2024.139418.

8 He H, Muhammad P, Guo Z, Peng Q, Lu H, Liu Z, Controllably prepared molecularly imprinted core-shell plasmonic nanostructure for plasmon-enhanced fluorescence assay, *Biosens. Bioelectron.* **146**, 111733 (2019). https://doi.org/10.1016/j.bios.2019.111733.

9 Lin J-S, Radjenovic PM, Jin H, Li J-F, Plasmonic core–shell nanoparticle enhanced spectroscopies for surface analysis, *ACS Publ.* **93**(17), 6573 (2021). https://doi.org/10.1021/acs.analchem.1c00233.

10 Xu J et al., Shell-isolated nanoparticle-enhanced Raman and fluorescence spectroscopies: synthesis and applications, *Adv. Optic. Mater.* **6**(4), 1701069 (2018). https://doi.org/10.1002/adom.201701069.

11 Iwanaga M, Productive biosensing techniques empowered by all-dielectric metasurfaces, *Front. Bioeng. Biotechnol.* **12**. (2025). https://doi.org/10.3389/fbioe.2024.1484638.

12 Khan SA et al., Optical sensing by metamaterials and metasurfaces: from physics to biomolecule detection, *Adv. Opt. Mater.* **10**(18), 2200500 (2022). https://doi.org/10.1002/ adom.202200500.

13 Montaño-Priede JL et al., An overview on plasmon-enhanced photoluminescence via metallic nanoantennas, *Nanophotonics* **13**(26), 4771–4794 (2024). https://doi.org/10.1515/nanoph-2024-0463.

14 Gartia MR, Hsiao A, Sivaguru M, Chen YC, Liu GL, Enhanced 3D fluorescence live cell imaging on nanoplasmonic substrate, *Nanotechnology* **22**(36), 365203–365203 (2011). https://doi.org/10.1088/0957-4484/22/36/365203.

15 Hohenberger E, Freitag N, Rosenmann D, Korampally V, A facile route towards large area self-assembled nanoscale silver film morphologies and their applications towards metal enhanced fluorescence, *Sens. Actuators B Chem.* **249**, 549–557 (2017). https://doi.org/10.1016/J.SNB.2017.04.012.

16 Lu H, et al., Broadband near-field enhancement in the macro-periodic and micro-random structure with a hybridized excitation of propagating Bloch-plasmonic and localized surface-plasmonic modes, *Nanoscale* **7**(40), 16798–16804 (2015). https://doi.org/10.1039/C5NR03391H.

17 González-Alcalde AK, Rojas-Martinez IY, Reyes-Coronado A, Broadband antireflective random metasurfaces, *J. Opt. Soc. Am. B* **38**(6), 1974–1983 (2021). https://doi.org/10.1364/JOSAB.420085.

18 Anăstăsoaie V et al., Influence of Random Plasmonic Metasurfaces on Fluorescence Enhancement, *Materials* **15** (4), 1429 (2022). https://doi.org/10.3390/ma15041429.

19 Tasolamprou AC, et al., Highly ordered laser imprinted plasmonic metasurfaces for polarization sensitive perfect absorption, *Sci. Rep.* **12**(1), 19769 (2022). https://doi.org/10.1038/s41598-022-21647-w.

20 Skoulas E et al., Laser induced periodic surface structures as polarizing optical elements, *Appl. Surf. Sci.* **541**, 148470 (2020). https://doi.org/10.1016/j.apsusc.2020.148470.

21 Khatua S, Paulo PMR, Yuan H, Gupta A, Zijlstra P, Orrit M, resonant plasmonic enhancement of single-molecule fluorescence by individual gold nanorods, *ACS Nano.* **8**(5), 4440 (2014). https://doi.org/10.1021/nn406434y.

22 Xu J et al., Significant quantum yield enhancement for near infrared fluorescence dyes by silica templated silver nanorods, *Appl. Phys. Rev.* **9**(3), 031406 (2022). https://doi.org/10.1063/5.0082187.

23 Brey LA, Schuster GB, Drickamer HG, High pressure studies of the effect of viscosity on fluorescence efficiency in crystal violet and auramine, *O. J. Chem. Phys.* **67**, 2648 (1977). https://doi.org/10.1063/1.435177.

24 Sackett DL, Wolff J, Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces, *Anal. Biochem.* **167**(2), 228 (1987). https://doi.org/10.1016/0003-2697(87)90157-6.

25 *Quest Database™ Quantum Yield [Rose Bengal]* (AAT Bioquest, Inc., 25 Jan. 2026). https://www.aatbio.com/resources/quantum-yield/rose_bengal.

26 Jones RR et al., Dense arrays of nanohelices: Raman scattering from achiral molecules reveals the near-field enhancements at chiral metasurfaces, *Adv. Mater.* **35**(34), 2209282 (2023). https://doi.org/10.1002/adma.202209282.

27 Synak A et al., Enhanced emission of Nile Red on plasmonic platforms, *Optical Mater.* **78**, 82 (2018). https://doi.org/10.1016/j.optmat.2018.02.009.

28 Swayam P et al., Microscopic perspective of synergy between localized surface plasmon resonance and disruption of dye aggregates in metal nanoparticle-enhanced fluorescence, *ACS Appl. Nano Mater.* **6**(19), 17539 (2023). https://doi.org/10.1021/acsanm.3c02734.

29 Gabudean AM, Focsan M, Astilean S, Gold nanorods performing as dual-modal nanoprobes via metal-enhanced fluorescence (MEF) and surface-enhanced Raman scattering

(SERS), *J. Phys. Chem. C* **116**(22), 12240 (2012). https://doi.org/10.1021/jp211954m.

30 Hsing-Hui L, Chen I-C, Study of the interaction between gold nanoparticles and rose bengal fluorophores with silica spacers by time-resolved fluorescence spectroscopy, *J. Phys. Chem. C* **119**(47), 26663 (2015). https://doi.org/10.1021/acs.jpcc.5b08477.

31 Knight MW et al., Aluminum for plasmonics, *ACS Nano.* **8** (1), 834 (2014). https://doi.org/10.1021/nn405495q.

32 Zhao X, Alizadeh MH, Reinhard BM, Harnessing leaky modes for fluorescence enhancement in gold-tipped silicon nanowires, *J. Phys. Chem. C.* **120**, 37 (2016). https://doi.org/10.1021/acs.jpcc.5b11702.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 33
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026036
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE

OPEN ACCESS

Atomic layer deposition for hafnium oxide-based meta-optics in the ultraviolet spectral range

Thomas Siefke[1,2,3,*], Kristin Gerold[1,2], Svetlana Shestaeva[2], Pallabi Paul[1,2], Shawon Alam[4], Daniel Franta[5], Adriana Szeghalmi[1,2], Sven Schröder[2], and Stefanie Kroker[6,7,8]

[1] Friedrich-Schiller-University, Institute of Applied Physics, Albert-Einstein-Strasse 15, 07745 Jena, Germany
[2] Fraunhofer Institute for Applied Optics and Precision Engineering IOF, Albert-Einstein-Str. 7, 07745 Jena, Germany
[3] Ernst-Abbe-Hochschule Jena, University of Applied Sciences, Carl-Zeiss-Promenade 2, 07745 Jena, Germany
[4] Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, 76131 Karlsruhe, Germany
[5] Department of Plasma Physics and Technology, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czechia
[6] TU Braunschweig, Institute of Semiconductor Technology, Hans-Sommer-Str. 66, 38106 Braunschweig, Germany
[7] TU Braunschweig, LENA Laboratory for Emerging Nanometrology, Langer Kamp 6a/b, 38106 Braunschweig, Germany
[8] Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

Received 30 January 2026 / Accepted 4 April 2026

Abstract. Hafnium oxide (HfO_2) is a high-index dielectric material of growing importance for optical coatings and meta-optical components operating from the ultraviolet (UV) to the visible spectral range. Its large bandgap, chemical stability, and compatibility with established semiconductor processes make it particularly attractive for applications requiring low absorption and precise thickness control. In this work, we investigate the optical material properties of HfO_2 thin films deposited by plasma-enhanced atomic layer deposition (PEALD). The complex refractive index is experimentally determined over a broad spectral range extending from the vacuum ultraviolet (140 nm) to the visible (600 nm) by combining spectroscopic ellipsometry, spectrophotometry, and infrared ellipsometry. A comprehensive dispersion model is employed to extract consistent optical constants and thickness values. The results highlight the suitability of PEALD-grown HfO_2 films for advanced optical and meta-optical applications.

Keywords: Hafniumoxide, meta-optics, ultra violet, refractive index, atomic layer deposition.

1 Introduction

Hafnium oxide (HfO_2) has attracted significant attention over the past two decades as a functional material in both microelectronics and photonics. In microelectronic devices, HfO_2 is well established as a high-permittivity (high-k) dielectric replacing SiO_2 in advanced metal-oxide-semiconductor field-effect transistors [1–3]. Beyond its electronic functionality, the optical properties of HfO_2, namely its high refractive index, wide optical bandgap, and high transparency from the ultraviolet to the infrared, render it attractive for optical coatings, interference filters, and dielectric metasurfaces [4–7]. In particular, the demand for optical components such as wire grid polarizer [8, 9], meta lenses [10] or diffractive optical elements [11] or grating couplers in photonic integrated circuits [12] operating in the ultraviolet and vacuum ultraviolet spectral ranges has increased substantially due to applications in spectroscopy – especially in quantum technologies [13–15] – space optics, and high-resolution microscopy. For these applications, materials must combine low optical absorption with high refractive index contrast and excellent environmental stability. HfO_2 fulfills these requirements and offers superior laser-induced damage threshold (LIDT) [16, 17] as well as resistance to radiation-induced damage [18, 19] and chemical degradation and corrosion compared to many alternative high-index oxides. A large variety of deposition techniques have been employed for HfO_2 thin films, such as electron beam evaporation [20, 21], magnetron sputtering [22], ion beam sputtering [23, 24], sol-gel [25–27] and atomic layer deposition (ALD) [28–33]. Among these, ALD is particularly suitable for the fabrication of HfO_2 thin films with precisely controlled thickness and excellent structural conformality. The self-limiting surface reactions inherent to ALD enables uniform coatings on both planar as well as three-dimensional nanostructured substrates, which is essential for emerging meta-optical concepts relying on subwavelength structuring.

* Corresponding author: thomas.siefke@uni-jena.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Numerous studies have been reported on the optical and structural properties of ALD-grown HfO_2, demonstrating its applicability in optical thin films across a wide spectra range, for instance as anti-reflection coatings at 266 nm, 355 nm, 532 nm and/or 1064 nm [31, 33], or as high-reflective mirrors at 355 nm and 532 nm [32]. Nevertheless, comprehensive experimental data on its optical constants over an extended spectral range, particularly toward the vacuum ultraviolet (VUV) remain comparatively scarce.

The present work aims to expand the accessible optical database by providing detailed optical properties of PEALD HfO_2 thin films down to a wavelength of 140 nm. By combining multiple complementary measurement techniques [34, 35], we extract reliable optical constants from the visible to the VUV range. The resulting dataset provides a solid foundation for the optical design and simulation of HfO_2-based coatings and meta-optical devices.

2 Thin film deposition

Hafnium oxide thin films are commonly deposited by atomic layer deposition using a variety of hafnium precursor chemistries in combination with either thermal or plasma-assisted oxidation steps. In thermal ALD, both inorganic precursors such as hafnium tetrachloride ($HfCl_4$) [36] and metal-organic precursors tetrakis(dimethylamido) hafnium (TDMAH), tetrakis(ethylmethylamido)hafnium (TEMAH) [37], or β-diketonate-based [38, 39] compounds have been widely employed together with water as the oxidant. While $HfCl_4$-based processes typically require higher deposition temperatures and may lead to halogen-related impurities, metal-organic precursors enable lower-temperature growth and improved process flexibility. In plasma-enhanced ALD (PEALD), the molecular oxidant is replaced by an oxygen plasma, which enhances ligand removal and film densification [40]. PEALD processes using TDMAH or TEMAH have been shown to produce dense, stoichiometric HfO_2 films with reduced impurity content at comparatively low substrate temperatures [30–32], making them especially attractive for optical coatings and nanophotonic applications.

Depending on the chosen precursor chemistry and oxidation scheme, the resulting films may exhibit significant differences in growth per cycle, density, impurity concentration, and ultimately optical material properties. The HfO_2 thin films investigated in this study were deposited at 100 °C using plasma-enhanced atomic layer deposition in an OpAL reactor (Oxford Instruments Plasma Technology). Tetrakis(dimethylamido)hafnium (TDMAH, $Hf[N(CH_3)_2]_4$) served as the metal precursor, while an oxygen plasma was employed as the oxidizing reactant. Each ALD cycle consisted of a TDMAH precursor pulse (0.4 s), followed by a purge (8 s) and pump-down (2 s) sequence to remove excess precursor and by-products, a gas stabilization period (3 s) followed by an oxygen plasma exposure (10 s) and a final purge (10 s). The details are summarized in Table 1.

The deposition was performed on fused silica substrates in order to enable transmission measurements over a wide spectral range. To ensure sufficient optical sensitivity in ellipsometry measurements while maintaining negligible interference from the substrate backside, a total film thickness of approximately 200 nm was targeted. The deposition process yielded smooth, homogeneous films with excellent macroscopic uniformity.

Table 1. ALD parameter for Hafnia deposition using TDMAH and oxygen plasma in an OPAL-tool.

Step	Parameter	Value
1	TDMAH pulse time	0.4 s
2	Purge time	8 s
3	Pump down time	2 s
4	Gas stabilization time	3 s
5	Oxygen plasma pulse	
	Time	10 s
	Flow	50 sccm
	Power	300 W
6	Purge time	10 s

3 Optical characterization methods

The optical properties of the HfO_2 films were characterized using a combination of spectroscopic ellipsometry and spectrophotometry to cover a broad spectral range. Vacuum ultraviolet and ultraviolet ellipsometric measurements were performed using a Jobin Yvon UVISEL2 ellipsometer at photon energies ranging from 1.5 to 8.7 eV at a fixed angle of incidence of 70°. Additional ellipsometric data in the range from 0.6 to 6.5 eV were acquired with a Jobin Yvon UVISEL ellipsometer using multiple angles of incidence between 55° and 75°. Infrared ellipsometry measurements covering the range from 300 to 6500 cm^{-1} were conducted using a Woollam IR-VASE ellipsometer, again employing variable angles of incidence. Complementary transmission and reflection spectra were recorded using a Perkin Elmer Lambda 1050 spectrophotometer and a Bruker Vertex 80v Fourier-transform infrared spectrometer. For reflectance measurements, illumination from both the film and substrate side was used to enhance sensitivity to interface effects. The experimental data were analyzed using the universal dispersion model implemented in the newAD2 software package [34]. All relevant parameters, including film thickness, dispersion coefficients, and interface roughness, were treated as free fitting parameters. This approach ensured a consistent description of the optical response across all spectral regions and measurement techniques.

4 Results and discussion

The combined optical analysis yielded film thickness values between 217 and 221 nm, depending on the specific measurement configuration and spot size. This small variation indicates a high degree of thickness uniformity across the

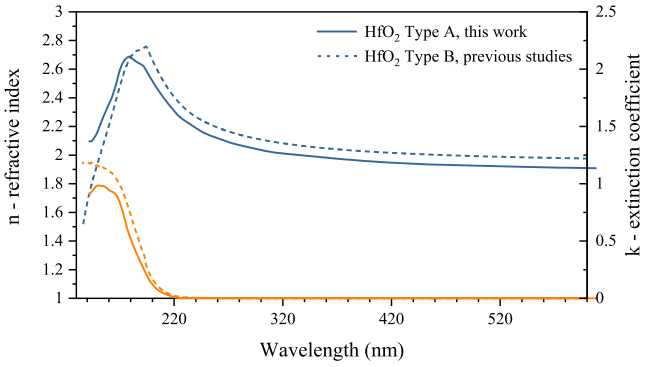


Fig. 1. Wavelength dependence of the refractive index of the fabricated hafnia thin films.

Table 2. Compared HfO2 thins films prepared by PEALD with different ALD tools.

	HfO_2 – Type A	HfO_2 – Type B
reference	this work; [29]	[33]
PEALD tool	OpAL	SILAYO-ICP330
precursors	TDMAH + O_2 plasma	TDMAH + O_2 plasma
substrate temperature	100 °C	100 °C
density	7.90 gcm-3	8.25 gcm-3
surface roughness	0.8 nm (XRR)	1.07 nm (AFM)

sample. The surface roughness of the top interface was determined to be approximately 1.51 nm (root mean square), with an autocorrelation length of about 5.3 nm. These values correspond to dense, smooth oxide layers produced with PEALD and are comparable to HfO_2 thin films on silicon reference samples deposited under similar conditions with 0.8 nm rms (XRR) [29]. The low surface roughness is critical for optical applications, as elevated topography would introduce scatter losses and degrade coating performance in the ultraviolet spectral range.

To evaluate film quality and optical functionality, the extracted refractive index was compared with a previous PEALD-HfO_2 study. The thin film examined in this work is designated Type A, while the comparison sample is designated Type B (compare Fig. 1 and Table 2). To minimize the influence of impurity contributions on the optical properties, we compare only processes performed with the same precursor (TDMAH) same oxidizing agent (O_2-plasma) and at the same deposition temperature (100 °C), since these are the main parameters affecting carbon and nitrogen levels [30]. The low deposition temperature was chosen to reduce crystallization in the thin films, even though lower temperatures often lead to increased carbon contamination because of incomplete precursor oxidation.

The refractive index of both HfO_2 thin films exhibits a pronounced wavelength dependence, with a normal dispersion i.e. high values in the visible and near-ultraviolet spectral range and a gradual decrease toward longer wavelengths. In the ultraviolet region, the refractive index increases significantly as the photon energy approaches the optical band edge, while absorption remains negligible over a wide spectral range down to 230 nm (compare band edge of 5.6 eV [33]). This behaviour is highly advantageous for optical coatings and metasurfaces requiring strong phase modulation with minimal optical losses.

While both HfO_2 thin films demonstrate excellent optical properties and comparable transparent regions, Type A and Type B show slight deviations in refractive index. These differences can be attributed to two distinct sources: differences in metrology and deposition conditions.

For Type A thin films, the universal dispersion model (UDM) was employed to extract optical constants [34]. In contrast, optical constants of Type B thin films were extracted from VUV transmittance and reflectance spectra using the Lorentz Calculator (LCalc) [41]. These different analytical methods can introduce systematic deviation, particularly in the VUV region where extinction is dominated by complex inter-band transitions. Nevertheless, both methods are sufficiently mature to provide reliable data and inherently capture the material response, including any influence of i.e. residual impurities.

Typically, different deposition parameters are the main source of such differences. The dominant factor separating the two film types are the different PEALD reactor configuration and associated plasma parameters, which are known to strongly influence HfO_2 film density and refractive index [31]. Type A films were deposited using an OpAL PEALD reactor with a remote Inductively Coupled Plasma (ICP) source, while Type B samples were produced on a SILAYO-ICP330 system (Sentech Instruments GmbH), equipped with a direct ICP configuration featuring a Planar Triple Spiral Antenna (PTSA) source. These distinct plasma architectures lead to markedly different energy input and ion bombardment characteristics. The SILAYO PTSA-ICP source delivers higher plasma power density and enhanced ion flux, enabling more effective energetic assistance during film growth. This results in denser films: Type B achieves 8.25 g/cm^3 compared to 7.90 g/cm^3 for Type A (Table 2). Film density is directly reflected in the refractive index; the denser Type B films exhibit higher n-values across the transparent spectral range. This density-dependent optical property improvement is a well-established phenomenon in oxide thin films and confirms that plasma parameter optimization is a critical control lever for tailoring optical performance in PEALD-HfO_2 systems.

X-ray diffraction measurements conducted at the deposition temperature (100 °C) indicate that both Type A and Type B films are X-ray amorphous (not shown here), consistent with low-temperature PEALD processes. However, X-ray amorphous coatings can contain nanocrystalline domains, short-range crystalline order, or point defects that influence optical properties. This may also explain the moderate deviations, especially observed below the absorption edge between the two film sets. The higher density of type B can encourage slight structural rearrangement, such as localized crystallization or structural densification, and affect the absorption profile below the band gap. Despite these structural subtleties, both films maintain low

absorption in the transparent region and preserve well-defined optical bandgaps, indicating that any nanocrystalline or defect-state contributions remain minimal and do not compromise optical performance.

The results of both HfO_2 thin films conclusively confirm that PEALD enables the fabrication of optically dense HfO_2 layers with properties suitable for advanced optical designs, including applications in ultraviolet meta-optics.

5 Conclusion

In summary, we have presented a comprehensive optical characterization of hafnium oxide thin films deposited by plasma-enhanced atomic layer deposition. By combining spectroscopic ellipsometry and spectrophotometry across a wide spectral range, reliable optical constants from the infrared to the vacuum ultraviolet were obtained. The films exhibit high refractive indices, low surface roughness, and excellent optical homogeneity. These findings underscore the suitability of ALD-grown HfO_2 for demanding optical applications, particularly in the ultraviolet spectral range and in nanostructured meta-optical systems. The presented dataset provides a valuable reference for optical modeling and design and supports the further development of HfO_2-based optical components.

Funding

The research has been supported by the Deutsche Forschungsgemeinschaft (DFG) (Emmy-Noether-Project SZ253/1-1, 287542364 and Collaborative Research Center (CRC/SFB) 1375), the European Space Agency (ESA) (Contract No. 4000109161/13/NL/RA), under Grant No. 13 N16897 GRADIENT, and BMFTR, Förderprogramm Fusion2040 – Forschung auf dem Weg zum Fusionskraftwerk, FKZ: 13F1009I, SHARP and ATIQ (Fkz: 13N16116), Cluster of excellence QuantumFrontiers (ExC-2123-390837967) as well as from the EMPIR programme 20IND04 ATMOC.

Conflicts of interest

The authors declare that they have no competing interests to report.

Data availability statement

The data that support the findings of this study will be made publicly available in the refractiveindex.info database (https://refractiveindex.info/) upon publication.

Author contribution statement

Thomas Siefke contributed to writing (original draft), formal analysis, funding acquisition, and resources. Kristin Gerold contributed to writing (original draft), visualization, formal analysis, and investigation. Svetlana Shestaeva contributed to writing (review and editing), formal analysis, and investigation. Pallabi Paul contributed to formal analysis, and investigation. Shawon Alam contributed to formal analysis, and investigation. Daniel Franta contributed to formal analysis and investigation. Adriana Szeghalmi contributed to funding acquisition and writing (review and editing). Sven Schröder contributed to writing (review and editing). Stefanie Kroker contributed to funding acquisition and writing (review and editing).

References

1 Yin YT, Huang CC, Chiu PH, Jiang Y Sen, Hoo JY, Chen MJ, High-quality HfO_2 high-K gate dielectrics deposited on highly oriented pyrolytic graphite via enhanced precursor atomic layer seeding, *ACS Appl. Electron. Mater.* **7**, 1943–1952 (2025). https://doi.org/10.1021/ACSAELM.4C02224.
2 Zhang S, Zhang T, Yu H, Li T, Li X, Cui C, et al., Wafer-scale high-κ HfO2 dielectric films with sub-5-Å equivalent oxide thickness for 2D MoS2 transistors, *Nat. Commun.* **17**, 1888 (2026). https://doi.org/10.1038/s41467-026-68584-0.
3 Choi JH, Mao Y, Chang JP, Development of hafnium based high-k materials – A review, *Mater. Sci. Eng. R. Rep.* **72**, 97–136 (2011). https://doi.org/10.1016/J.MSER.2010.12.001.
4 Chow R, Falabella S, Loomis GE, Rainer F, Stolz CJ, Kozlowski MR, Reactive evaporation of low-defect density hafnia, *Appl. Opt.* **32**, 5567 (1993). https://doi.org/10.1364/AO.32.005567.
5 Lehan JP, Mao Y, Bovard BG, Macleod HA, Optical and microstructural properties of hafnium dioxide thin films, *Thin Solid Films* **203**, 227–250 (1991). https://doi.org/10.1016/0040-6090(91)90131-G.
6 Balog M, Schieber M, Michman M, Patai S, Chemical vapor deposition and characterization of HfO2 films from organo-hafnium compounds, *Thin Solid Films* **41**, 247–259 (1977). https://doi.org/10.1016/0040-6090(77)90312-1.
7 Ritala M, Leskelä M, Niinistö L, Prohaska T, Friedbacher G, Grasserbauer M, Development of crystallinity and morphology in hafnium dioxide thin films grown by atomic layer epitaxy, *Thin Solid Films* **250**, 72–80 (1994). https://doi.org/10.1016/0040-6090(94)90168-6.
8 Stock C, Siefke T, Zeitner U, Metasurface-based patterned wave plates for VIS applications, *J Opt Soc Am B* **36**, D97 (2019). https://doi.org/10.1364/JOSAB.36.000D97.
9 Siefke T, Kley E-B, Tünnermann A, Kroker S, Design and fabrication of titanium dioxide wire grid polarizer for the far ultraviolet spectral range, in: *Proceedings of SPIE – The International Society for Optical Engineering*, vol. **9927**, edited by Campo EM, Dobisz EA, Eldada LA (2016). https://doi.org/10.1117/12.2237644.
10 Ossiander M, Meretska ML, Hampel HK, Lim SWD, Knefz N, Jauk T, et al., Extreme ultraviolet metalens by vacuum guiding, *Science* **1979**, 380, 59–63 (2023). https://doi.org/10.1126/SCIENCE.ADG6881.
11 Kaufmann J, Siefke T, Ronning C, Zeitner UD, *Fabrication of EUV gratings via ion irradiation* JW4A.15 (2024).
12 Sorokina A, Meyer AA, Grimpe CF, Du G, Sauer S, Jordan JE, Mehlstäubler TE, Kroker S, Numerical analysis of a high-efficiency dual-material waveguide-grating coupler system for ultraviolet photonics, *Opt. Continuum.* **4**, 10 (2025). https://doi.org/10.1364/OPTCON.568352.
13 Alombert-Goget G, Trichard F, Li H, Pezzani C, Silvestre M, Barthalay N, et al., Titanium distribution profiles obtained by luminescence and LIBS measurements on Ti: Al2O3 grown by Czochralski and Kyropoulos techniques, *Opt. Mater. (Amst)* **65**, 28–32 (2017). https://doi.org/10.1016/J.OPTMAT.2016.09.049.
14 Trichard F, Moncayo S, Devismes D, Pelascini F, Maurelli J, Feugier A, et al., Evaluation of a compact VUV spectrometer

for elemental imaging by laser-induced breakdown spectroscopy: application to mine core characterization, *J. Anal. At. Spectrom.* **32**, 1527–134 (2017). https://doi.org/10.1039/C7JA00185A.
15 Bassel L, Motto-Ros V, Trichard F, Pelascini F, Ammari F, Chapoulie R, et al., Laser-induced breakdown spectroscopy for elemental characterization of calcitic alterations on cave walls, *Environ. Sci. Pollut. Res.* **24**, 2197–2204 (2017). https://doi.org/10.1007/S11356-016-7468-5/FIGURES/5.
16 Alvisi M, Di Giulio M, Marrone SG, Perrone MR, Protopapa ML, Valentini A, et al., HfO2 films with high laser damage threshold, *Thin Solid Films* **358**, 250–258 (2000). https://doi.org/10.1016/S0040-6090(99)00690-2.
17 Stolz CJ, Thomas MD, Griffin AJ, BDS thin film damage competition, **7132**, 107–113 (2008). https://doi.org/10.1117/12.806287.
18 Fan Y, Liu J, Jiang J, Jiang LM, Ion radiation effects on the stability of hafnium oxide-based ferroelectric thin films: mechanisms and regulation, *IEEE Trans. Device Mater. Rel.* **25**, 314–322 (2025). https://doi.org/10.1109/TDMR.2025. 3550950.
19 Ding M, Liu X, Damage effect of hafnium oxide gate dielectric based metal-oxide-semiconductor structure under gamma-ray irradiation, *AIP Adv.* **11**, 65304 (2021). https://doi.org/10.1063/5.0048080/993914.
20 Kant S, Kumari N, Kumar M, Effect of deposition conditions on the morphological, optical, and corrosion behavior of electron beam evaporated high-performance HfO2 thin films, *Appl. Phys. A Mater. Sci. Process.* **131**, 631 (2025). https://doi.org/10.1007/s00339-025-08755-w.
21 Ramzan M, Rana AM, Ahmed E, Bhatti AS, Hafeez M, Ali A, et al., Optical description of HfO2/Al/HfO2 multilayer thin film devices, *Curr. Appl. Phys.* **14**, 1854–1860 (2014). https://doi.org/10.1016/J.CAP.2014.10.023.
22 Araiza J de J, Álvarez-Fraga L, Gago R, Sánchez O, Surface morphology and optical properties of hafnium oxide thin films produced by magnetron sputtering, *Materials* **16**, 5331 (2023) https://doi.org/10.3390/MA16155331.
23 Thielsch R, Gatto A, Kaiser N, Mechanical stress and thermal-elastic properties of oxide coatings for use in the deep-ultraviolet spectral region, *Appl. Opt.* **41**, 16 (2002) https://doi.org/10.1364/AO.41.003211.
24 Bundesmann C, Neumann H, Tutorial: The systematics of ion beam sputtering for deposition of thin films with tailored properties, *J. Appl. Phys.* **124**, 231102 (2018). https://doi.org/10.1063/1.5054046.
25 Jiang K, Anderson JT, Hoshino K, Li D, Wager JF, Keszler DA, Low-energy path to dense HfO_2 thin films with aqueous precursor, *Chem. Mater.* **23** (4), 945–952 (2011). https://doi.org/10.1021/cm102082j.
26 Nishide T, Honda S, Matsuura M, Ide M, Surface, structural and optical properties of sol-gel derived HfO_2 films, *Thin Solid Films* **371**, 61 (2000) https://doi.org/10.1016/S0040-6090(00)01010-5.
27 Zaharescu M, Teodorescu VS, Gartner M, Blanchin, MG, Barau A, Anastasescu M, Correlation between the method of preparation and the properties of the sol–gel HfO_2 thin films, *J. Non-Cryst. Solids* **354**, 409 (2008) https://doi.org/10.1016/j.jnoncrysol.2007.07.097.
28 Kull M, Piirsoo HM, Tarre A, Mändar H, Tamm A, Jõgiaas T, Hardness, modulus, and refractive index of plasma-assisted atomic-layer-deposited hafnium oxide thin films doped with aluminum oxide, *Nanomaterials* **13**, 1607 (2023). https://doi.org/10.3390/NANO13101607/S1.
29 Shestaeva S, Bingel A, Munzert P, Ghazaryan L, Patzig P, Tünnermann A, Szeghalmi A, Mechanical, structural, and optical properties of PEALD metallic oxides for optical applications, *Appl. Opt.* **56**, 4 (2017) https://doi.org/10.1364/AO.56.000C47.
30 Kim K-M, Jang JS, Yoon S-G, Yun J-Y, Chung N-K, Structural, optical and electrical properties of HfO_2 thin films deposited at low-temperature using plasma-enhanced atomic layer deposition, *Materials* **13**(9), 2008 (2020). https://doi.org/10.3390/ma13092008.
31 Lapteva M, Beladiya V, Riese S, Hanke P, Otto F, Fritz T, Schmitt P, Olaf Stenzel O, Andreas Tünnermann A, Szeghalmi A, Influence of temperature and plasma parameters on the properties of PEALD HfO_2, *Opt. Mater. Express* **11**, 1918–1942 (2021). https://doi.org/10.1364/OME.422156.
32 Beladiya V, Faraz T, Schmitt P, Munser AS, Schröder S, Riese S, Mühlig C, Schachtler D, Steger F, Botha R, Otto F, Fritz T, Helvoirt C, Kessels WMM, Gargouri H, Szeghalmi A, Plasma-enhanced atomic layer deposition of HfO_2 with substrate biasing: thin films for high-reflective mirrors, *ACS Appl. Mater. Interfaces* **14**, 14677−14692 (2022). https://doi.org/10.1021/acsami.1c21889.
33 Alam S, Paul P, Beladiya V, Schmitt P, Stenzel O, Trost M, Wilbrandt S, Mühlig C, Schröder S, Matthäus G, et al., Heterostructure Films of SiO_2 and HfO_2 for High-Power Laser Optics Prepared by Plasma-Enhanced Atomic Layer Deposition, *Coatings* **13**, 278 (2023). https://doi.org/10.3390/coatings13020278.
34 Franta D, Nečas D, Ohlídal I, Universal dispersion model for characterization of optical thin films over a wide spectral range: application to hafnia, *Appl. Opt.* **54**, 9108 (2015). https://doi.org/10.1364/AO.54.009108.
35 Siefke T, Kroker S, Pfeiffer K, Puffky O, Dietrich K, Franta D, et al., Materials Pushing the Application Limits of Wire Grid Polarizers further into the Deep Ultraviolet Spectral Range, *Adv. Opt. Mater.* **4**, 1780–1786 (2016). https://doi.org/10.1002/adom.201600250.
36 Puurunen RL, Analysis of hydroxyl group controlled atomic layer deposition of hafnium dioxide from hafnium tetrachloride and water, *J. Appl. Phys.* **95**, 4777–4786 (2004). https://doi.org/10.1063/1.1689732.
37 Triyoso DH, Hegde RI, White BE, Tobin PJ, Physical and electrical characteristics of atomic-layer-deposited hafnium dioxide formed using hafnium tetrachloride and tetrakis (ethylmethylaminohafnium), *J. Appl. Phys.* **97**, 124107 (2005). https://doi.org/10.1063/1.1947389.
38 Pasko SV, Hubert-Pfalzgraf LG, Abrutis A, Richard P, Bartasyte A, Kazlauskiene V, New sterically hindered Hf, Zr and Y β-diketonates as MOCVD precursors for oxide films, *J. Mater. Chem.* **14**, 1245–1251 (2004). https://doi.org/10.1039/B401052C.
39 Zherikova KV, Morozova NB, Zelenina LN, Sysoev SV, Chusova TP, Igumenov IK, Thermal properties of hafnium(IV) and zirconium(IV) β-diketonates, *J. Therm. Anal. Calorim.* **92**, 729–734 (2008). https://doi.org/10.1007/s10973-008-9027-x.
40 Martínez-Puente MA, et al., ALD and PEALD deposition of HfO2 and its effects on the nature of oxygen vacancies, *Mater. Sci. Eng. B* **285**, 115964 (2022). https://doi.org/10.1016/J.MSEB.2022.115964.
41 Stenzel O, Wilbrandt S, Friedrich KF, Kaiser N, Realistische Modellierung der NIR/VIS/UV-optischen Konstanten dünner optischer Schichten im Rahmen des Oszillatormodells, *VIP* **21**, 15–23 (2009). https://doi.org/10.1002/vipr.200900396.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 56
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026049
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

RESEARCH ARTICLE OPEN ACCESS

Benchmarking dual-polarization silicon nitride photonic integrated circuits for trapped-ion quantum technologies

Carl-Frederik Grimpe[1], Anastasiia Lüßmann-Sorokina[1,2,3], Guochun Du[1,*], Pragya Sah[4,5], Steffen Sauer[1,2,3], Elena Jordan[1], Rijil Thomas[4], Pascal Gehrmann[2,3], Maksim Lipkin[4,5], Stephan Suckow[4], Max C. Lemme[4,5], Stefanie Kroker[1,2,3], and Tanja E. Mehlstäubler[1,6,7]

[1] Physikalisch-Technische Bundesanstalt, Bundesallee 100, Braunschweig, 38116, Germany
[2] Technische Universität Braunschweig, Institute of Semiconductor Technology, Hans-Sommer-Str. 66, Braunschweig, 38106, Germany
[3] Laboratory for Emerging Nanometrology (LENA), Langer Kamp 6a/b, Braunschweig, 38106, Germany
[4] AMO GmbH, Otto-Blumenthal-Straße 25, 52074 Aachen, Germany
[5] Chair of Electronic Devices, RWTH Aachen University, Otto-Blumenthal Str. 25, Aachen 52074, Germany
[6] Leibniz Universität Hannover, Institut für Quantenoptik, Welfengarten 1, Hannover, 30167, Germany
[7] Leibniz Universität Hannover, Laboratorium für Nano- und Quantenengineering, Welfengarten 1, Hannover, 30167, Germany

Received 29 January 2026 / Accepted 26 May 2026

Abstract. Trapped ions are one of the most advanced platforms for quantum technologies, with applications ranging from quantum computing to precision timekeeping. A crucial step towards more compact and scalable systems involves integrating photonic integrated circuits (PICs) into surface ion traps to enable on-chip light delivery and optical addressing of individual ions. Currently, most implementations rely solely on transverse-electric (TE) mode grating couplers, where the emitted light is polarized in the plane of the chip. In this work, we design, fabricate and characterize key silicon nitride (Si_3N_4) PIC components required for scaling trapped-ion based quantum systems to multiple operating zones, including incoupling structures, splitters, and grating couplers that support both TE and transverse-magnetic (TM) modes with comparable optical losses. We benchmark the PIC at 760 nm, which is a typical wavelength for Yb^+-applications. The fabricated grating couplers enable the outcoupling of collimated free-space beams for both polarizations, exhibiting distinct emission angles. This dual-polarization capability gives more flexibility in polarization control and expands the accessible optical design space for trapped-ion quantum technologies.

Keywords: Grating coupler, Silicon nitride, Ion traps, Dual-polarization.

1 Introduction

Trapped ions are a leading platform for numerous quantum applications, including sensing [1, 2], communications [3, 4], timekeeping [5, 6], and computing [7, 8].

The integration of nanophotonics into surface ion trap platforms has emerged as a promising route toward scalable quantum computing architectures and quantum sensors [9–16]. Monolithically integrated waveguides (WGs) and grating couplers (GCs) enable stable, well-defined optical beam delivery with a small footprint, mitigating the alignment sensitivity of free-space optical systems. Importantly, these photonic components are compatible with CMOS processes, enabling wafer-scale fabrication [17].

Among available material platforms, silicon nitride (Si_3N_4) has been developed as a mature and versatile option for visible and near-infrared operation [18–24]. Si_3N_4 offers low propagation loss and high optical-power handling [18], together with CMOS-compatible fabrication suitable for both passive and nonlinear photonic devices. Advances in film-stress control, etch uniformity, and wafer-scale processing have improved reproducibility and yield [19, 21, 25], reinforcing Si_3N_4 as a leading dielectric platform for visible integrated quantum technologies.

To date, most Si_3N_4 photonic implementations in ion traps are optimized for transverse-electric (TE) polarization [9–11, 14, 16], where the electric field points parallel to the chip surface. The polarization of light determines the coupling to atomic transitions. Therefore, TE-optimized operation provides limited coupling directions, and placement options of the GC to deliver the correct polarization with

* Corresponding author: guochun.du@ptb.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

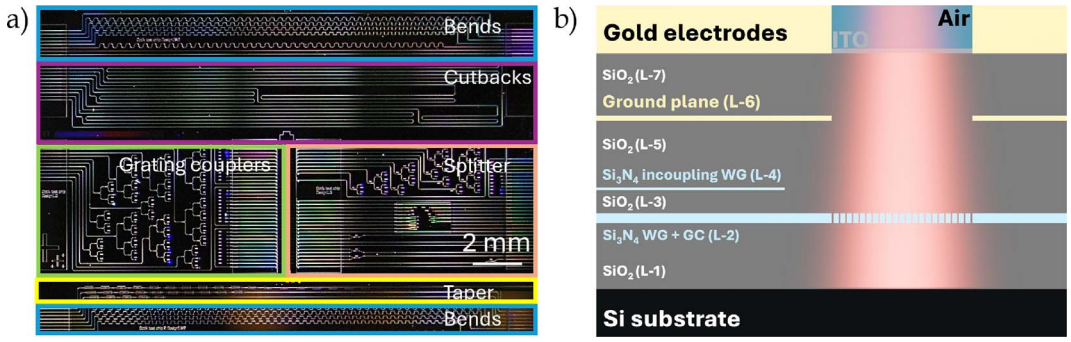


Fig. 1. a) Optical micrograph of the fabricated Si_3N_4 PIC, showing different functional regions: WG cutback structures (purple), bend and bend-shift loss structures (cyan), EBL–OL taper structures (yellow), grating coupler (green) and splitter test structures (orange). b) Targeted multilayer stack. From bottom to top: Si substrate, SiO_2 BOX (L-1), Si_3N_4 WG-GC (L-2) and WG for light incoupling layers (L-4) separated by a 1.25 µm thick SiO_2 buffer layer (L-3). Above these, an additional SiO_2 buffer layer is followed by a gold ground plane, 3 µm thick SiO_2 spacer, and a top gold layer used as the electrodes. Holes in the electrodes are covered with an ITO layer. The red cone illustrates the beam emitted by the GC.

respect to the ions quantization axis. Furthermore, TE-optimized operation limits accessible beam geometries. More recently, implementing transverse-magnetic (TM) modes, with the electric field oriented out of plane, have seen increased attention in integrated quantum and precision systems, such as Si_3N_4 resonator platforms for laser stabilization [26], and ion trap applications leveraging GCs for polarization gradient cooling [27].

We report the design, fabrication, and experimental characterization of the performance of a Si_3N_4 photonic integrated circuit (PIC) consisting of components such as incoupling structures, splitters, bends, taper sections, and GCs, that are key components for scaling trapped-ion based quantum technologies such as quantum computers and clocks to multiple operating zones. As an example, we use a wavelength of 760 nm for applications with trapped Yb^+ ions [28, 29]. The Yb^+ system underpins a wide range of state-of-the-art applications, including precision optical clocks [28, 29], trapped-ion quantum information processing [30, 31], and tests of fundamental physics [32–35]. We characterize the PIC with TE and TM modes and discuss how the observed polarization dependence may be leveraged to extend the photonic toolbox for ion trap quantum technology applications by offering additional design flexibility and enabling dual-polarization architectures.

2 TE and TM light guiding

2.1 Chip design and fabrication

The chip layout (Fig. 1a) comprises several functional sections, including WG cutback structures for the characterization of propagation losses, bend and bend-shift test structures, taper test structures, GC studies, and splitter test structures designed to operate at 760 nm. Together, these sections enable a comprehensive evaluation of polarization dependence, coupling efficiency, and loss mechanisms across the PIC. Each section is highlighted in a different color, corresponding to distinct device groups used to characterize optical performance and assess design variations.

The targeted multilayer trap stack comprises both dielectric and metallic layers, as depicted in Figure 1b. Starting from the silicon (Si) substrate, a thermally grown 3 µm thick buried oxide (BOX) silicon dioxide (SiO_2) layer acts as the lower cladding (L-1). The thickness of this bottom oxide layer is chosen to suppress optical leakage into the silicon substrate. On top of this, a 200 nm thick low-pressure chemical vapor deposition (LPCVD) Si_3N_4 layer (L-2) is deposited. An intermediate 1.25 µm thick low-temperature oxide (LTO) SiO_2 spacer (L-3) separates this layer from a 15 nm thick LPCVD Si_3N_4 layer (L-4). Owing to its larger optical mode, this thin Si_3N_4 layer is positioned at the top of the stack to maximize the separation from the silicon substrate and thereby minimize substrate leakage. An additional 3 µm thick LTO SiO_2 layer (L-5) forms the upper cladding of the PIC and is designed to be sufficiently thick to prevent optical leakage of the guided modes into the metallic layers introduced above. The targeted stack is completed by a 100 nm thick gold ground plane (L-6) introduced to provide electrical isolation and shielding, followed by a 3 µm thick SiO_2 buffer layer (L-7). A top gold layer forming the ion trap electrodes completes the envisioned ion trap architecture, with a thin indium tin oxide (ITO) layer included as a transparent conductive material covering holes in the electrodes that could otherwise affect the trapped ion [36].

For the fabrication of the PIC, a mix-and-match lithography approach was employed. Features larger than 500 nm, such as WGs and splitters, were realized using optical lithography (OL), whereas smaller features, such as incoupling tapers, were patterned with electron-beam lithography (EBL). Consequently, the Si_3N_4 GCs were fabricated in two sequential lithographic steps, the taper region defined by OL and the grating teeth subsequently patterned by EBL to achieve the required submicrometer periodicity and duty-cycle precision. The Si_3N_4 layer thickness of 200 nm was chosen as a compromise between compatibility with OL-based waveguide fabrication, GC coupling efficiency, and the minimum feature sizes achievable by EBL.

All photonic designs and simulations presented in this work are carried out with the full target layer stack in mind.

The experimental characterization reported here is performed on representative photonic test structures fabricated only up to layer L-5, before the deposition of the metallic layers.

2.2 Photonic component design

2.2.1 Alignment-tolerant incoupling

The incoupling region consists of three functional parts employing two vertically separated Si_3N_4 layers. The thinner upper layer serves for mode matching to the input fiber (see Fig. 2a), providing an expanded optical mode comparable to the fiber mode field diameter, while the thick-core lower layer serves as the routing layer for on-chip propagation. Light is transferred adiabatically between the two layers by a vertically tapered section (see Fig. 2b), which minimizes mode conversion loss through gradual effective index evolution [37]. Such dual-layer and vertically adiabatic coupling concepts are widely adopted in Si_3N_4 and Si-based photonic platforms to achieve efficient fiber-to-chip coupling, sub-dB loss, and relaxed alignment tolerances [38–46]. Finally, a second taper region is introduced to compensate for potential lateral misalignment between the EBL and OL process levels, ensuring robust and efficient coupling (see Fig. 2c).

The designs were carried out using Ansys Lumerical, employing the Finite Difference Eigenmode (FDE) solver for mode-overlap simulations to optimize fiber-to-chip coupling and eigenmode expansion (EME) simulations were used to analyze adiabatic transitions and lithography overlap sections. The simulated total incoupling loss is 1.95 dB for TE and 1.65 dB for TM, assuming perfect alignment across all fabrication steps. When considering a 500 nm lateral misalignment at each coupling stage, the total incoupling loss increases to 2.26 dB (TE) and 2.03 dB (TM), demonstrating the alignment-tolerance of the incoupling design. Resulting design parameters, coupling efficiencies, and simulations are summarized in the Supplementary material.

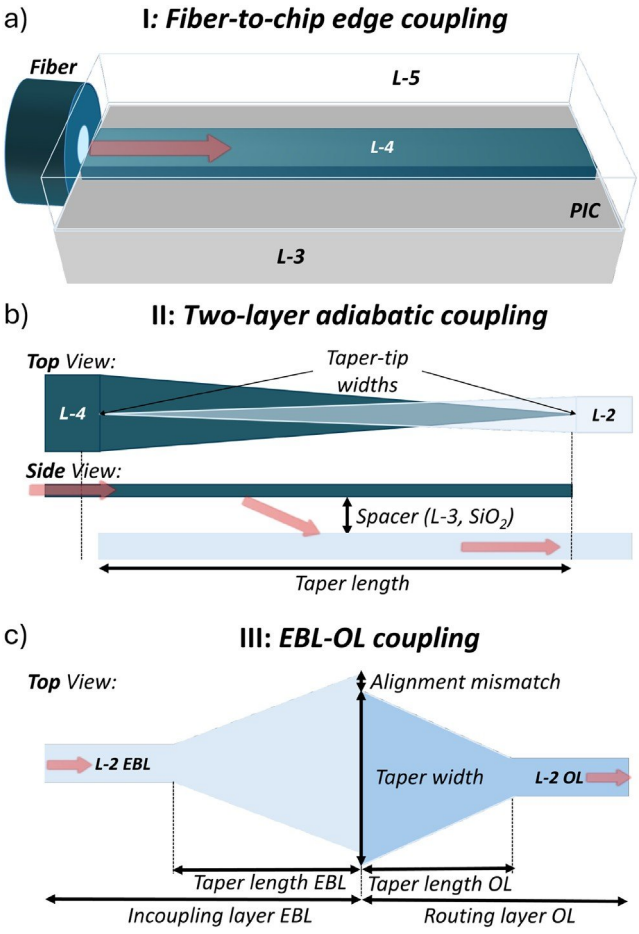


Fig. 2. Schematics of vertical fiber-to-chip coupling scheme. (a) Illustration of the fiber-to-chip edge coupling into the PIC. (b) Geometry of the two-layer adiabatic taper for efficient mode conversion between L-4 and L-2 waveguides. (c) Schematic of the EBL-to-OL coupling stage. The red arrows depict the light propagation through the structures.

2.2.2 Routing

To deliver light to the ions, it is routed across the chip via WGs, bends, and splitters before being directed into the GC.

2.2.2.1 Waveguides

WG modes were simulated using the FDE solver and the final waveguide geometry was then fixed to 200 nm × 520 nm (height × width) to secure single-mode operation and avoid higher-order contributions in the GC emission (see Supplementary material).

2.2.2.2 Bends

Bend structures with radii of 10, 20, 40, and 60 µm were designed, along with bend-shift structures optimized for TE mode matching inbetween bent and straight WGs using the FDE solver [47]. Bend shifts Λ of 30, 10, 10 nm were implemented for the 10, 20, 40 µm bends, respectively. They offer a promising approach to lowering mode-mismatch losses in compact bends, which in turn increases the flexibility of the photonic design [47].

2.2.2.3 Splitter

multimode interference (MMI) splitters are common on-chip components used for optical power splitting based on self-imaging effects in multimode waveguides [48]. The MMI splitters were designed using EME simulations for TE-polarized light (see Supplementary material). Each splitter features a core length of 38 µm, a core width of 6 µm, taper sections of 10 µm length with 1.6 µm width, and an output channel separation of 3.14 µm.

2.3 Experimental comparison of TE and TM losses

2.3.1 Setup

A fiber-coupled 760 nm distributed Bragg reflector (DBR) laser[1] was used as the light source for the characterization of the PIC. The input polarization was adjusted, and

[1]Thorlabs DBR760PN.

coupling was achieved using a single-mode (SM) polarization-maintaining (PM)[2] fiber aligned with a nanopositioner[3]. A powermeter was used to measure the optical power, while a pick-off on a photodiode was measured to monitor laser power fluctuations during the measurements. The polarization state was set using a polarimeter[4] (see Fig. 3a).

Experiments were performed with either a custom-made fiber array with PM fibers in a U-configuration, aligned to dedicated reference structures at the outer waveguide channels, or by coupling light through a single PM fiber of the array and collecting the output with a SM fiber[5] positioned behind the chip using a three-axis alignment system[6].

Optical measurements and characterization of the GC emission were performed using a commercial microscope system[7] equipped with a high-NA objective[8] (0.95 NA) and a CMOS camera[9]. The PIC is mounted on a tip, tilt and rotation stage[10] to minimize tilt-induced angle errors. In addition, the grating efficiency was determined by summing the pixel intensities of the outcoupled light on the CMOS camera and converting them to optical power using a pre-established calibration factor. Prior to these measurements, a reference calibration was performed using a fiber-coupled source to evaluate the objective transmission and to determine the conversion coefficient relating pixel intensity to optical power per exposure time.

2.3.2 Measurements

2.3.2.1 Propagation losses

Propagation losses were determined by measuring the transmitted power through waveguides of varying lengths arranged in a U-configuration. The exponential decay of transmitted power with length was fitted to extract the loss parameters, where the slope corresponds to the propagation loss and the intercept represents twice the combined incoupling and bend losses (see Fig. 3b). Measured propagation losses are (2.91 ± 0.33) dB/cm for the TE mode and (2.05 ± 0.18) dB/cm for the TM mode.

2.3.2.2 Taper losses

Ol-EBL taper losses have been characterized to be (1.17 ± 0.07) dB per taper for TE and (1.48 ± 0.04) dB per taper for TM by fitting the decay of the power transmission through waveguides with a different number of taper sections (see Fig. 3c). The OL-EBL taper losses are contributing to the total incoupling loss since they directly capture the loss introduced by the mix-and-match lithography approach.

2.3.2.3 Bend and bend-shift losses

Bend losses were determined by comparing waveguides containing bends with varying radii to a straight reference

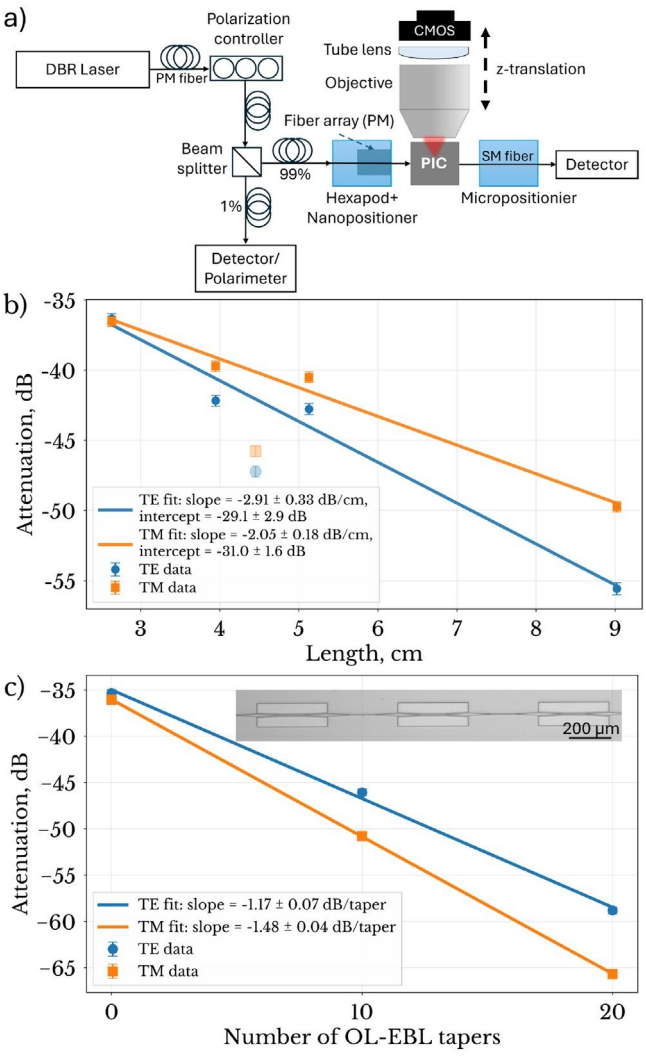


Fig. 3. a) Schematic illustration of the optical characterization setup. The red cone illustrates the beam emitted from the PIC. b) Cutback measurement to deduce the coupling and propagation losses for the TE and TM modes at a wavelength of 760 nm. Data points shown with reduced opacity were excluded from the fit. c) Characterization of the EBL–OL transition taper, showing the extracted taper loss. The inset displays the cascaded taper section.

waveguide. The extracted values include the intrinsic propagation loss within the curved sections. The total excess loss was then divided by the number of bends (200) to obtain the bend loss per bend.

The corresponding bend-loss measurements are shown in Figures 4a and 4b for TE- and TM-polarized light, respectively. For both polarizations, the bend loss strongly depends on the bend radius and is significantly reduced by the bend-shift design. For TE polarization (Fig. 4a), bends with radii of 20, 40, and 60 µm exhibit losses below 0.06 dB per 90° bend. The lowest loss is observed for the 20 µm bend, with approximately 0.04 dB, while the corresponding bend-shift design further reduces the loss to below 0.03 dB. At larger radii, the loss remains nearly constant, indicating that radiation and mode mismatch losses are already well

[2]PM780-HP.
[3]PI P-616 NanoCube® XYZ-Nanopositionierer.
[4]SK010PA-NIR.
[5]SM 780HP.
[6]NanoTrak® Auto-Alignment Controller.
[7]Olympus BXFM.
[8]Olympus MPLAPON50X.
[9]Hamamatsu ORCA-spark C11440-36U.
[10]Thorlabs TTR001.

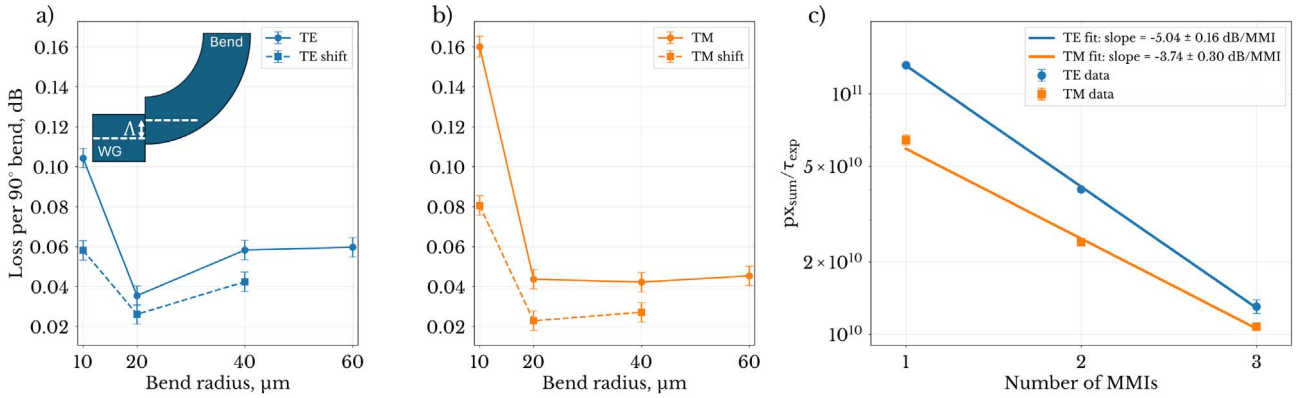


Fig. 4. Per-bend loss for the TE a) and TM b) polarization, comparing standard bends to laterally shifted bend geometries across different bend radii. In the shifted designs, Λ denotes the lateral offset, as illustrated in the inset of a). c) Measured attenuation of cascaded MMI splitters for the TE and TM modes. The linear fits provide the attenuation per MMI device and include additional contributions from bend and propagation losses.

suppressed in this regime. For TM polarization (Fig. 4b), a similar trend is observed. For bend radii of 20 µm and above, losses remain low, with values of approximately 0.04 dB for a 40 µm bend and as low as 0.02 dB for the 20 µm bend-shift design. As for TE polarization, the bend-shift geometry consistently outperforms the standard bend, confirming its effectiveness in reducing mode mismatch loss for both polarizations. At a bend radius of 10 µm, a pronounced increase in loss is observed for both polarizations, exceeding 0.10 dB per bend. In this regime, TE losses increase to approximately 0.11 dB, while TM losses reach about 0.16 dB, indicating the onset of radiation and mode mismatch losses due to insufficient mode confinement at tight bend radii.

2.3.2.4 MMI splitter

The splitter losses were evaluated using an MMI splitter tree terminated by grating couplers, each configuration containing a different number of MMIs. The total transmitted optical power was obtained from the summed pixel intensity px_{sum} per exposure time τ_{exp} monitored by the CMOS camera and plotted logarithmically as a function of the number of MMIs. The extracted slope represents the combined loss per MMI arm of one MMI, two bends, and a 150 µm straight waveguide section. From this, the individual MMI insertion loss was determined to be (1.88 ± 0.16) dB for TE and (0.62 ± 0.30) dB for TM (see Fig. 4c).

2.3.2.5 Incoupling

The incoupling loss was estimated in two ways: (i) from propagation-loss measurements by subtracting the contributions from bends and EBL–OL taper sections from the y-intercept, and (ii) from direct transmission measurements through a straight waveguide across the chip, with propagation losses subtracted. Method (i) yields single-sided incoupling losses of (13.0 ± 1.5) dB for TE and (14.08 ± 0.81) dB for TM, while method (ii) gives (13.89 ± 0.43) dB for TE and (14.71 ± 0.43) dB for TM (see Fig. 3b). These incoupling loss values correspond to the sum of the fiber-to-chip edge-coupling loss and the loss introduced by the two-layer adiabatic coupler. The total incoupling loss per facet is given by these contributions, including additional EBL-OL taper loss. The measured losses are significantly higher than those predicted by simulation, which may be attributed to fabrication-induced deviations in the taper-tip widths.

3 Design and characterization of the grating couplers

The operational principle of the structure is illustrated schematically in Figure 5a. Two beams corresponding to orthogonal TE and TM polarizations are emitted from a single GC and through an opening in the gold electrode layer. An optical microscope image of the fabricated device is presented in Figure 5b.

3.1 Design

The design optimization was performed using a combination of 2D and 3D finite-difference time-domain (FDTD) simulations via Ansys Lumerical. In the 2D model, the grating periods and duty cycles were varied to control the emission angle and far-field beam shape for the TE polarization [49–51]. To finalize the design, a 3D FDTD simulation was used to define the grating taper geometry and grating taper width and to extract the far-field intensity profile from the resulting structure. The GC design is primarily optimized for TE operation, resulting in a 30 µm-long grating section for efficient TE outcoupling, while simultaneously supporting a well-defined TM-polarized response, enabling a comprehensive assessment of polarization selectivity, angular separation of the emitted beams, efficiency and their quality.

3.2 Measurement

The outcoupled intensity distribution was recorded slice by slice in the xy-plane by translating the microscope system along the z-direction with a step size of 0.5 µm. At each z-slice, the intensity profile of the emitted beam was

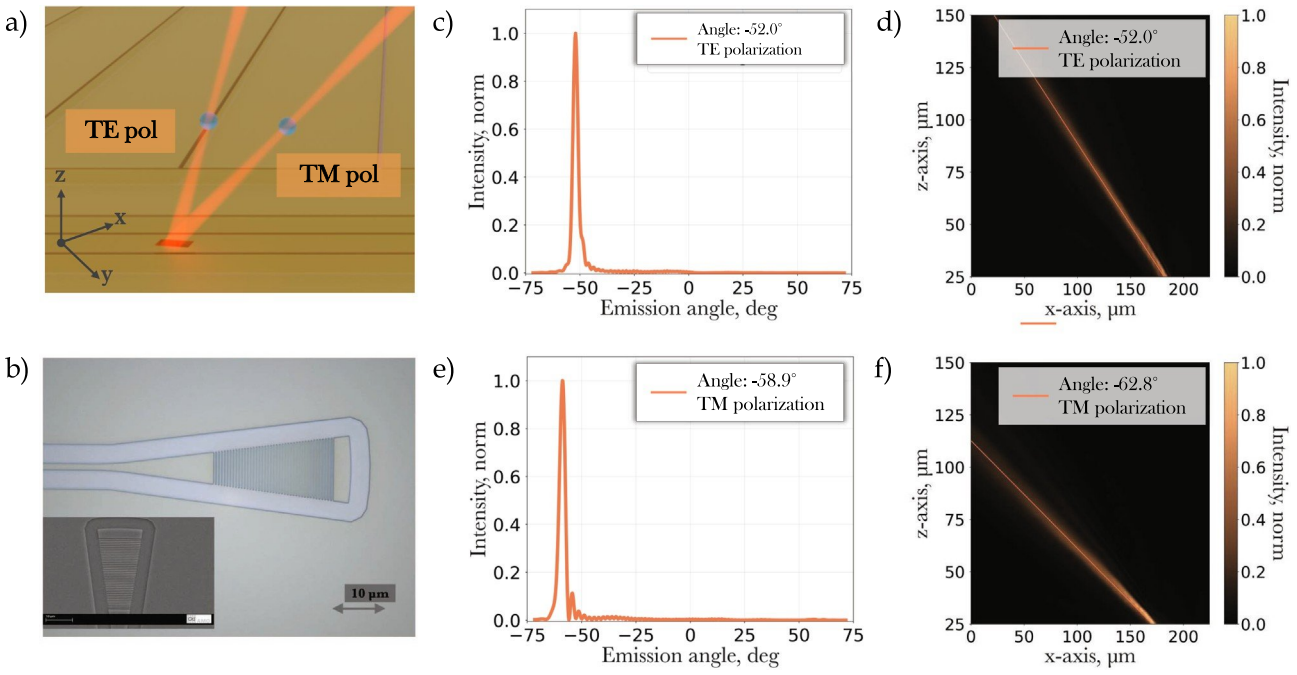


Fig. 5. a) Schematic illustration of the two beams emitted from the GC, corresponding to orthogonal TE and TM polarizations. b) Top-view optical microscope image of the fabricated Si_3N_4/SiO_2 GC. (c, e) Simulated far-field intensity distributions as a function of emission angle, evaluated in a plane 100 µm above the chip surface, for TE and TM polarizations, respectively. (d, f) Experimental intensity maps acquired from 25 µm to 150 µm above the chip surface were fitted to reconstruct the emission profiles of the GC for TE and TM polarizations, respectively.

analyzed by determining the position of maximum intensity. This maximum was taken as the beam position in the transverse plane. By repeating this procedure for all measured z-slices, we obtained a set of beam-center coordinates along the propagation direction. The emission angle was then extracted by performing a linear fit to these beam-center positions. The resulting angles for both polarizations were then compared with simulations, as shown in Figures 5c–5f. The measured emission angle of $(-52.0 \pm 0.4)°$ for TE polarization, agrees well with the simulated value of $-52.0°$.For TM polarization, the measured angle of $(-62.8 \pm 0.4)°$ shows a larger deviation from the simulated value of $-58.8°$. This discrepancy may be attributed to the increased sensitivity of steeper emission angles to fabrication tolerances. The measurement uncertainty accounts for the individual angle fit error per measurement ($\pm 0.1°$ for both TE and TM), the uncertainty in the z-translation calibration of the microscope system ($\pm 0.3°$ for both TE and TM) and the chip-tilt uncertainty ($\pm 0.2°$). Angles are reported with respect to the surface normal to ensure consistency between simulations and the experimental analysis. Negative angles correspond to a backward-emitting GC.

A detailed comparison between the simulated and measured far-field beam profiles presented in Figure 6, where all data sets were analyzed using the elliptical contour method and 1D Gaussian fits along the principal axes. The ellipse-based method provides a geometric estimate of the beam diameters along the principal axes of the far-field spot, while the Gaussian model fitting yields beam waists with one-standard-deviation uncertainties. The TE beam exhibits simulated waists of (7.96 ± 0.01) µm (13.09 ± 0.01) µm, compared to measured values of (9.07 ± 0.24) µm (10.79 ± 0.24) µm. For the TM polarization, the simulated beam waists are (12.29 ± 0.01) µm (19.17 ± 0.01) µm and the measured waists of (15.47 ± 0.24) µm (17.59 ± 0.24) µm. The simulated grating-coupler efficiencies are -3.61 dB for TE polarization and -8.79 dB for TM polarization. These values are in good agreement with the experimentally measured efficiencies of (-3.55 ± 0.58) dB for TE and (-9.02 ± 0.69) dB for TM, which were obtained by measuring the emitted free-space power relative to the power after the fiber array and subtracting the losses of the remaining PIC components.

4 Discussion

4.1 Propagation losses

Both TE- and TM-polarized light were guided in the fabricated Si_3N_4 waveguides. The measured propagation loss is 0.86 ± 0.38 dB/cm higher for the TE polarization than for the TM polarization. This behavior indicates that the propagation loss is dominated by sidewall roughness, as the TE mode exhibits a larger modal overlap with etched sidewalls. The measured losses remain higher than state-of-the-art Si_3N_4 waveguides of similar thickness, where values as low as 2 dB at 729 nm have been demonstrated [10]. This indicates that further optimization of the fabrication process is required. The TM propagation loss may be attributed to leakage into the ground plane or substrate, since their larger vertical mode extent demands sufficiently thick

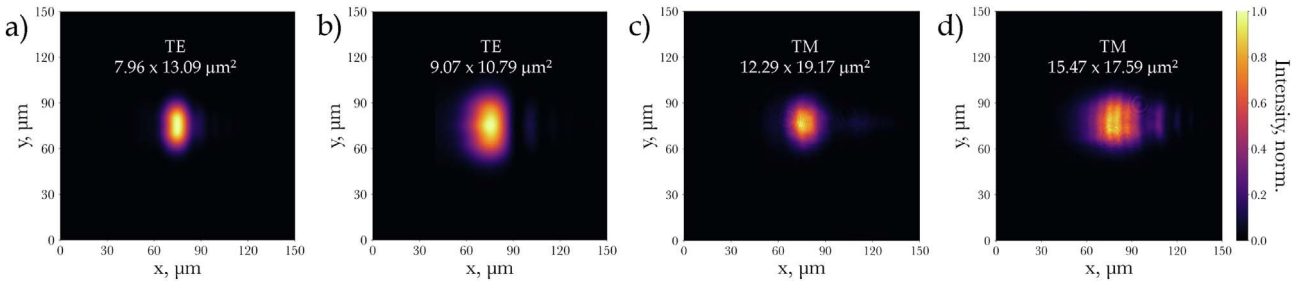


Fig. 6. Simulated (a, c) and measured (b, d) far-field intensity distributions of the laser-beam waist at 100 µm above the chip surface for TE (a, b) and TM polarization (c, d).

oxide layers to suppress. Transitioning to fused silica substrates would further mitigate substrate-related losses [52].

4.2 Incoupling

Incoupling losses of around 14 dB exceed the simulated values, yet both evaluation methods yielded similar absolute coupling efficiencies, with TE coupling slightly lower than TM. For comparison, visible-wavelength dual-layer Si_3N_4 couplers have been reported with losses of about 4 dB/facet [38], while single-layer couplers typically reach around 7.5–8.5 dB/facet [20]. The values obtained here are higher than those reported, suggesting that additional loss mechanisms are present in the fabricated devices. This behavior is plausibly linked to the two-layer edge-coupler architecture, which is highly sensitive to the taper-tip width of the thick-core Si_3N_4 layer [38]. A taper-tip wider than the optimized target value reduces the adiabaticity of the local index transition near the taper entrance and diminishes modal overlap (see Supplementary material). This indicates that future designs require an improved fabrication of the taper tip, or a thinner thick-core Si_3N_4 layer to reduce sensitivity to the taper-tip width. Measurements with the fiber array yielded coupling efficiencies comparable to the single-fiber configuration, indicating that alignment constraints introduce no significant additional penalty.

4.3 Bend and bend-shift losses

The measured bend losses of around 0.04 dB per 20-µm-radius bend match results reported in visible-wavelength Si_3N_4 experiments [20]. TM-polarized light exhibited noticeably higher loss for small-radius bends, a consequence of the 200 nm Si_3N_4 thickness, which was optimized for TE operation and provides insufficient confinement for TM, thereby increasing curvature-induced radiation loss. Such polarization-dependent bend performance aligns with earlier analyses of thin-core Si_3N_4 waveguides [19]. The shift of the bend-loss minima to larger radii – from 20 µm for TE to 40 µm for TM modes – is consistent with the lower propagation loss observed for TM.

Introducing bend-shift structures proved particularly beneficial for tight bends, where they reduced excess loss associated with mode mismatch at straight-to-bend transitions. This strategy is valuable for compact photonic-circuit layouts or applications requiring minimal stray light and crosstalk, as it enables dense routing without compromising optical efficiency.

4.4 MMI splitter

Significant deviations from simulated values were observed for TE MMI splitter insertion losses, while TM performance remained close to predictions. Enhanced TE losses can be attributed to stronger sensitivity to sidewall scattering within the multimode region, which perturbs the self-imaging condition. This observation is consistent with the higher TE propagation losses across the platform. TM modes, with weaker confinement, are less susceptible to such imperfections. Further optimization of the MMI length would likely reduce the residual insertion loss (see Supplementary material).

4.5 Taper losses

The OL-EBL taper transitions show good agreement with simulations for the TE polarization, while larger deviations are observed for the TM mode. These deviations may arise from lateral and longitudinal misalignment, as well as propagation losses, which are not accounted for in the simulations. Additionally, the observed behavior can be attributed to polarization-dependent changes in the modal field distribution induced by the taper. For the TE mode, tapering reduces the overlap with the etched sidewalls, thereby lowering sidewall-scattering loss. In contrast, for the TM mode, tapering may enhance propagation loss by increasing the overlap of the modal field with horizontal layer interfaces, thereby amplifying interface-related loss mechanisms. Extending the taper length or employing higher-order polynomial or spline profiles could further improve adiabaticity and reduce insertion loss.

4.6 Grating coupler

The measured grating emission profiles closely match the simulated far-field profiles in both overall shape and divergence. When quantified using Gaussian beam waists along the fitted principal axes, the average absolute deviation between simulated and measured data is 17–18% for both polarization regimes. The TM profiles exhibit a side lobe originating from the substrate back-reflected beam, which becomes visible because the lateral offset between the upward- and downward-reflected beams increases at larger angles. The measured emission angles are likewise consistent with the simulated values: for TE polarization, the emission angle matches the simulated value and agrees within experimental uncertainty, whereas for TM polarization, a deviation of approximately 4° is observed. These

Table 1. Comparison between simulated and measured losses for different PIC components for TE- and TM-polarized modes at 760 nm.

Component	TE loss		TM loss	
	Simulated	Measured	Simulated	Measured
Waveguide		2.91 ± 0.33 dB/cm		2.05 ± 0.18 dB/cm
Incoupling	1.95 dB	13.89 ± 0.43 dB	1.65 dB	14.71 ± 0.43 dB
EBL-OL taper	1.09 dB	1.17 ± 0.07 dB	0.92 dB	1.48 ± 0.04 dB
MMI splitter	0.77 dB	1.88 ± 0.16 dB	0.42 dB	0.62 ± 0.30 dB
GC	3.61 dB	3.55 ± 0.58 dB	8.79 dB	9.02 ± 0.69 dB

differences may be attributed to fabrication tolerances, such as slight variations in the etch depth of the grating and the duty cycle of the non-uniform grating design. They may also be attributed to deviations from the targeted layer stack, which arose from the omission of the upper metallic and dielectric layers in the devices that were characterized experimentally. The more pronounced deviation observed for the TM mode is consistent with the increased sensitivity of grating couplers emitting at larger angles, where small angular variations within the SiO_2 cladding translate into larger deviations in free space. Additional contributions may arise from the finite numerical aperture of the collection optics, minor phase-front distortions, and aberrations in the imaging system. Overall, the beam morphology, emission direction, and polarization-dependent divergence remain in good agreement, confirming that the fabricated grating accurately reproduces the simulated optical performance.

The GC efficiencies also align well with simulated predictions (see Tab. 1). Small deviations may originate from small fabrication-induced variations. The lower coupling efficiency observed for the TM grating coupler is attributed to its weaker grating coupling strength. This behavior of TM GCs may be beneficial in the design of highly focusing GCs or GCs that emit large beams, where precise control over the outcoupling strength is essential. The observed 8–10° angular separation between the TE and TM beams enables two distinct operation regimes: (i) individual addressing of ions or emitters at spatially separated locations when the beam waist is small, or (ii) polarization-state discrimination by monitoring the relative output intensity at the TE- and TM-specific emission angles. The angular separation corresponds to a spacing of roughly 14–18 µm at an ion height of 100 µm. This separation can be tuned through the waveguide geometry and grating duty cycle, which set the offset in effective index between the modes and thus control the resulting beam spacing. Accordingly, this effect is more pronounced in high-index platforms. This effect is fundamentally limited by the requirement that both modes satisfy the Bragg condition; consequently, at large backward emission angles, the other polarization may not yet be present.

5 Conclusion

We designed key Si_3N_4 PIC components and experimentally characterized their polarization-dependent performance for ion trap applications. The fabricated GCs support outcoupling for both polarizations and exhibit distinct emission angles.

This dual-polarization capability expands the accessible optical design space and may be used for multi-ion addressing through angular multiplexing, enabling more compact and versatile on-chip architectures while reducing the number of required GCs in integrated ion trap platforms. At the same time, the ability to excite both modes highlights the importance of robust polarization management, as inadvertent TM excitation can generate parasitic beams and increase stray light in the trapping region. Robust solutions such as TE–TM splitters or polarization converters can help to ensure a well-defined input state [53, 54]. The weaker coupling strength of the TM mode is promising for the design of highly focusing GCs for single-ion addressing in trapped-ion quantum processors or GCs that emit large beams for multi-ion addressing in multi-ion clocks.

Furthermore, since TM modes exhibit stronger sensitivity to substrate leakage, transitioning to fused-silica substrates [52] or employing thicker SiO_2 BOX cladding layers may be advantageous in future implementations. Taken together, TE–TM operation introduces a useful form of polarization diversity that complements spatial-multiplexing strategies explored in multimode photonics [55] and provides a promising pathway toward densely integrated, flexible, and scalable photonic interfaces for next-generation multi-zone trapped-ion quantum technologies.

Acknowledgments

We thank Markus Kromrey for taking images of the chips and Fatemeh Salahshoori for preparation of the chips.

Funding

The authors gratefully acknowledge the support by the Quantum Valley Lower Saxony, the BMFTR-project ATIQ (FKZ: 13N16116, 13N16126, 13N16130), Braunschweig International Graduate School of Metrology B-IGSM, the cluster of Excellence Quantum Frontiers (EXC-2123-390837967), and the Cluster of Excellence PhoenixD (EXC-2122).

Conflicts of interest

The authors declare that they have no competing interests to report.

Data availability statement

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Author contribution statement

Conceptualization, C.-F.G., G.D., S.Sa., A.L-S., E.J., T.M., S.K., S.Su.; Methodology, C.-F.G., G.D., A.L-S., P.G., S.Sa., M.L., E.J., T.M., S.K., S.Su; Software, C.-F.G., G.D., A.L-S., P.G., M.L.; Investigation, C.-F.G., G.D., A.L-S., E.J.; Resources, P.S., R.T.; Data Curation, C.-F.G., A.L-S., G.D.; Writing – Original Draft Preparation, C.-F.G., A.L-S., P.S.; Writing – Review & Editing, C.-F.G., G.D., A.L-S., E.J., S.Sa., S.Su., S.K., T.M.; Visualization, C.-F.G., A.L-S.; Supervision, M.Le., S.Su., S.K., T.M.; Funding Acquisition, E.J., M.Le., S.Su., S.K., T.M.

Supplementary material

S1: Incoupling simulations.
S2: Waveguide and MMI splitter simulations.
S3: GC efficiency simulation.
The supplementary material of this article is available at https://jeos.edpsciences.org/10.1051/jeos/2026049/olm.

References

1 Biercuk MJ et al., Ultrasensitive detection of force and displacement using trapped ions, *Nature Nanotechnology* **5**, 646 (2010). https://doi.org/10.1038/nnano.2010.165.
2 Harlander M et al., Trapped-ion probing of light-induced charging effects on dielectrics, *New Journal of Physics* **12**, 093035 (2010). https://doi.org/10.1088/1367-2630/12/9/093035.
3 Maunz P et al., Quantum interference of photon pairs from two remote trapped atomic ions, *Nature Physics* **3**, 538 (2007). https://doi.org/10.1038/nphys644.
4 Higginbottom DB et al., Pure single photons from a trapped atom source, *New Journal of Physics* **18**, 093038 (2016). https://doi.org/10.1088/1367-2630/18/9/093038.
5 Hausser HN et al., $^{115}In^{+}$–$^{172}Yb^{+}$ Coulomb crystal clock with 2.5×10(-18) systematic uncertainty, *Phys. Rev. Lett.* **134**, 023201 (2025). https://doi.org/10.1103/PhysRevLett.134.023201.
6 Lindvall T et al., Coordinated international comparisons between optical clocks connected via fiber and satellite links, *Optica* **12**, 843 (2025). https://doi.org/10.1364/OPTICA.561754.
7 Monroe C, Kim J, Scaling the ion trap quantum processor, *Science* **339**, 1164 (2013). https://doi.org/10.1126/science.1231298.
8 Moses SA et al., A race-track trapped-ion quantum processor, *Phys. Rev. X* **13**, 041052 (2023). https://doi.org/10.1103/PhysRevX.13.041052.
9 Mehta KK et al., Integrated optical addressing of an ion qubit, *Nature Nanotechnol.* **11**, 1066 (2016). https://doi.org/10.1038/nnano.2016.139.
10 Mehta KK et al., Integrated optical multi-ion quantum logic, *Nature* **586**, 533 (2020). https://doi.org/10.1038/s41586-020-2823-6.
11 Niffenegger RJ et al., Integrated multi-wavelength control of an ion qubit, *Nature* **586**, 538 (2020). https://doi.org/10.1038/s41586-020-2811-x.
12 Bruzewicz CD et al., Trapped-ion quantum computing: Progress and challenges, *Appl. Phys. Rev.* **6**, 021314 (2019). https://doi.org/10.1063/1.5088164.
13 Ivory M et al., Integrated optical addressing of a trapped Ytterbium ion, *Phys. Rev. X* **11**, 041033 (2021). https://doi.org/10.1103/PhysRevX.11.041033.
14 Kwon J et al., Multi-site integrated optical addressing of trapped ions, *Nature Commun.* **15**, 1 (2024). https://doi.org/10.1038/s41467-024-47882-5.
15 Knollmann FW et al., Integrated photonic structures for photon-mediated entanglement of trapped ions, *Opt. Quantum* **2**, 230 (2024). https://doi.org/10.1364/opticaq.522128.
16 Mordini C et al., Multizone trapped-ion qubit control in an integrated photonics QCCD device, *Phys. Rev. X* **15**, 011040 (2025). https://doi.org/10.1103/PhysRevX.15.011040.
17 Zesar A et al., Industrial Ion Trap Chips with Integrated Optics, in Witzens J, Poon J, Zimmermann L, Freude W (eds) *The 25th European Conference on Integrated Optics* (Springer Nature, Switzerland, Cham, 2024), p. 222. https://doi.org/10.1007/978-3-031-63378-2_37.
18 Blumenthal DJ et al., Silicon nitride in silicon photonics, *Proc. IEEE* **106**, 2209 (2018). https://doi.org/10.1109/JPROC.2018.2861576.
19 Muñoz P et al., Silicon nitride photonic integration platforms for visible, near-infrared and mid-infrared applications, *Sensors* **17**, 2088 (2017). https://doi.org/10.3390/s17092088.
20 Sacher WD et al., Visible-light silicon nitride waveguide devices and implantable neurophotonic probes on thinned 200 mm silicon wafers, *Opt. Express* **27**, 37400 (2019). https://doi.org/10.1364/OE.27.037400.
21 Leinse A, Zhang S, Heideman R, TriPleX: The versatile silicon nitride waveguide platform, in *2016 Progress in Electromagnetic Research Symposium (PIERS)*, p. 67 (2016). https://doi.org/10.1109/PIERS.2016.7734240.
22 Chauhan N et al., Ultra-low loss visible light waveguides for integrated atomic, molecular, and quantum photonics, *Opt. Express* **30**, 6960 (2022). https://doi.org/10.1364/OE.448938.
23 Smith JA et al., SiN foundry platform for high performance visible light integrated photonics, *Opt. Mater. Express* **13**, 458 (2023). https://doi.org/10.1364/OME.479871.
24 Blasco-Solvas M et al., Silicon nitride building blocks in the visible range of the spectrum, *J. Lightwave Technol.* **42**, 6019 (2024). https://doi.org/10.1364/JLT.42.006019.
25 Ji X et al., Efficient mass manufacturing of high-density, ultra-low-loss Si_3N_4 photonic integrated circuits, *Optica* **11**, 1397 (2024). https://doi.org/10.1364/optica.529673.
26 Isichenko A et al., Multi-laser stabilization with an atomic-disciplined photonic integrated resonator, arXiv:2509.09124 (2025). https://arxiv.org/abs/2509.09124.
27 Corsetti SM et al., Integrated-photonics-based systems for polarization-gradient cooling of trapped ions, *Light Sci. Appl.* **15**, 57 (2026). https://doi.org/10.1038/s41377-025-02094-4.
28 Huntemann N et al., High-accuracy optical clock based on the octupole transition in $^{171}Yb^{+}$, *Phys. Rev. Lett.* **108**, 090801 (2012). https://doi.org/10.1103/PhysRevLett.108.090801.
29 Tofful A et al., $^{171}Yb^{+}$ optical clock with 2.2×10^{-18} systematic uncertainty and absolute frequency measurements, *Metrologia* **61**, 045001 (2024). https://doi.org/10.1088/1681-7575/ad53cd.
30 Olmschenk S et al., Manipulation and detection of a trapped Yb^{+} hyperfine qubit, *Phys. Rev. A* **76**, 052314 (2007). https://doi.org/10.1103/PhysRevA.76.052314.
31 Piltz C et al., A trapped-ion-based quantum byte with 10^{-5} next-neighbour cross-talk, *Nature Commun.* **5**, 4679 (2014). https://doi.org/10.1038/ncomms5679.
32 Dreissen LS et al., Improved bounds on Lorentz violation from composite pulse Ramsey spectroscopy in a trapped ion,

Nature Commun. **13**, 7314 (2022). https://doi.org/10.1038/s41467-022-34818-0.
33 Hur J et al., Evidence of two-source King plot nonlinearity in spectroscopic search for new boson, *Phys. Rev. Lett.* **128**, 163201 (2022). https://doi.org/10.1103/PhysRevLett.128.163201.
34 Filzinger M et al., Ultralight dark matter search with space-time separated atomic clocks and cavities, *Phys. Rev. Lett.* **134**, 031001 (2025). https://doi.org/10.1103/PhysRevLett.134.031001.
35 Door M et al., Probing new bosons and nuclear structure with ytterbium isotope shifts, *Phys. Rev. Lett.* **134**, 063002 (2025). https://doi.org/10.1103/PhysRevLett.134.063002.
36 Du G, Jordan E, Mehlstäubler TE, Electric field distortions in surface ion traps with integrated nanophotonics, *Phys. Rev. Appl.* **25** (6) (2026) 064001. https://doi.org/10.1103/l1cn-28kv.
37 Sun X, Liu H-C, Yariv A, Adiabaticity criterion and the shortest adiabatic mode transformer in a coupled-waveguide system, *Opt. Lett.* **34**, 280 (2009). https://doi.org/10.1364/OL.34.000280.
38 Lin Y et al., Low-loss broadband bi-layer edge couplers for visible light, *Opt. Express* **29**, 34565 (2021). https://doi.org/10.1364/OE.435669.
39 Zhang W et al., Buried 3D silicon photonics spot-size convertors, in *2021 IEEE 17th International Conference on Group IV Photonics (GFP)*, p. 1 (2021). https://doi.org/10.1109/GFP51802.2021.9673849.
40 Dai D, He S, Tsang H-K, Bilevel mode converter between a silicon nanowire waveguide and a larger waveguide, *J. Lightwave Technol.* **24**, 2428 (2006). https://doi.org/10.1109/JLT.2006.874554.
41 Jia L et al., Efficient suspended coupler with loss less than 1.4 dB between Si-photonic waveguide and cleaved single mode fiber, *J. Lightwave Technol.* **36**, 239 (2018). https://doi.org/10.1109/JLT.2017.2779863.
42 Yi X et al., Asymmetric bi-level dual-core mode converter for high-efficiency and polarization-insensitive O-band fiber-chip edge coupling: breaking the critical size limitation, *Nanophotonics* **13**, 4149 (2024). https://doi.org/10.1515/nanoph-2024-0320.
43 Chen Z et al., Modeling dual-SiOxN thin-film edge coupler with ultra-low loss and large alignment tolerance, *Photonics* **12**, 136 (2025). https://doi.org/10.3390/photonics12020136.
44 He A et al., Low loss, large bandwidth fiber-chip edge couplers based on silicon-on-insulator platform, *J. Lightwave Technol.* **38**, 4780 (2020). https://doi.org/10.1109/JLT.2020.2995544.
45 Marchetti R et al., Coupling strategies for silicon photonics integrated chips, *Photon. Res.* **7**, 201 (2019). https://doi.org/10.1364/PRJ.7.000201.
46 Mu J et al., Monolithic integration of Al_2O_3 and Si_3N_4 toward double-layer active–passive platform, *IEEE J. Sel. Top. Quantum Electron.* **25**, 1 (2019). https://doi.org/10.1109/JSTQE.2019.2908559.
47 Sakai A, Go H, Baba T, Sharply bent optical waveguide silicon-on-insulator substrate, in *Physics and Simulation of Optoelectronic Devices IX (SPIE)* **4283**, pp. 610–618 (2001). https://doi.org/10.1117/12.432614.
48 Soldano LB, Pennings ECM, Optical multi-mode interference devices based on self-imaging: principles and applications, *J. Lightwave Technol.* **13**, 615 (1995). https://doi.org/10.1109/50.372474.
49 Taillaert D, Bienstman P, Baets R, Compact efficient broadband grating coupler for silicon-on-insulator waveguides, *Opt. Lett.* **29**, 2749 (2004). https://doi.org/10.1364/OL.29.002749.
50 Beck GJ, Home JP, Mehta KK, Grating design methodology for tailored free-space beam-forming, *J. Lightwave Technol.* **42**, 4939 (2024). https://doi.org/10.1109/JLT.2024.3381785.
51 Oton CJ, Long-working-distance grating coupler for integrated optical devices, *IEEE Photonics Journal* **8**, 1 (2016). https://doi.org/10.1109/JPHOT.2015.2511098.
52 Dietl M et al., Test and characterization of multilayer ion traps on fused silica, *Advanced Quantum Technologies* **8**, e00412 (2025). https://doi.org/10.1002/qute.202500412.
53 Hattori A et al., Integrated visible-light polarization rotators and splitters for atomic quantum systems, *Opt. Lett.* **49**, 1794 (2024). https://doi.org/10.1364/OL.509747.
54 Gallacher K et al., Silicon nitride waveguide polarization rotator and polarization beam splitter for chip-scale atomic systems, *APL Photon.* **7**, 046101 (2022). https://doi.org/10.1063/5.0077738.
55 Momenzadeh M et al., Individual trapped-ion addressing with adjoint-optimized multimode photonic circuits, *NPJ Nanophoton.* **3, 3** (2026). https://doi.org/10.1038/s44310-025-00102-4.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 41
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026033
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

Verifying optical simulations for infrared curing of powder coatings

Monika Kroneberger[1,★,*], Eva Stanik[2,3,★], and Thorsten Doehring[2]

[1] digitX GbR, Östlicher Dammweg 36, D-83052 Bruckmühl, Germany
[2] Technische Hochschule Aschaffenburg, D-63743 Aschaffenburg, Germany
[3] NETZ Technologietransferzentrum für nachhaltige Energien, D-63755 Alzenau, Germany

Received 29 December 2025 / Accepted 26 March 2026

Abstract. In line with the energy transition, it is desirable to replace fossil fuels in the curing process of industrial powder coatings. Infrared heating is the considered method here. Near Infra Red-emitter arrangement in the reflective heating chamber, their geometry and the dwell time of the components are crucial for a successful and efficient process. Process optimization has to consider the optical parameter variation of the powder coating during the procedure. Measurements in the NAPUBEST ("**Na**chhaltige **Pu**lver**bes**chichtungs-**T**echnologie": sustainable powder coating technology) prototype are compared with optical simulations of the setup to get the simulation parameters in agreement with reality and to provide a foundation for the layout of industrial processes.

Keywords: IR-curing, Optical simulation, Powder coating, Sustainable energy transition, Infrared heating.

Introduction

The recent energy crisis revealed the dependence of many manufacturing companies on fossil fuels. Industrial processes need to be transformed into sustainable alternatives. For powder coatings, which traditionally rely on gas or oil heating systems, short wave infrared radiation e.g. from infrared heating emitters can deliver the required process heat. It is an approved method for flat steel strip surfaces in the steel industry [1]. The strong dependence of the IR absorption on the geometry of the irradiated parts as well as on the colour and sort of the coating powder [2, 3] has to be considered before the deployment of this innovative heating. The aim of the NAPUBEST project (German abbreviation for "Nachhaltige Pulverbeschichtungs-Technologie" which means "Sustainable Powder Coating Technology") is to get a better understanding of these correlations. In our laboratory, we created a prototype of an octagonal IR oven for test measurements. The high reflectivity of silver-coated reflector sheets facilitates the concentration of infrared radiation inside the oven. On the other hand, the absorption of powder-coated samples in different paint colours has an impact on the heating efficiency. The arrangement of the Near Infra Red (NIR)-emitters in the reflective heating chamber, the geometry and the dwell time of the components are therefore crucial for a successful and efficient process. Process optimisation has to consider various optical and mechanical parameters for different coatings and components. Some of the parameters are also subject to change during the curing procedure.

The objective of this investigation is the verification of a simple simulation model with simplified optical parameters that can be easily adapted and used during production.

Simulation model and heating tests

The NAPUBEST heating chamber is made of Alanod Miro Silver sidewalls and a Heraeus Noblelight M220x511 NIR/SWIR heating module with 4 fast short wave emitters (1.5 kW, 1.4 µm peak wavelength each) [4]. The high reflectivity of the Alanod sheets guarantees a homogeneous radiance inside the chamber (except in front of the infrared emitters). For the optical simulations, the program FRED from Photon Engineering was used. The geometry of the heating chamber was exactly reproduced in the simulation model. The heating module was represented by the aluminum backside of the emitter-mount, the fused silica glass cover and the 4 emitter tubes. The filaments of the emitters were simulated as an emitting tube.

Since the heating chamber has no facility (yet) to hang the samples into the middle of the chamber they were placed onto the bottom sheet. The same position was used in the simulation (see Fig. 1). Free hanging as in conventional gas ovens will be beneficial since the light will reach

* Corresponding author: monika.kroneberger@digitx.de
★ These authors contributed equally to this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

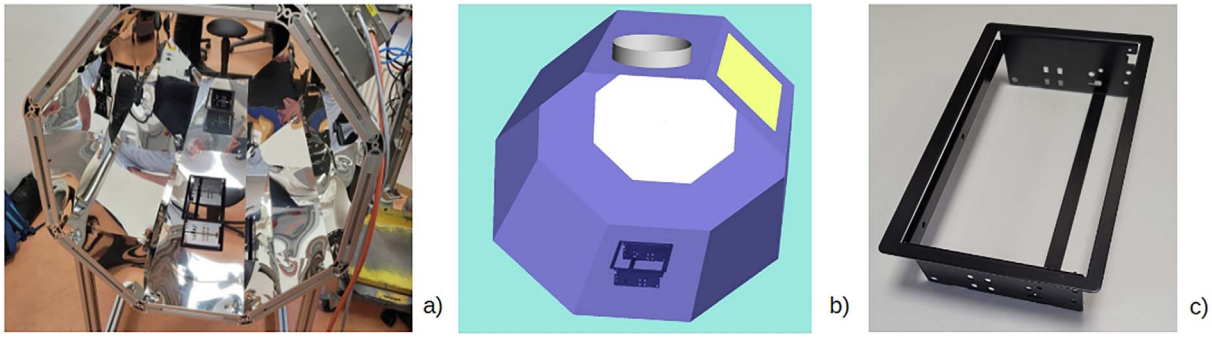


Figure 1. (a) NAPUBEST IR-oven with sample placed at the bottom of the chamber, (b) same arrangement in the simulation model of NAPUBEST IR-oven including the sample placed at the bottom of the chamber, (c) black coated sample.

all sides of the sample and is planned to be implemented as a next step.

For the verification, the CAD model of the specimen used in industrial fabrication and in the NAPUBEST oven was imported into FRED. The full surface of the specimen was used as the detector surface. The absorbed irradiance at the surface was simulated and the resulting irradiance map was exported as a CSV-file for post processing.

The optical properties of all materials in the heating chamber were either taken from vendor data (fused silica transmittance and refractive index, reflectivity of emitter mount and gold coating of emitter backside) or measured using an Ulbricht Sphere for all other materials (see Fig. 2). The black and white coating were found to be perfectly Lambertian: the HDR and DDR measurements had the same result, thus there is no specular fraction. For Alanod the two quantities have quite different values, resulting in a high specular fraction of the reflectivity, which is expected.

Measurements and vendor data were used to create material, coating and scatter data [5] in FRED for the peak wavelength of the emitters only. This simplification largely reduced the simulation time.

As a first test of the equipment and the correspondence between experiment and simulation, a black and a white coated sample were heated for 2 min each in the NAPUBEST heating chamber. The sample was placed into the oven chamber as shown in Figure 1 and the chamber was closed by the Alanod-covered front lid. All four emitters were running with full power for 2 min before the lid was re-opened and the temperature distribution was captured by an infrared camera 0, 1 and 2 min later (Testo 883, 640 × 480 pixel). The corresponding simulation takes into account four emitters at full power and the respective BSDF of the coatings.

Results

The temperatures of the specimen after 2 min of infrared irradiation were measured with the Testo infrared camera. Especial care had to be taken to not measure heat reflections of the lamps or of hot spots which get reflected on colder parts of the sample. The reflecting sheets of the heating chamber create a mirror image of the sample with the advantage of making the down facing surfaces of the sample visible.

The simulation results were displayed in an in-house developed software allowing the readout of surface irradiance point or linewise. The results for maximum and minimum temperature as well as the corresponding irradiance are shown in Table 1. Regarding ratio and over all distribution, the measurement and simulation are in good accordance (see Fig. 3). As expected, the temperature of the black sample was much higher than that of the white sample after the same dwell time in the oven due to the higher infrared absorption.

The temperature distribution though showed differences compared to the irradiance distribution (see Fig. 4) that we attribute to heat flow and heat transfer in the system, e.g. at the point of support, the temperature is lower than the mean surface temperature while the irradiance is more or less homogenously distributed over the side surface of the sample.

Discussion

Earlier simulations with a simple T-shaped sample [6] showed that inhomogeneities in irradiance at the sample surface can be prevented by illuminating the sample from different positions. It allowed to reduce the max/min ratio of the irradiance from 4.2 to 1.9 for the T-shaped sample. In the real world this means to turn the sample in the oven and/or install multiple heating modules in the oven.

The optical parameters of cured powder coating can be measured easily with an Ulbricht sphere and converted into simplified scatter models. The results of this investigation show that this is sufficient to map the simulations to the measurements. Heating results strongly depend on the coating colour and sample geometry as expected. The thermal effects like thermal conduction exceed the expected magnitude and need to be taken into account.

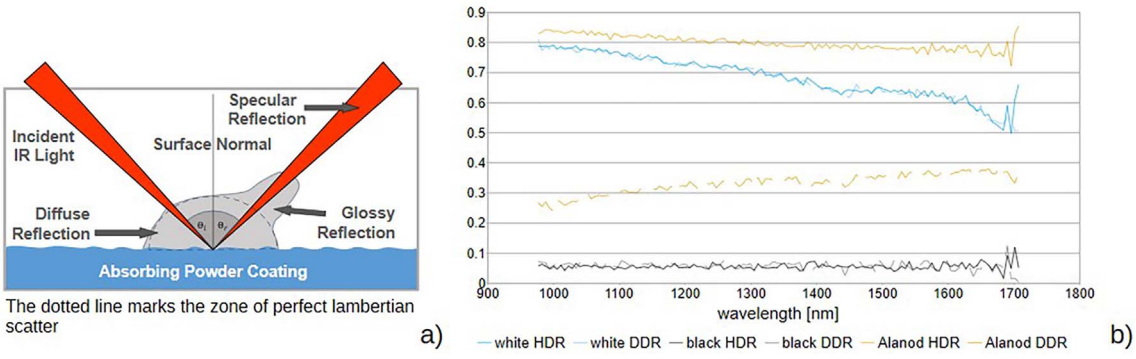


Figure 2. (a) Description of different parts of reflection on a surface, (b) comparison of HDR (Hemispherical Directional Reflectivity) and DDR (Diffuse Hemispherical Directional Reflectivity) measurement in an Ulbricht Sphere of cured white and black powder coating and Alanod sheet. HDR and DDR are identical for the two coatings indicating a Lambertian scatter whereas the two are quite different for the high specular Alanod.

Table 1. Results for maximum/minimum values of simulated irradiance and measured temperature of the white and black sample (Values in the holes are excluded).

	Max irradiance	Min irradiance	Ratio max/min	Max temperature	Min temperature	Ratio max/min
	Simulated values			Measured values		
White sample	3200 W/m^2	2300 W/m^2	1.39	152 °C	115 °C	1.32
Black sample	7500 W/m^2	4500 W/m^2	1.67	266 °C	190 °C	1.40

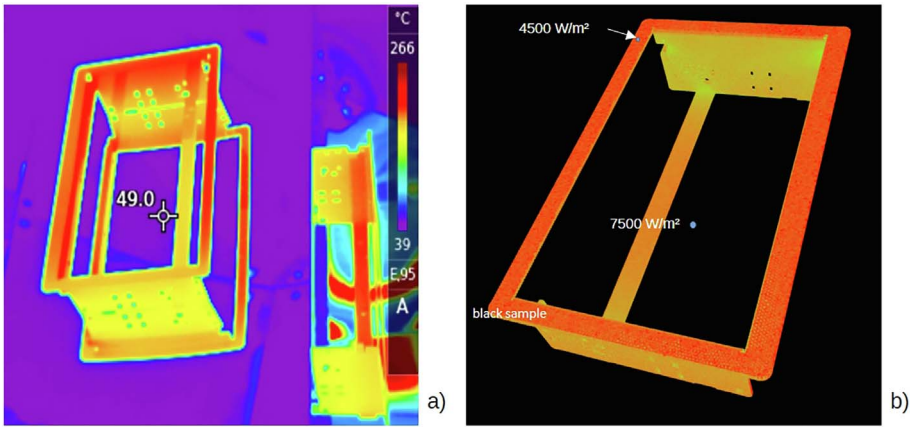


Figure 3. Simulation (b) and heating results (a) for the black sample. The ratio of temperatures matches the ratio of irradiances at the main surfaces of the sample. The red color corresponds here to higher temperatures while yellow/green resembles lower temperatures.

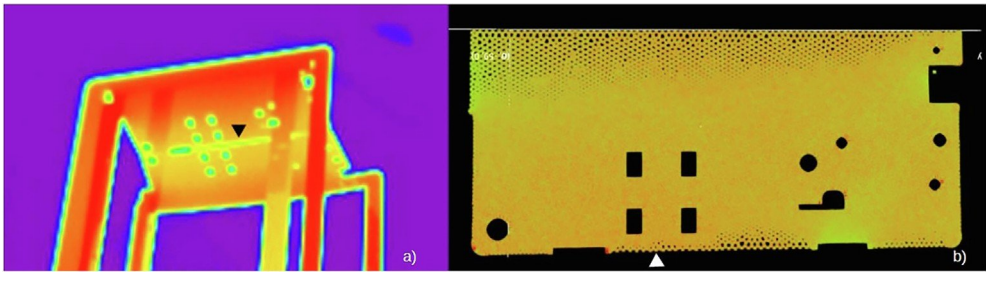


Figure 4. Detailed results of simulation and infrared camera show differences due to heat flow inside the sample. The side wall in the IR-picture (a) shows a heat sink at the bottom: heat flow to Alanod sheets. The pure ray trace simulation (b) does not show a lower irradiance at the bottom of the side surface.

Conclusion

These first results show that it is possible to predict the heating of complicated geometries by IR irradiation with sufficient precision. The simulation model with simplified scatter models for the powder coating delivers accurate results to predict the required heating power. Differences between the temperature distribution in the thermal images and the irradiance distributions are evidence for the need to include thermal effects into the model for more accurate predictions.

The next challenge will be to create a process and as a long term objective a tool which helps to define irradiation power sequences that lead to a more homogenous temperature distribution for easy and fast use at the production line.

Outlook

Next steps will be the investigation of different heating profiles with periods of reduced power to keep the temperature level after initial heating and to verify sufficient temperature homogeneity at the sample surface. This is vital for the application in powder coating with respect to the narrow curing process window of industrial powders. An additional thermal simulation using the results of the optical simulation as input will be carried out to include heat flow inside the sample and convection. When feasibility of the process in principle has proven, a larger research project will be initialised to investigate processes during the curing of the powder and possible impacts of the coating colour on the process.

Glossary

NIR	Near Infra Red
SWIR	Short Wave Infra Red
NAPUBEST	Nachhaltige Pulverbeschichtungs-Technologie: sustainable powder coating technology
NETZ	Technology Transfer Center for Sustainable Energies
BSDF	Bidirectional Scatter Distribution Function
FRED	Optical simulation software for non sequential ray tracing by Photon Engineering
HDR/THR	Hemispherical Directional Reflectivity (includes specular reflection) / Total Hemispherical Reflection
DDR/DHR	Diffuse Hemispherical Directional Reflectivity (excludes specular reflection) / Diffuse Hemispherical Reflection

Funding

This research was funded by The Bavarian State Ministry of Science and the Arts within the scope of the transfer initiative Hightech Transfer Bayern. The Bavarian State Ministry of Science and the Arts has granted start-up funding of € 5 million to Aschaffenburg University of Applied Sciences to establish a technology transfer centre for sustainable energy (NETZ) in Alzenau / Germany, together with regional companies and municipal partners.

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

The research data are available on request from the authors.

Author contribution statement

Conceptualization, T.D., E.S. and M.K..; Methodology, T.D., E.S. and M.K..; Validation, T.D., E.S. and M.K.; Formal Analysis, M. K.; Writing – Original Draft Preparation, M.K.; Writing – Review & Editing, T.D., E.S..; Visualization, M.K.; Project Administration, M.K.; T.D.; Funding Acquisition, T.D.

References

1 Stewart A.L., et al., Characterization of infrared absorption by powder coatings on steel panels. *J. Coat. Technol.* **71**, 71–84 (1999). https://doi.org/10.1007/BF02697891.
2 Barletta M., Flash IR pre-curing of the decorative layer in metal-flake powder coatings. *J. Prog. Org. Coat.* **72**(3), 498–510 (2011). https://doi.org/10.1016/j.porgcoat.2011.06.009.
3 von Horstig M.-W. et al., A perspective on innovative drying methods for energy-efficient solvent-based production of lithium-ion battery electrodes, *Energy Technol.* **10** (2022). 2200689. https://doi.org/10.1002/ente.202200689.
4 Stanik E. et al., Infrared heating for a sustainable powder coating technology, *DGaO Proc.* 2025.
5 Fest E.C. (2013). *Stray light analysis and control.* https://doi.org/10.1117/3.1000980.
6 Kroneberger M. et al., Optik-Simulation für die Energiewende am Beispiel eines industriellen Infrarot-Heizofens, *DGaO Proc.* 2025.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 34
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026039
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

SHORT COMMUNICATION OPEN ACCESS

Monkeypox virus (MPXV) neutralizing antibodies diagnosis with spectral-phase surface plasmon resonance 3D imaging biosensor

Chi Lok Wong*

Department of Electrical Engineering, Chang Gung University, Taiwan R.O.C

Received 9 December 2025 / Accepted 12 April 2026

Abstract. In this paper, we present the detection of Monkeypox virus (MPXV) neutralizing antibodies with a spectral-phase surface plasmon resonance (SPR) 3D imaging biosensor. To the best of our knowledge, this is the first demonstration of SPR imaging for MPXV neutralizing antibodies detection in the literature. The imaging biosensor provides three-dimensional molecular binding information in the *x*-, *y*- and time dimensions, enables real-time multiplex molecular bindings detection. In the optical design, a plasmonic sensing surface is placed in between a pair of polarizers with perpendicular orientation to block light transmission. At plasmonic resonance, an additional phase value is introduced for p-polarized light, while s-polarized light remains unchanged. This rotates the polarization ellipse and enables light transmission at specific plasmonic resonance wavelengths, generating the spectral-phase SPR image. In refractive index (RI) measurements, the sensor RI sensitivity was calculated to be 1191.2 Hue units/RIU. The biosensor was further applied to detect MPXV neutralizing antibodies with a 4 × 4 gold biosensor array immobilized with specific MPXV proteins and bovine serum albumin as control. The limit of detection was found to be 0.27 ng/μL with detection time of less than 16 min. This value is significantly lower than the neutralizing antibodies level (3.0 ng/μL) stimulated by MPXV vaccine reported in previous study.

Keywords: Monkeypox virus, MPXV, Neutralizing antibodies, SPR imaging, Biosensor, Spectral-phase.

1 Introduction

According to World Health Organization (WHO), an outbreak of Monkeypox happened in May 2022 and rapidly spread across Africa countries, Europe, the Americas and Southeast Asia. Monkeypox cases have been reported in over 120 countries worldwide. Between January 2022 to August 2024, cumulative infection cases reached 1 million and 220 deaths was reported. The Monkeypox virus (MPXV) belongs to the orthopoxvirus genus within the poxviridae family of vaccinia virus, variola virus and cowpox virus [1]. The diagnosis of MPXV neutralizing antibodies in patients contributes not only to vaccination development and efficacy evaluation, but also to governmental public health planning during outbreak. Current detection methods include Plaque-Reduction Neutralization Test (PRNT), Focus Reduction Neutralization Test (FRNT) and Enzyme-Linked Immunosorbent Assay (ELISA) [2]. However, PRNT and FRNT suffer from long detection time of 3–5 days and 1–2 days, respectively, while ELISA involves complex labeling and sandwich detection process [2]. Furthermore, trained staffs and centralized laboratory facilities are required, which are often lacking in Monkeypox outbreak areas, such as The Democratic Republic of the Congo and other Africa countries.

In this paper, we propose a spectral-phase surface plasmon resonance (SPR) 3D imaging biosensor for rapid MPXV neutralizing antibodies diagnosis. The biosensor provides rapid detection results in 16 min without requiring any labeling process. Furthermore, it is a potential point-of-care diagnostic tool, which contributes to patient care in developing countries experiencing Monkeypox outbreaks. In the optical design, the plasmonic sensor head is placed between two crossed polarizers for light blocking. At plasmonic resonance wavelength, a differential phase response is introduced between p- and s-polarized light [3], which rotates the polarization ellipse and allows light at resonance wavelengths [4] to pass through, resulting in a spectral-phase plasmonic color image. In addition, the excitation beam is expanded to cover a two-dimensional (2D) gold sensor spots array (4 × 4 elements). As video data are captured over the sensor surface, the biosensor can provide three-dimensional (3D) molecular binding information in the x-, y- and time dimensions, which enables real-time acquisition of multiplex molecular binding curves for

* Corresponding author: clwong@mail.cgu.edu.tw; davewongastar@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

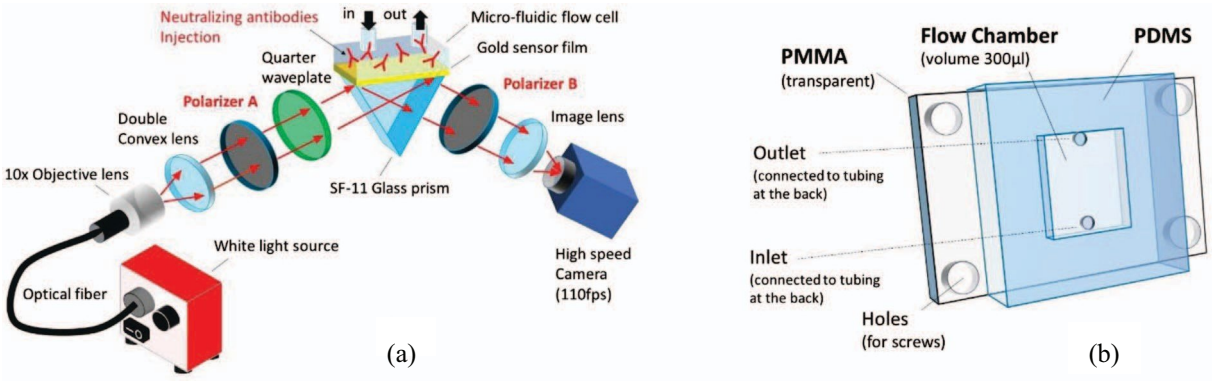


Fig. 1. (a) Optical scheme of the spectral-phase SPR 3D imaging biosensor. (b) Diagram of flow-cell.

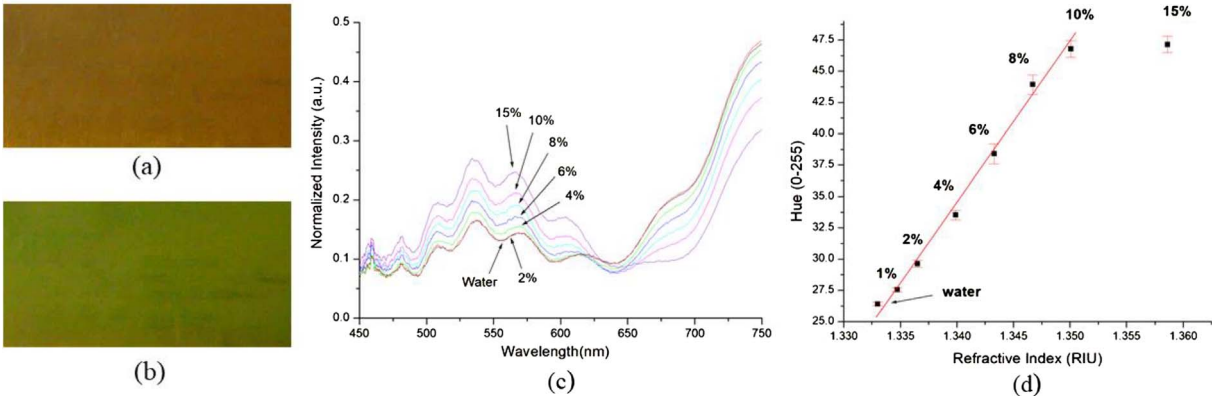


Fig. 2. Spectral-phase SPR images: (a) water, (b) 15% NaCl solution. (c) Reflection spectra for different concentrations of NaCl solutions ranged from 0% (water) to 15%. (d) Sensor response curve for different reflective index samples.

MPXV protein-neutralizing antibody interactions. Compared to the spectra-phase SPR imaging works we previously reported in [5, 6], this paper contributes to the first demonstration of MPXV neutralizing antibody detection using SPR imaging in the literature [7], while bovine serum albumin (BSA) and COVID-19 antibody detection were reported in [5, 6] respectively.

2 Material and methods

The experimental optical setup was constructed as described in Figure 1a. A fiber coupled halogen lamp was used as the broadband light source, which covers the spectral range of 350–1800 nm (Thorlab, model:OSL2). Subsequently, a telescope optical system consisting of a 10× objective lens (Olympus, model:10× Olympus plan achromat objective) and a double convex achromatic lens (Edmund Optics, model:50 mm dia. × 75 mm) expanded the beam to approximately 30 mm in diameter. A 4 × 4 gold sensor array was sputtered onto an SF-18 equilateral glass prism for plasmonic biosensing. It is integrated with a polydimethylsiloxane (PDMS) based microfluidic flow cell for sample injection. The plasmonic sensing element was placed between polarizer A and B with near perpendicular polarization angles to block light transmission. However, at plasmonic resonance, additional light phase difference was introduced between the p- and s-polarized light, that rotated the polarization ellipse and allowed light to pass through. As plasmonic excitation is a wavelength dependent process, a plasmonic spectral profile was encoded into the spectral-phase SPR image. Finally, a high-speed CMOS camera (Basler, model: acA1440-220uc) was used to capture sequence of plasmonic images up to 113 fps (corresponding to a temporal resolution of 0.0088 s), which were further analyzed using a custom-built MATLAB program. Spectral-phase SPR imaging enabled 2D detection. At video format, the sequence of SPR images can provide 3D multiplexed molecular bindings curves in the time domain. Figure 1b shows structure of the flow cell, where a flow chamber was patterned at the center of a PDMS layer. The size of the chamber was approximately 1.5 × 1.5 cm^2 and with the fluidic volume of ~300 µL. During detection, a 5 mL syringe was connected to the flow cell through tubing for static sample injection.

MPXV L1R neutralizing antibodies (40889-T62, 1 µg/µL), L1R protein (40889-V07E, 0.25 µg/µL) and control BSA (810651) protein samples were purchased from

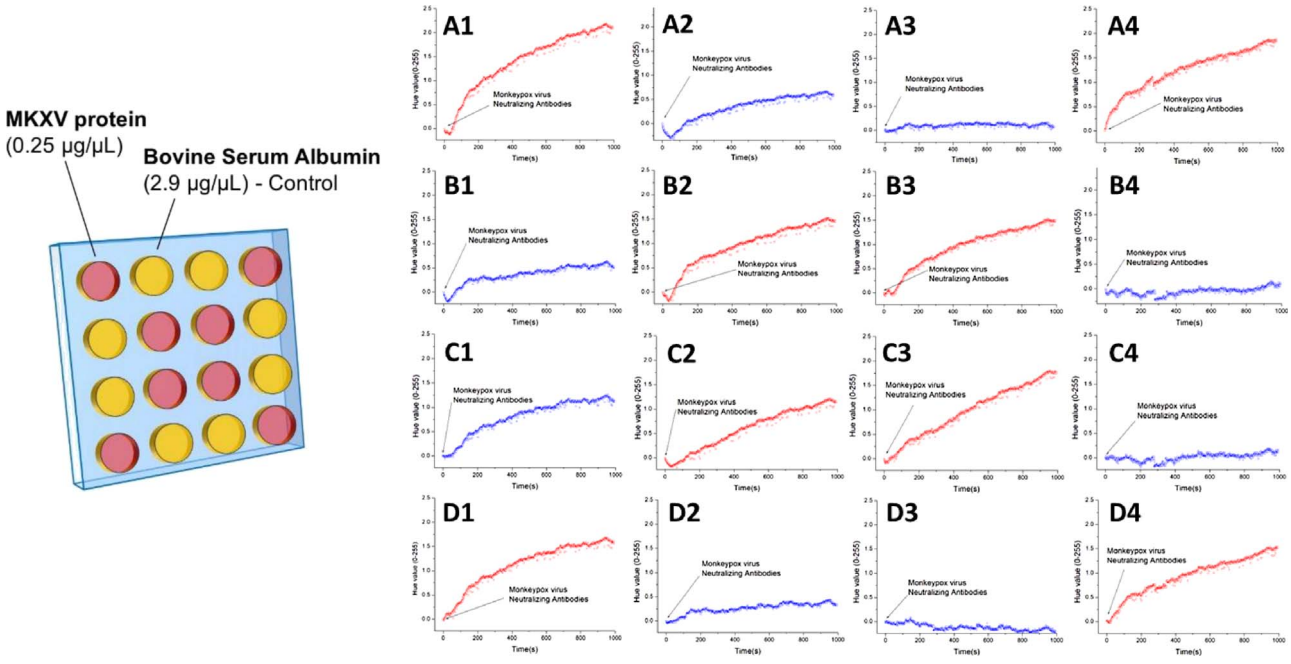


Fig. 3. Distribution map of the specific Monkeypox virus protein and control bovine serum albumin protein molecules on the biosensor array spots. A1–D4 shows 16 molecular binding curves recorded for all sensor spots in parallel.

Sino Biological and Sigma-Aldrich, respectively. In the biosensing experiment, MPXV proteins (0.25 μg/μL) and control BSA proteins (2.94 μg/μL) were immobilized onto the 4 × 4 gold sensor array using the following surface chemistry procedures. Initially, all gold sensor spots were incubated in an 11-mercatoundecanoic acid solution (10.8 mM) for 28 h to produce a self-assembled monolayer. Subsequently, mixture of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (0.4 M) and N-hydroxysuccinimide (1 M) solution were applied to activate the sensor surfaces. It was followed by a protein immobilization process of 3 h. Finally, the sensor surface was rinsed with phosphate-buffered saline (PBS) buffer before performing MPXV neutralizing antibodies detection.

3 Results and discussion

Performance of the spectral-phase 3D SPR imaging sensor was first characterized with NaCl solutions ranged from 0% to 15%, which corresponding to refractive index values between 1.3330 to 1.3585 RIU [8]. Figures 2a and 2b are the experimental spectral-phase SPR images of water (0%, 1.3330 RIU) and 15% NaCl solution (1.3585 RIU), respectively. A distinct green shift is observed with increasing refractive index values. In addition, a fiber-based spectrometer (Ocean Insight, model: Ocean HR high-resolution spectrometer) was used to analyze the visible range spectral profiles. Figure 2c presents the plasmonic excitation spectra of 0% (water), 2%, 4%, 6%, 8%, 10% and 15% NaCl solutions. The spectra exhibit a significant intensity increase within the green (500–570 nm) and yellow (570–590 nm) spectral ranges, while the red spectral region (620–750 nm) is suppressed. The spectral-phase SPR images were then quantified by extracting Hue component of the HSV color space, where the Hue value represents the dominate wavelength of color [9]. Figure 2d illustrates the Hue values as a function of refractive index. It demonstrates a linear relationship within 1.3330–1.3501 RIU. Furthermore, slope of the sensor response curve yields a sensor sensitivity of 1191.2 Hue units/RIU.

The spectral-phase imaging biosensor was further applied for the detection of MPXV neutralizing antibodies. Figure 3 shows the distribution map of MKXV protein (red sensor spots) and the control BSA protein (yellow sensor spots) in the biosensor array. The functionalized sensor surface was initially kept at PBS, followed by the injection of MPXV neutralizing antibodies (13 ng/μL) to facilitate MPXV protein–antibody bindings. Figure 3 presents the molecular binding curves recorded for all 16 sensor spots. On average, the injection of 13 ng/μL MPXV antibody produced a sensor response of 1.6 Hue units. In Figure 3, the discernible signals observed at sensor spots A2, B1, C1 and D2 are attributed to non-specific bindings, potentially resulting from physical adsorption of MPXV proteins. The non-specific sensor spots can serve as control sensor sites. The final sensor response for MPXV antibody was subtracted with the average non-specific sensor responses (0.33 Hue units). In addition, the sensor measurement stability was determined to be 0.033 Hue units based on 500 measurements in PBS buffer. According to the literature [10], the biomolecular detection sensitive can be estimated to be 0.27 ng/μL using equation (1). This value is significantly lower than the previously reported neutralizing antibodies level (3.0 ng/μL) induced by MPXV vaccine [11].

$$\text{Sensor biosensing resolution} = \frac{\text{Concentration of biomolecules}}{\text{Sensor response}} \times \text{Sensor measurement stability}. \quad (1)$$

Furthermore, the observed binding kinetics difference between A1 and C3 sensor spots may cause by spatial non-uniformity of MPXV antibody molecules within the flow cell (1.5 × 1.5 cm^2). Such variations could be minimized by using a syringe pump to maintain a continuous flow during detection. In addition, different concentrations of MPXV neutralizing antibodies ranged from 0 ng/μL (PBS buffer), 1.56 ng/μL, 3.125 ng/μL, 6.25 ng/μL to 12.5 ng/μL were sequentially injected onto the MPXV protein functionalized surface. Each antibodies sample was incubated for 45 min within the detection chamber to facilitate MPXV protein–antibody binding interactions. After subtracting the PBS baseline, the normalized sensor responses were plotted in Fig. A1 (Appendix). The results demonstrate a linear relationship between the sensor signals and MPXV neutralizing antibodies concentrations. Slope of the sensor response curve also provides the detection sensitivity of the biosensor. It is determined to be 8.45 ng/μL / Hue units.

Acknowledgments

Chi Lok Wong would like to give his sincere thanks to Yeshua for his guidance and supports in all experiments and ideas.

Funding

This research was funded by National Science and Technology Council (NSTC), Taiwan ROC grant number NSTC 112–2221-E-182-054- and NSTC 113–2221-E-182-010-.

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

Data underlying the results presented in this paper can be obtained from the author upon reasonable request.

References

1 Wang Y, Yang K, Zhou H, Immunogenic proteins and potential delivery platforms for mpox virus vaccine development: a rapid review, *Int. J. Biol. Macromol.* **245**, 125515 (2023). https://doi.org/10.1016/j.ijbiomac.2023.125515.
2 Grossegesse M, Stern D, Hofmann N, Surtees R, Kohl C, Michel J, Nitsche A, Serological methods for the detection of antibodies against Monkeypox virus applicable for laboratories with different biosafety levels, *J. Med. Virol.* **95**(12), e29261 (2023). https://doi.org/10.1002/jmv.29261.
3 Wong CL, Ho HP, Kong YK, Chen QL, Yuan W, Wu SY, Real-time protein biosensor arrays based on surface plasmon resonance differential phase imaging, *Biosens. Bioelectron.* **24**, 606 (2008). https://doi.org/10.1016/j.bios.2008.06.013.
4 Wong CL, Ho HP, Chan KS, Wu SY, Application of surface plasmon resonance sensing to studying elastohydrodynamic lubricant films, *Appl. Optics* **44**(23), 4830 (2005). https://doi.org/10.1364/AO.44.004830.
5 Wong CL, Multiplex spectral-phase surface plasmon resonance 3D imaging biosensor, *Opt. Com.* **584**, 131815 (2025) https://doi.org/10.1016/j.optcom.2025.131815.
6 Wong CL, Spectral-phase 3D surface plasmon resonance imaging biosensor for COVID-19 neutralizing antibody detection, *Opt. Express* **33**(8), 18089 (2025). https://doi.org/10.1364/OE.550815.
7 Hashemi Z, Ranjbari F, Ranc V, Fathi F, Multiplexed biosensing: a review of surface plasmon resonance platforms for biomarker analysis, *Talanta* **304**, 129564 (2026). https://doi.org/10.1016/j.talanta.2026.129564.
8 Weast RC, *CRC Handbook of Chemistry and Physics*, 69th edn. (CRC Press, New York, 1988).
9 Jahne B, *Digital Image Processing*, 4th edn. (Springer, Berlin, Heidelberg, 1997).
10 Wong CL, Olivo M, Surface plasmon resonance imaging sensors: a review, *Plasmonic* **9**, 809 (2014). https://doi.org/10.1007/s11468-013-9662-3.
11 Shi YT, Wang S, Hao Y, Shen X, Zhang J, Wang S, Zhang J, Fu Y, Chen R, Wang D, Shao Y, Li D, Liu Y, Isolation and characterization of e8 monoclonal antibodies from donors vaccinated with recombinant vaccinia vaccine with efficient neutralization of authentic Monkeypox virus, *Vaccines* **13**(5), 471 (2025). https://doi.org/10.3390/vaccines13050471.

Appendix: Detection results for different concentrations of MPXV antibody samples

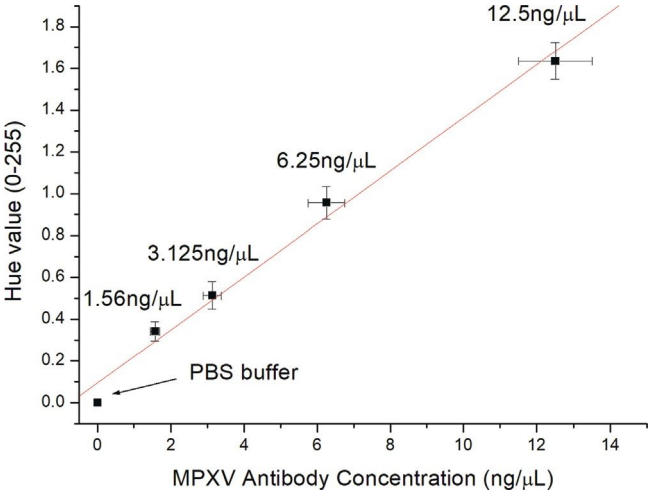


Fig. A1. Different concentrations of MPXV neutralizing antibodies ranged from 0 ng/μL (PBS buffer), 1.56 ng/μL, 3.125 ng/μL, 6.25 ng/μL to 12.5 ng/μL were sequentially injected onto the MPXV protein functionalized surface. Each antibodies sample was incubated for 45 min within the detection chamber to facilitate MPXV protein-antibody binding interactions. After subtracting the PBS baseline, the normalized sensor responses were plotted in Fig. A1 (Appendix). The results demonstrate a linear relationship between the sensor signals and MPXV neutralizing antibodies concentrations. Slope of the sensor response curve also provides the detection sensitivity of the biosensor. It is determined to be 8.45 ng/μL /Hue unit.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 53
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026046
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE OPEN ACCESS

Determining the health state of a single exosome via nano-coated silica microresonator

Maíra Martins Garcia, Brahim Bessif, Daniel Erni, and Mandana Jalali*

General and Theoretical Electrical Engineering (ATE), University of Duisburg-Essen (UDE) and Center for Nanointegration Duisburg-Essen (CENIDE), D-47048 Duisburg, Germany

Received 30 January 2026 / Accepted 10 May 2026

Abstract. We present a numerical study of a nanocoated whispering-gallery-mode (WGM) silica microdisk as a label-free platform for single-exosome detection and refractive-index–based health-state classification. The device is modeled as a fiber-coupled silica microdisk in water, functionalized with a thin nanocoating of either polystyrene (PS) or the metal-halide perovskite $CsPbI_3$. Using full-wave driven-mode simulations in Comsol Multiphysics, we show that nanocoatings reshape the WGM field distribution and improve sensing-relevant figures of merit by enhancing surface-field confinement while preserving high-Q_f operation. Beyond the field-pulling mechanism provided by polymer coatings, we demonstrate a distinct sensitivity enhancement enabled by perovskites: spectral alignment of the WGM with an excitonic resonance in $CsPbI_3$ supports a hybrid excitonic–photonic mode that concentrates optical energy at the sensing interface and increases the transduction of minute effective-refractive-index (ERI) variations into measurable resonance shifts. To connect the exosome composition to the optical response, we introduce a physics-based workflow to estimate dispersive ERIs of individual exosomes from their protein and nucleic-acid content using a Barer-type relation and a core–shell geometry, and we map these ERIs to resonance-wavelength shifts for single exosomes at the sensing position. The resulting resonance signatures provide separable responses for healthy-like, borderline, and cancer-like exosomes, indicating that the proposed excitonically engineered WGM microresonator can not only detect single exosomes but also classify their health state, supporting a route toward non-invasive liquid-biopsy diagnostics.

Keywords: Whispering gallery mode (WGM) microresonator, Nanocoating, Polystyrene, Perovskite, Exosome health state, Excitonic-photonic coupling.

1 Introduction

Optical biosensors that enable direct, real-time, non-invasive and label-free detection of bioelements are highly desired in biological studies. These sensors require ultra-sensitive mechanisms to detect single nanoscale particles and, when the goal is specificity, should also detect and distinguish even small variations in the particle's composition. In this regard, optical microresonators, sustaining whispering gallery modes (WGMs) have been established as ultra-sensitive sensors and are the appropriate candidates for such task.

The first demonstrations of using WGM microresonators for biosensing emerged in the early 2000s, such as the 2002 report by Vollmer et al. on protein detection using a resonant microcavity and the 2003 study by Arnold et al. discussing the frequency shifts induced by protein adsorption in microspheres sustaining WGMs [1, 2]. Since then, interest in using WGM-based biosensors has increased substantially, and major advances in both sensing mechanisms and experimental techniques have been made [3–6].

Therefore, WGM biosensing has developed into a central platform for label-free biological and chemical sensing, owing to their ultra-high quality factors (Q_f), small mode volumes (V_{mode}), and strong electric field intensity at the sensing location $|\boldsymbol{E}|$. A wide variety of ultra-high Q_f WGM microresonator designs have already been fabricated, with various geometries including microspheres, microdisks, microrings, microtoroids, microbubbles, microbottles, and microcapillaries, presenting ultra-high Q_f values in the range of 10^4–10^{10} [3–5, 7, 8].

Comprehensive reviews report applications of WGM microresonators ranging from bulk refractometry and protein binding to single virus and nanoparticle detection, as well as recent progress towards integrated and optofluidic implementations for clinical diagnostics [3, 6, 7, 9–11]. More specifically, for single-molecule and single-particle detection, WGM biosensors have also been applied to sense

* Corresponding author: mandana.jalali@uni-due.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

nucleic acids (NAs), antibodies, and extracellular vesicles (EVs) such as exosomes [12–14]. For EVs and exosomes in particular, microtoroid WGM resonators have enabled real-time, label-free single-exosome detection in complex media [13], complementing other optical platforms such as interferometric imaging and nanoparticle-tracking-based methods [15–17]. Collectively, these characteristics established WGM-mode microresonators as an attractive platform for single-nanoparticle detection and characterization.

Exosomes are of particular interest for diagnostics because their molecular cargo reflects the physiological and pathological state of their parent cells, making them promising biomarkers for a range of diseases, including cancer [18, 19]. This property supports early-stage disease detection through non-invasive or minimally invasive liquid biopsies, as exosomes can be isolated from readily accessible body fluids such as blood, saliva, and urine [20, 21]. Tumor-derived exosomes often carry altered protein and nucleic-acid signatures compared with exosomes from healthy cells, which can manifest as subtle shifts in their effective refractive index (ERI); such shifts are, in principle, detectable with WGM-based biosensors and may enable cancer diagnosis [22–24]. Nevertheless, direct phenotyping and robust health-state classification remain challenging due to the small size of exosomes (typically 30–150 nm) and their low refractive-index contrast relative to the surrounding medium [25–27].

Recent optical studies have begun to quantify the ERI of extracellular vesicles (EVs) and to correlate subtle ERI variations with differences in biochemical composition [17, 25, 26]. In parallel, a range of hybrid and coated WGM architectures, including plasmonic-enhanced resonators, high-index coatings, and polymer microlaser platforms, have been proposed to further localize the optical field at the sensor surface and to push detection limits toward the single-molecule regime [28–30]. Metal-halide perovskites have emerged as particularly attractive high-index, excitonic materials, enabling high-Q_f WGM lasing and strong exciton–photon coupling in $CsPbX_3$ and $FAPbX_3$ microcavities [31–35]. However, to the best of our knowledge, perovskite nanocoatings have not yet been leveraged to engineer WGM microresonators specifically for exosome sensing, nor has ERI-based classification of individual exosomes been explored using such hybrid structures.

The goal of this work is to discriminate the health state (cancerous versus healthy) of single exosomes using coated WGM microresonators. Realizing this capability would enable an early-stage, non-invasive, label-free biosensing approach to cancer diagnostics based on liquid biopsy. Achieving robust single-exosome classification requires further improvements in the sensing limit of optical microresonators. A particularly effective strategy is to introduce a carefully engineered nanocoating with a refractive index higher than that of the underlying resonator. We have previously shown that such high-index nanocoatings can enhance sensing performance by pulling the supported modes toward the coating region, increasing Q_f and $|\boldsymbol{E}|$, and simultaneously reducing the effective mode volume V_{mode} [36–38]. Results show that, depending on the coating material, $|\boldsymbol{E}|$ can be amplified by up to a factor of 2.25 relative to the uncoated microresonator, V_{mode} can be reduced by 21%–77%, and Q_f can increase from approximately 5×10^4 (uncoated) up to 8×10^7 for the case of Si_3N_4 nanocoating [36].

In this work, microdisks with tailored nanocoatings are modeled and simulated using the finite-element method (FEM)-based simulation platform Comsol Multiphysics. Full-wave, driven-mode simulations are performed, in which WGMs are excited in the microresonator via fiber optic coupling. Although computationally demanding, this modeling approach allows a comprehensive investigation of the impact of the nanocoating on the microresonator characteristics and enables precise calculation of the resonance wavelength shifts induced by the presence of healthy and cancerous exosomes. A 30 μm silica microdisk in an aqueous medium is considered, to which a 100 nm-thick nanocoating of either a polymer or a perovskite material is added. The sensing performance of the uncoated microresonator, the polymer-coated microresonator and the perovskite-coated microresonator is evaluated in terms of Q_f, mode volume, coupling efficiency and electric field strength at the sensing location.

Additionally, we introduce and implement a novel approach for determining the effective refractive index (ERI) of exosomes via numerical simulations. This method yields a dispersive ERI for healthy and cancerous exosomes, as well as for exosomes whose ERI lies in the borderline region between the two health states. These exosomes are then placed in the near field of the modeled microresonators and the resulting resonance wavelength shifts are computed.

Our results show that selecting a suitable perovskite material with a bandgap around the WGM resonance wavelengths of the microresonator enables strong photonic–excitonic interaction between the perovskite coating and the resonator, leading to a substantial enhancement of the quality factor. These improvements, in turn, enable the detection of very small variations in the exosomes' ERI and thereby the determination of the health state of individual exosomes.

The rest of the paper is organized as follows. In Section 2, the nanocoating concept is discussed in detail and the choice of nanocoating materials is explained. Section 3 introduces exosomes and our approach to determine their ERI in different health states; the modeling details and the resulting dispersive ERIs are presented there. The microresonator modeling and simulations are described in Section 4, while the impact of the nanocoating on the microresonator performance is analyzed in Section 5. In Section 6, single exosomes are positioned next to the microresonators and the corresponding resonance wavelength shifts are evaluated. Finally, Section 7 summarizes the main conclusions and provides an outlook on future work.

2 The nanocoating

The rationale for introducing a high-refractive-index nanocoating is to directly engineer the three figures of merit that determine the resonance shift of whispering-gallery–

mode (WGM) microresonators, namely the quality factor Q_f, the mode volume V_{mode}, and the local electric field intensity at the sensing location $|E(\boldsymbol{r})|^2$. Rather than modifying the resonator geometry, we add a thin nanocoating with refractive index $n_{coat} > n_{SiO_2}$ around the silica microresonator. An extensive material study among dielectrics, polymers, quantum dots, and perovskites, based on their dispersive refractive indices and, in the cases of quantum dots and perovskites, their bandgaps, showed that appropriate candidates for our purpose are silicon nitride Si_3N_4 and titanium dioxide TiO_2 (dielectrics), polystyrene (PS; polymer), cadmium selenide (CdSe) and cadmium zinc (CdZn) quantum dots, as well as cesium lead triiodide ($CsPbI_3$) perovskite. The corresponding dispersive refractive indices and extinction coefficients of these materials, together with the refractive index of silica as the baseline, obtained from Ref. [39, 40], are illustrated in Figure 1. Among the investigated materials, the PS polymer and the $CsPbI_3$ perovskite are chosen as the nanocoating materials in this work.

2.1 Polymer coating (polystyrene)

In our earlier eigenmode analysis of coated silica microspheres, we systematically explored several high-refractive-index dielectric and polymer materials as candidate nanocoatings, namely TiO_2, Si_3N_4, and PS, with various thicknesses [36]. For each coating, the quality factor Q_f, the mode volume V_{mode}, and the normalized electric field at the sensing interface $|E_{norm}| = |E|_{surf}/|E|_{surf}^{(uncoated)}$ were obtained from our eigenmode models. All three nanocoatings led to an increase in Q_f and a reduction in V_{mode} compared to the uncoated microsphere, confirming the general effectiveness of the nanocoating concept for improving the sensitivity. However, the different materials exhibited distinct trade-offs between radiation-loss suppression and evanescent-field penetration into the surrounding sensing medium, i.e., the water in which the exosomes are suspended.

The highest-index TiO_2 nanocoating provided a strong suppression of radiative leakage and thus a substantial Q_f enhancement, together with a reduced mode volume, but the WGM was largely confined inside the nanocoating, which resulted in a diminished electric field at the outer interface. The Si_3N_4 nanocoating offered the best purely optical figures of merit within the eigenmode framework: it yielded the largest increase in Q_f, the smallest V_{mode} and an enhancement of the surface field by approximately a factor of two relative to the uncoated case. PS, on the other hand, produced a somewhat smaller Q_f enhancement, yet still reduced the mode volume and increased the surface field by a factor of ~2.25, while its refractive index, being closer to that of water ($n_{water} = 1.3264 + i4.29 \times 10^{-6}$ in the corresponding wavelength range), allowed the evanescent tail of the mode to penetrate more efficiently into the external sensing region [36]. This behaviour was also reflected in the radial field profiles, where the PS nanocoating shifted the mode towards the surface without overly confining it within the coating.

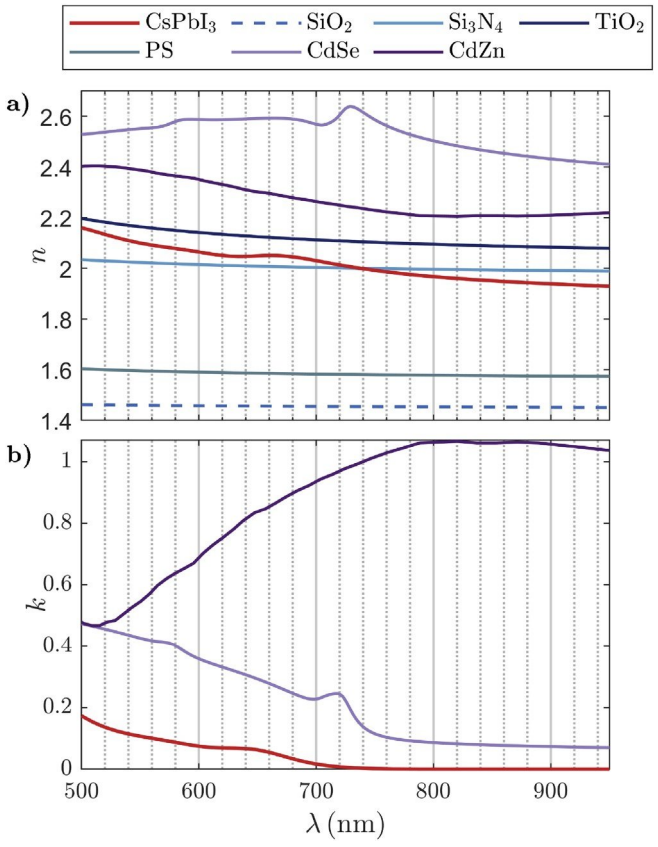


Figure 1. a) Dispersive refractive indices of SiO_2, Si_3N_4, TiO_2, PS, CdSe, CdZn, and $CsPbI_3$ versus wavelength, and b) the extinction coefficients for CdSe, CdZn, and $CsPbI_3$.

Beyond these optical metrics, practical fabrication considerations strongly favour polymer coatings for implementing nanocoatings on curved microresonators. High-index dielectric layers such as TiO_2 or Si_3N_4 typically require vacuum-based deposition techniques (e.g. CVD, ALD, sputtering) that can introduce surface roughness and residual stress, which in realistic devices reintroduce scattering losses and may compromise the high Q_f values predicted by ideal eigenmode simulations. In contrast, PS can be deposited by standard spin-coating or dip-coating from solution on silica substrates, yielding conformal nanocoatings with low surface roughness and good thickness control, and thus preserving the high intrinsic Q_f of the underlying resonator. Taking into account both optical performance and fabrication feasibility, we therefore selected PS as the polymer nanocoating material in the present work and use it as the reference system for comparison with higher-index perovskite coatings.

It is important to note that the initial material screening in Ref. [36] was performed in an eigenmode setting, which is ideal for extracting intrinsic Q_f and V_{mode} but does not explicitly include the excitation and out-coupling configuration. In an eigenmode formulation, the resonator is treated as a source-free open cavity, yielding complex eigenfrequencies $\tilde{\omega}$ and hence an intrinsic quality factor

$$Q_{\mathrm{int}} = \frac{\Re\left\{\tilde{\omega}\right\}}{2\,\Im\left\{\tilde{\omega}\right\}}, \tag{1}$$

together with the corresponding modal field profiles and V_{mode} for the isolated structure. This framework does not capture coupling-induced loading, mode selectivity under a specific excitation, nor the transmitted-spectrum line shape. By contrast, our full-wave driven-mode simulations explicitly launch the single-mode optical fiber (SMF) fundamental mode and compute the transmission T of the coupled fiber–microdisk system. The extracted linewidth therefore corresponds to the loaded quality factor Q_{L},

$$\frac{1}{Q_{\mathrm{L}}} = \frac{1}{Q_{\mathrm{int}}} + \frac{1}{Q_{\mathrm{ext}}}, \tag{2}$$

where Q_{ext} accounts for external coupling. This driven formulation enables optimizing the coupling gap d_x (under-/critical-/over-coupling) based on T, and it provides power-normalized field amplitudes at the sensing region under realistic excitation conditions, which is essential for quantitatively predicting exosome-induced resonance shifts. Accordingly, in the present study, these eigenmode-based design guidelines are validated and extended using full driven-mode simulations with explicit fiber-optic coupling. The driven-mode results confirm the qualitative trends found previously and demonstrate that PS nanocoatings indeed yield the expected enhancement of the exosome-induced resonance wavelength shifts under realistic excitation conditions.

2.2 Perovskite nanocoating and exciton–photon coupling

To further enhance the sensing performance in our approach beyond what is achievable with purely dielectric or polymer nanocoatings, we investigated whether an excitonic–photonic coupling between the microresonator whispering-gallery mode (WGM) and an excitonic transition in the nanocoating can be exploited. The basic idea is to choose a coating material that supports a pronounced excitonic resonance in the same spectral range as the fundamental WGM of the silica microresonator, such that the photonic mode hybridizes with the excitonic mode and forms mixed exciton–photon states with enhanced field confinement and a dispersive response at the sensor surface.

As a first step, we carried out a material screening of potential candidates for excitonic material systems, focusing on commercially available quantum-dot materials (CdSe, CdZn-based quantum dots) and metal-halide perovskites. Using reported dispersion data and bandgap energies (cf. Fig. 1b), we identified materials whose bandgaps lie in the vicinity of ~700 nm, i.e. close to the WGM resonance of our 30 µm silica microresonator in water. This analysis confirmed that CdSe and CdZn quantum dots as well as the perovskite $CsPbI_3$ exhibit excitonic transitions in the targeted spectral range. In a subsequent eigenmode study, each of these materials was introduced as a nanocoating with thicknesses in the range of 100–200 nm around the silica microresonator, and the resulting resonator figures of merit – the quality factor Q_{f}, the mode volume V_{mode} and the surface electric field $|\boldsymbol{E}|$ – were systematically evaluated.

The quantum-dot coatings, due to their relatively high extinction coefficients (cf. Fig. 1b), not only failed to achieve any clear photonic–excitonic coupling but also substantially degraded the photonic mode of the microresonator and, in several cases, prevented the formation of well-defined resonances. In contrast, the $CsPbI_3$ perovskite nanocoating led to a distinct photonic–excitonic coupling and accordingly provided the most pronounced simultaneous improvement of all sensing figures of merit. In the eigenmode simulations, the perovskite-coated resonator exhibited a substantial enhancement of Q_{f}, a compression of V_{mode} and a strong increase of the evanescent field at the outer interface compared to the uncoated resonator and the resonators with the best-performing polymer (PS) and dielectric (TiO_2, Si_3N_4) coatings [41]. The spectral dependence of the eigenfrequencies and field distributions indicated that the WGM of the silica core hybridizes with the excitonic resonance of the $CsPbI_3$ layer: close to the perovskite band edge, the mode is pulled into the coating, the effective index exhibits a pronounced dispersive feature and the field intensity at the coating–water interface is strongly amplified. These observations are consistent with an excitonic–photonic coupling mechanism rather than a purely geometric or index-contrast effect.

A natural concern when employing excitonic materials as microresonator coatings is their intrinsic loss: the strong oscillator strength that gives rise to the excitonic resonance is accompanied by a non-negligible imaginary part of the refractive index, which could in principle degrade the microresonator's Q_{f} and offset any sensitivity gain. For this reason, we explicitly compared all sensing criteria of the $CsPbI_3$-coated resonator against those of the PS and dielectric (TiO_2, Si_3N_4) nanocoatings. Despite its higher material loss, the $CsPbI_3$ coating was found to outperform all other options: in the spectral region where the WGM and excitonic resonance overlap, the perovskite nanocoating yields the highest Q_{f}, the smallest V_{mode} and the strongest surface field enhancement, which together translate into the largest resonance wavelength shift induced by a single exosome. These eigenmode-based findings, first reported in our earlier perovskite-coating study [41], are in the present work corroborated and extended by full driven-mode simulations with explicit fiber coupling. The results obtained within this fully numerical framework should be interpreted as a predictive assessment of the sensing performance under realistic excitation conditions, indicating that the excitonic $CsPbI_3$ nanocoating could significantly enhance the sensing functionality of silica microresonators.

The analysis presented here focuses on the idealized optical response of an intact $CsPbI_3$ nanocoating within a numerical design framework. Long-term chemical stability of bare $CsPbI_3$ in aqueous or electrolyte-containing biological environments is not assumed and remains a separate materials-engineering challenge. Practical implementation in biofluids will therefore require appropriate stabilization and encapsulation strategies for the perovskite shell, along with careful control of the local chemical environment.

3 Exosomes

Exosomes are nanoscale extracellular vesicles (EVs), typically ∼30–150 nm in diameter, released by most cell types through the endosomal pathway. They carry a rich molecular cargo of proteins, lipids, metabolites, and nucleic acids (DNA, mRNA, miRNA), which closely reflects the physiological and pathological state of their parental cells [18, 19]. Because exosomes are abundant and stable in virtually all biofluids (blood, urine, saliva, cerebrospinal fluid), they can be collected in a minimally invasive manner by liquid biopsy and are therefore highly attractive as early-stage cancer biomarkers [20, 21]. Tumour-derived exosomes exhibit characteristic alterations in their cargo composition – in particular in protein and nucleic-acid content – and encode cancer-type-specific signatures that can be exploited for diagnosis, patient stratification, and therapy monitoring [17, 23, 24].

From an optical-sensing perspective, these biochemical alterations manifest as subtle changes in the exosome's effective refractive index (ERI), which can in principle be detected by sufficiently sensitive label-free photonic biosensors [16, 42]. The present work builds on this concept by linking the protein and nucleic-acid (NA) content of single exosomes to their ERI and, in a second step, to the resonance response of high-Q_f WGM microresonators.

3.1 Core–shell geometric model of the exosome

A single exosome is modeled as a concentric core–shell sphere embedded in an aqueous host medium and numerically analyzed in Comsol Multiphysics. The core represents the lumen, an aqueous solution of proteins and nucleic acids (NA), while the shell represents the lipid bilayer membrane. Consistent with electron microscopy and membrane biophysics, the outer radius of the exosome is taken as $a = 45.5$ nm, and the lipid bilayer shell thickness as $t_{\text{shell}} = 5$ nm, so that the lumen radius is $b = a - t_{\text{shell}} = 40.5$ nm [43]. The lipid shell is modeled as a non-absorbing layer with slightly dispersive refractive index around $n_{\text{shell}} \simeq 1.48$ in the visible/near-IR, in line with spectroscopic studies of supported lipid bilayers and vesicles [44–47]. The host medium is taken as water with dispersive refractive index $n_{\text{water}}(\lambda)$ [48].

Within this framework, the only degrees of freedom that encode the biological state of the exosome are the lumen composition (protein and NA concentrations) and, to a lesser extent, the detailed lipid composition of the membrane. In this study we focus on the lumen composition as the primary contrast mechanism between healthy-like and cancer-derived exosomes.

3.2 Lumen refractive index from the Barer relation

The refractive index of the exosome lumen is computed in MATLAB using the classical two-substance (Barer) relation for aqueous macromolecular solutions [49]. For an aqueous mixture of proteins and nucleic acids at mass concentrations c_{prot} and c_{NA} (in g mL^{-1}), the lumen refractive index can be written as

$$n_{\text{core}}(\lambda) = n_{\text{water}}(\lambda) + \left(\frac{\mathrm{d}n}{\mathrm{d}c}\right)_{\text{prot}} c_{\text{prot}} + \left(\frac{\mathrm{d}n}{\mathrm{d}c}\right)_{\text{NA}} c_{\text{NA}}, \quad (3)$$

where $(\mathrm{d}n/\mathrm{d}c)_{\text{prot}}$ and $(\mathrm{d}n/\mathrm{d}c)_{\text{NA}}$ are the refractive-index increments (RII) of proteins and nucleic acids, respectively. In the visible range at ambient temperature, robust consensus values are $(\mathrm{d}n/\mathrm{d}c)_{\text{prot}} \approx 0.190$ mL g^{-1} and $(\mathrm{d}n/\mathrm{d}c)_{\text{NA}} \approx 0.170$ mL g^{-1}, with only weak dispersion [50, 51].

Guided by experimental reports that place the effective refractive index of extracellular vesicles in the range $n_{\text{eff}} \approx 1.37$–$1.40$ in the visible [16, 26, 42], we select biologically plausible concentration bands for healthy-like and cancer-like exosomes. Healthy-like exosomes are assigned protein and NA concentrations in the lower part of the band, while cancer-derived exosomes are assigned elevated concentrations, consistent with the increased protein and NA cargo typically observed in tumour-derived EVs [21, 52]. Evaluating equation (3) for these concentration bands yields three representative lumen refractive indices: $n_{\text{core}}^{(\text{H})}(\lambda)$ for healthy exosomes, $n_{\text{core}}^{(\text{C})}(\lambda)$ for cancerous exosomes, and intermediate values $n_{\text{core}}^{(\text{B})}(\lambda)$ for "borderline" exosomes whose protein/NA content lies between the two regimes.

Figure 2 summarizes the resulting $n_{\text{core}}(\lambda)$ for these three classes in the wavelength range relevant to our microresonator WGMs.

3.3 Effective refractive index of the whole exosome

To obtain an effective refractive index for the entire exosome (core plus lipid shell), a full-wave driven-mode simulation of the core–shell sphere in Comsol Multiphysics is carried out. The reason for considering exosomes as core–shell structures is to increase the sensitivity to the slightest changes in the core, i.e. variations in the protein/NA content. Consistent with this modeling objective, variations in the lipid bilayer are not explicitly included, and the membrane is treated as having fixed thickness and refractive index. While biologically relevant variations in membrane composition may also contribute to the overall effective refractive index of the exosome, this effect is beyond the scope of the present study and will be addressed in future sensitivity analyses. Within this framework, the model can be regarded as conservative, as it isolates the lumen contribution and enables a clear characterization of lumen-composition-induced contrast.

The core–shell exosome model described above is illuminated by a plane wave in the 500–900 nm spectral range, and the electric field $\boldsymbol{E}(\boldsymbol{r})$ and the electric flux density $\boldsymbol{D}(\boldsymbol{r})$ inside the whole exosome volume V_p (core plus shell) are computed. The effective relative permittivity of the exosome is then retrieved via

$$\varepsilon_{\text{eff}} = \frac{\langle \boldsymbol{D}(\boldsymbol{r}) \rangle}{\varepsilon_0 \langle \boldsymbol{E}(\boldsymbol{r}) \rangle}, \quad (4)$$

where ε_0 is the vacuum permittivity and $\langle \cdot \rangle$ denotes a volume average over V_p. The effective refractive index of the exosome is then obtained as

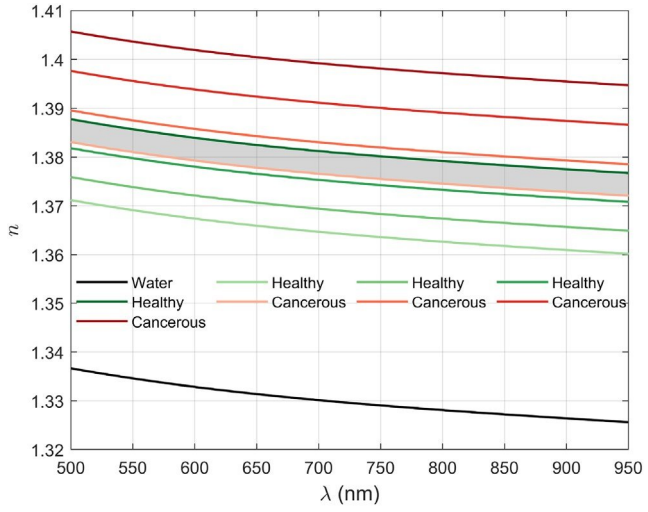


Fig. 2. Computed lumen refractive index n_{core} (λ) for healthy-like (H), borderline (B), and cancer-like (C) exosomes based on the Barer relation and protein/NA concentration bands.

$$n_{\text{eff}}(\lambda) = \sqrt{\varepsilon_{\text{eff}}(\lambda)}. \tag{5}$$

This procedure is applied to the healthy-like, borderline, and cancer-like lumen compositions defined above, thereby yielding three effective refractive indices $n_{\text{eff}}^{(\text{H})}(\lambda)$, $n_{\text{eff}}^{(\text{B})}(\lambda)$, and $n_{\text{eff}}^{(\text{C})}(\lambda)$.

The retrieved indices are consistent with the expected extracellular vesicle (EV) optical responses and fall within the experimentally observed EV index range (∼1.37–1.40 in the visible) reported by nanoparticle tracking analysis and flow cytometry [16, 26, 42, 53].

Figure 3 shows the resulting effective refractive indices for the three exosome classes at the wavelengths of interest. Cancer-derived exosomes systematically exhibit a slightly higher n_{eff} than healthy-like exosomes, with borderline exosomes (gray shaded area in the Fig. 2) populating the intermediate range. These distinct ERI bands form the basis for the single-exosome classification strategy pursued in this work, where the exosomes, modeled as homogeneous spheres with the retrieved n_{eff}, are placed in the near field of uncoated, PS- and $CsPbI_3$-coated silica microresonators to calculate the corresponding resonant wavelength shifts.

4 Modeling and simulation

All models are built in Comsol Multiphysics, which relies on the Finite-Element Method (FEM). The proposed biosensor is a silica microresonator (microdisk, 30 μm radius) with a 100 nm coating, placed in an aqueous medium. Light is coupled into the microdisk via a single-mode optical fiber (SMF), consisting of a silica core of 500 nm radius and a Teflon cladding of 2 μm thickness, designed and optimized in our previous works [37, 38]. A two-dimensional modeling approach was used, and the input power applied in the SMF optical fiber was 1 W/m for all models. A schematic representation of the modeled microresonator is shown in Figure 4.

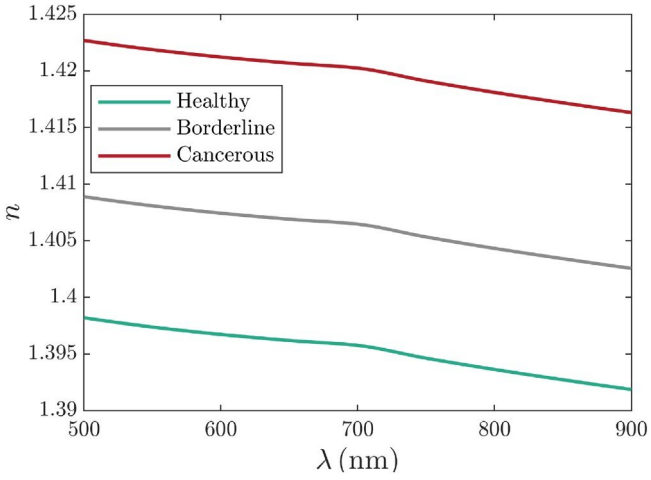


Fig. 3. Effective refractive indices n_{eff} of healthy-like (H), borderline (B), and cancer-like (C) exosomes retrieved from the full-wave Comsol Multiphysics simulations.

In order to ensure the accuracy of the results and explicitly account for the effect of light coupling into the microresonator, all simulations are performed in a full-wave driven-mode configuration. To the best of our knowledge, we are the first group to demonstrate such full-field driven-mode simulations of coated WGM microresonators [37, 38]. Three modeling scenarios are investigated: an uncoated microdisk, which provides the baseline for all comparisons; a PS-coated microdisk; and a $CsPbI_3$-coated microdisk. In each case, the coupling distance d_x, the distance from the fiber core to the microresonator, is optimized to maximize the coupling efficiency. Specifically, we choose a d_x that maximizes the transmitted power T while maintaining a high quality factor Q_{f}.

For every model, the relevant figures of merit are extracted, namely the resonance wavelengths λ_{res}, the transmitted power T, the quality factor Q_{f}, the mode volume V_{mode}, and the electric-field at the sensing location $|\boldsymbol{E}|$. These quantities are used to quantify and compare the sensing functionality of the biosensor in the uncoated, PS-coated, and $CsPbI_3$-coated cases. Afterwards, a single exosome is placed in the immediate vicinity of the microresonator equator for the three health states – healthy (H), borderline (B), and cancerous (C) – and the corresponding resonance wavelength shifts $\Delta\lambda_{\text{res}}$ are calculated. By immediate vicinity, we mean that, in each case, the exosome is positioned in direct, tangential contact with the microresonator boundary at the equator, on the right-hand side of the resonator in Figure 4. This enables a direct assessment of the sensing performance of uncoated and nanocoated microresonators for single-exosome classification.

It is worth mentioning that such multiscale full-wave driven-mode simulations (microresonator in the micrometer regime, coating and exosome in the nanometer regime) require an extremely well-designed and fine mesh, in particular in the fiber–microresonator coupling region, and around the coating and exosome. To ensure accuracy and consistency, the same meshing strategy is used for all models. This requirement results in 1,766,528 mesh elements

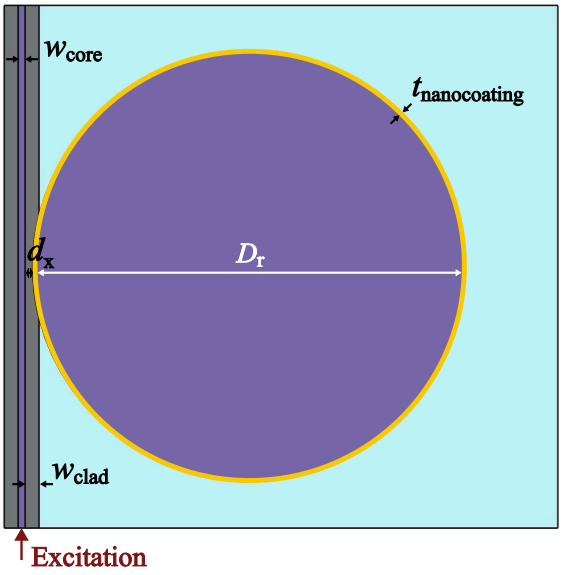


Fig. 4. Schematic of the biosensor: a 30 μm microdisk with a 100 nm coating in aqueous medium. Light is coupled into the microdisk via an optical fiber, where the coupling distance d_x is the key coupling parameter.

with an average mesh quality of 0.89. Each simulation run takes about 328 min (approximately 5 h 28 min) on a workstation equipped with an AMD Ryzen 9 7950X 16-core processor, and 96 GB of RAM.

5 Effects of the nanocoating

As a first step, the uncoated microresonator is modeled and simulated. It is found that the optimal coupling distance between the core of the optical fiber and the microresonator is $d_{x,u} = 350$ nm, for which 85.6% of the incoming power is coupled into the microresonator. The corresponding resonance wavelength is $\lambda_{res,u} = 786.92$ nm with a quality factor of $Q_f = 1.81 \times 10^4$ [38]. The electric-field intensity of this configuration at $\lambda_{res,u}$ is shown in Figure 5, together with a zoomed view of the z-component of the electric field that clearly illustrates the excited mode.

5.1 PS coating

Next, the PS-coated microresonator is modeled and simulated. In this case, the best coupling is obtained for the distance $d_{x,PS} = 400$ nm, for which 92.4% of the incoming power is coupled into the microresonator. The resonance wavelength is $\lambda_{res,PS} = 789.38$ nm, and the electric-field intensity at resonance is shown in Figure 6. It can be observed that the presence of the coating attracts the optical mode towards the coating layer and thereby amplifies the evanescent tail of the electric field in the water region.

The sensing figures of merit for this configuration are summarized in Table 1. The results show that the PS nanocoating enhances the biosensor performance compared to the uncoated case by increasing Q_f by about 15%, reducing the mode volume by 25.8%, and amplifying the electric-field intensity at the sensing location by a factor of 1.75, as already reported in [38]. These results demonstrate that the addition of the 100 nm PS coating significantly improves the sensing functionality of the microresonator.

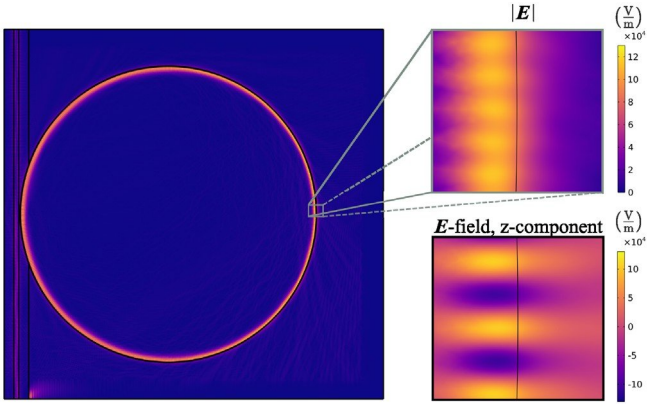


Fig. 5. Electric-field intensity of the uncoated microresonator coupled to the optical fiber at the resonance wavelength 786.92 nm, and zoomed mode profiles, including the z-component of the electric field around the microresonator surface.

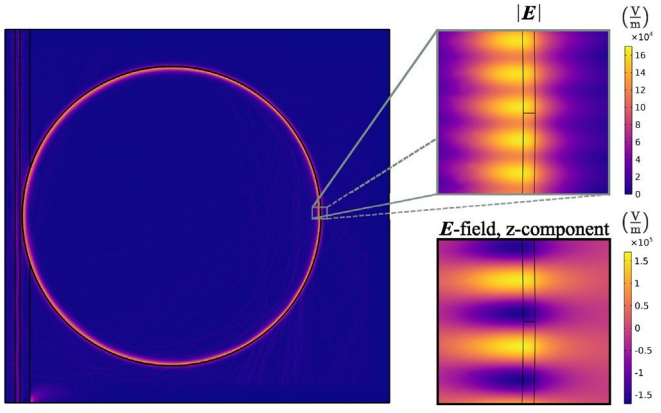


Figure 6. The electric-field at the resonance wavelength $\lambda_{res,PS} = 789.38$ nm for the PS-coated microresonator coupled to the optical fiber, with zoomed mode profile of the E_z field near the microresonator surface.

5.2 $CsPbI_3$ coating

Finally, the $CsPbI_3$-coated microresonator is modeled and numerically analyzed. Here, the first and most important step is to identify the desired excitonic–photonic coupling between the microresonator WGM and the perovskite nanocoating. Based on the eigenmode simulations in Ref. [41], a wavelength range from 886 nm to 900 nm with a step size of 0.01 nm is explored. It is observed that at $\lambda_{res,CsPbI_3} = 886.60$ nm an effective excitonic–photonic coupling between the microresonator photonic mode and the $CsPbI_3$ coating excitonic mode is achieved. In this configuration, the hybrid excitonic–photonic mode is mainly confined within the nanocoating rather than in the silica core. However, since the refractive index of the $CsPbI_3$

Table 1. Figures of merit for the sensing performance of the uncoated, 100 nm PS-coated, and $CsPbI_3$-coated microresonators.

	d_x (nm)	λ_{res} (nm)	Mode no.	T (%)	Q_f	V_{mode} (μm^2)	$\lvert \boldsymbol{E} \rvert$ (V/m)
Uncoated	350	786.92	335	85.56	1.81×10^4	73.88	0.87×10^5
PS-coated	400	789.38	339	92.39	2.09×10^4	54.79	1.52×10^5
$CsPbI_3$-coated	350	886.60	321	51.89	6.21×10^4	40.29	3.20×10^5

coating is not much higher than that of water, the evanescent tail of the electric field still penetrates strongly into the surrounding medium, and the field intensity at the sensing location is significantly increased (cf. Fig. 7).

It is worth mentioning that the electric field distribution $|\boldsymbol{E}|$ for the $CsPbI_3$-coated microresonator looks different compared to the usual WGM electric field distributions, observed in both uncoated and PS-coated microresonators. The reason behind this is the presence of the hybrid excitonic-photonic mode in the $CsPbI_3$-coated microresonator, which, due to the excitonic mode, does not depict the characteristic teardrop shapes typically present in WGM electric field distributions. Such teardrop features are visible in the E_z component resulting from the WGM of the microresonator; however, in this case, the E_x and E_y electric field components associated with the excitonic mode of the $CsPbI_3$ coating modify the electric field distribution $|\boldsymbol{E}|$.

In the next step, the coupling distance for the $CsPbI_3$-coated microresonator is optimized, and an optimal value of $d_{x,CsPbI_3} = 350$ nm is found, yielding a transmitted power of 51.9% within the microresonator. Although this transmitted power is substantially lower than in the uncoated and PS-coated cases, the other sensing-related figures of merit are markedly improved. As a result, the $CsPbI_3$-coated microresonator, due to the hybrid excitonic–photonic coupling, still outperforms the other configurations in terms of sensing functionality. All key sensing figures of merit for the three investigated configurations (uncoated, PS-coated, and $CsPbI_3$-coated microresonators) are calculated and presented in Table 1. As can be seen, the Q_f is about 2.4 times larger in the case of the $CsPbI_3$-coated microresonator compared to the uncoated case, which is a direct consequence of the hybrid excitonic–photonic mode formation. In addition, the relatively high refractive index of the $CsPbI_3$ coating reduces the mode volume by about 45% compared to the uncoated microresonator.

A critical figure of merit for quantifying the performance of our optical biosensor is the electric-field strength at the sensing location $|\boldsymbol{E}|$, as this parameter determines the excess polarizability of the exosome and, consequently, the magnitude of the resonance wavelength shift. Figure 8 shows the electric-field intensity distribution at the respective resonance wavelengths for all three configurations (uncoated, PS-coated, and $CsPbI_3$-coated microresonators) along the radial direction, zoomed into a region close to the microresonator surface. It is evident that the addition of the nanocoating shifts the optical mode profile towards the surface, thereby enhancing the field intensity at the sensing location. Moreover, the results show that the perovskite nanocoating both confines the mode and amplifies the electric-field intensity at the sensing location by a factor of 2.11 compared to the PS-coated case and by a factor of 3.68 compared to the uncoated case. This pronounced improvement is directly related to the formation of the hybrid excitonic–photonic mode within our proposed biosensor.

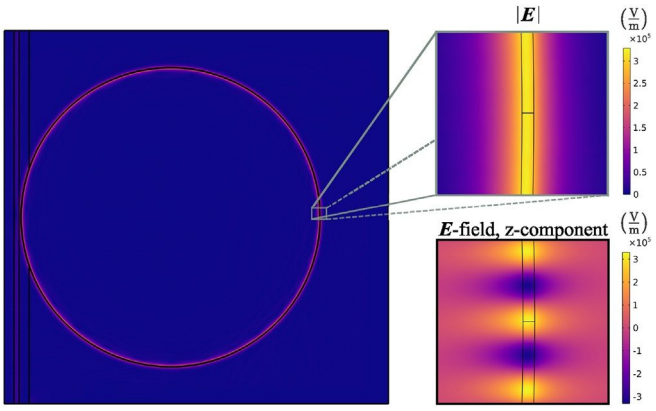


Fig. 7. Electric-field intensity of the $CsPbI_3$-coated microresonator coupled to the optical fiber at the resonance wavelength 886.60 nm. Zoomed views show $|\boldsymbol{E}|$ and E_z near the resonator surface.

The results summarized in Table 1 further underline the effectiveness of the $CsPbI_3$ coating compared to both the uncoated and the PS-coated microresonator.

Since the $CsPbI_3$ coating is lossy, there is a natural concern that absorption in the coating layer could degrade the microresonator performance. To address this, we evaluate the absorption within the nanocoating by integrating the total power dissipation density over the coated region. Because the model is two-dimensional, the integration is carried out over a cross-section, and the resulting absorbed power in the coating region is reported per unit length (W/m). It is found that the absorption losses in the coating are generally low, on the order of 10^{-5} W/m. However, as shown in Figure 9, the local absorption losses within the water medium are significantly larger in the presence of the coating, which is a direct consequence of the larger evanescent field tails (i.e. penetration depth) into the water solution.

6 Exosome detection and characterization

The final and decisive figure of merit for evaluating the $CsPbI_3$-coated microresonator is the resonance-wavelength shift induced by a single exosome. To this end, healthy

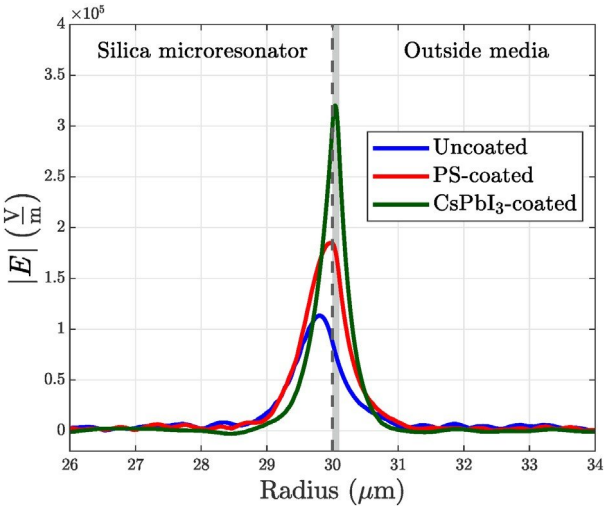


Fig. 8. Electric-field intensity along the resonator radius at the resonance wavelengths for the uncoated, PS-coated, and $CsPbI_3$-coated microresonators. The gray area represents the coating region for both PS- and $CsPbI_3$-coated cases.

(H), borderline (B), and cancerous (C) exosomes are placed in the immediate vicinity of the uncoated, PS-coated, and $CsPbI_3$-coated microresonators, and the corresponding resonance-wavelength shifts are computed for all three sensor configurations. The resulting shifts are summarized in Table 2. It can already be seen that the $CsPbI_3$ coating increases the exosome-induced wavelength shift by more than a factor of two compared to the uncoated microresonator.

Figure 10 shows the electric-field intensity distribution for the three microresonators when a single healthy exosome is placed at the sensing position. The field enhancement in the vicinity of the exosome is visibly stronger for the coated cases, with the $CsPbI_3$-coated microresonator exhibiting the largest local field intensity. Note that for the $CsPbI_3$-coated case the color scale (vertical) is adjusted, as the peak field is almost twice as large as in the uncoated and PS-coated configurations. The polarization magnitude $|\boldsymbol{P}|$ inside the healthy exosome is also shown in Figure 10 using the horizontal colorbars. The $CsPbI_3$-coated microresonator yields a significantly larger polarization response inside the exosome, presenting values approximately 3 and 6 times higher than for the PS-coated and uncoated cases, respectively (see the different color scales ranges).

When the exosome health state is changed (healthy, borderline, and cancerous), the induced polarization within the exosome varies. This behavior was observed for all three microresonator configurations. Figure 11 shows the polarization response inside healthy, borderline and cancerous exosomes placed at the sensing location for the $CsPbI_3$-coated microresonator. A monotonic increase in the polarization magnitude is observed, starting from the healthy case to the borderline case, reaching the largest values for the cancerous case. This trend is consistent with larger resonance wavelength shifts calculated for the borderline and cancerous cases, as shown in Table 2 for all the three microresonators.

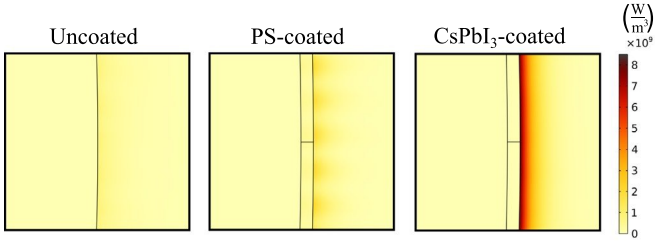


Fig. 9. Power dissipation density profiles for the uncoated, 100 nm PS-coated, and $CsPbI_3$-coated microresonators on the equator of the microresonator surface.

Table 2. Resulting resonance-wavelength shifts for a single exosome placed at the immediate vicinity of each microresonator, for healthy (H), borderline (B), and cancerous (C) exosomes.

	$\Delta\lambda_H$ (pm)	$\Delta\lambda_B$ (pm)	$\Delta\lambda_C$ (pm)
Uncoated	0.75	0.87	1.08
PS-coated	1.08	1.33	1.58
$CsPbI_3$-coated	1.79	2.02	2.38

Results from Table 2 show that compared to the uncoated microresonator, the PS-coated sensor exhibits increases of 44.0%, 52.9%, and 46.3% in the resonance-wavelength shift for the healthy, borderline, and cancerous exosomes, respectively. Changing the coating material to $CsPbI_3$ leads to even larger enhancements: the corresponding shifts increase by 138.7%, 132.2%, and 120.4% for the healthy, borderline, and cancerous exosomes, respectively, relative to the uncoated case. Thus, both coatings improve the absolute sensitivity, with the perovskite nanocoating providing the largest enhancement.

Our main goal, however, is not only to detect the presence of a single exosome but also to discriminate its health state. In this context, the separation between the wavelength shifts associated with different health classes is crucial. For the uncoated microresonator, the difference between the cancerous and healthy cases is $\Delta\lambda_C - \Delta\lambda_H = 0.33$ pm. This separation increases to 0.50 pm for the PS-coated microresonator and to 0.59 pm for the $CsPbI_3$-coated microresonator, i.e. an additional increase of about 18% in the H–C contrast compared to the PS case. Similar trends are observed for the H–B and B–C separations.

These results demonstrate that the presence of the nanocoating not only increases the absolute resonance-wavelength shifts but also enhances the contrast between healthy, borderline, and cancerous exosomes. In particular, the $CsPbI_3$-coated microresonator offers both the largest overall shifts and the largest separation between health states. This validates that our proposed excitonic–photonic hybrid microresonator is capable of identifying the health state of a single exosome and therefore supports liquid-biopsy based, non-invasive, early-stage cancer diagnostics.

In the present work, a single exosome is positioned at the equatorial region of the resonator to maximize

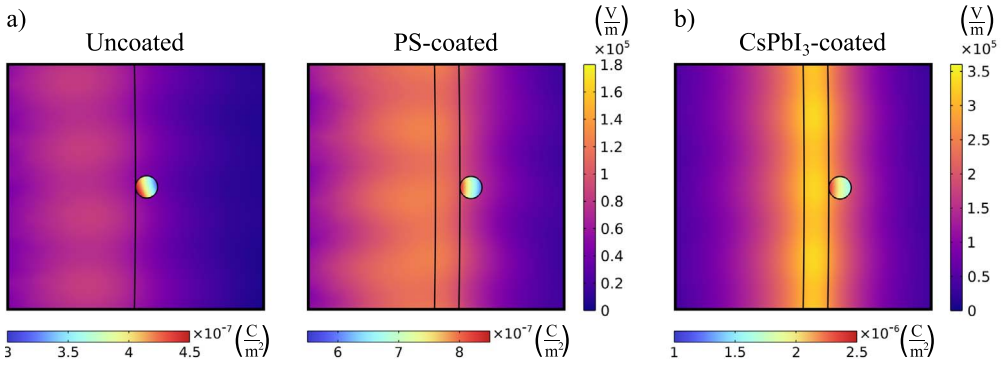


Fig. 10. Electric-field intensity distribution when a single healthy exosome (H) is placed at the immediate vicinity of (a) the uncoated and the PS-coated microresonators and (b) the $CsPbI_3$-coated microresonator, plotted together with the polarization magnitude $|\boldsymbol{P}|$ inside the exosome. The vertical colorbars denote $|\boldsymbol{E}|$. Note that the color scale for the $CsPbI_3$-coated microresonator is different and the maximum value is twice as high as in the other cases. The horizontal colorbars indicate $|\boldsymbol{P}|$ inside the healthy exosome. Note the different magnitude ranges for each plot.

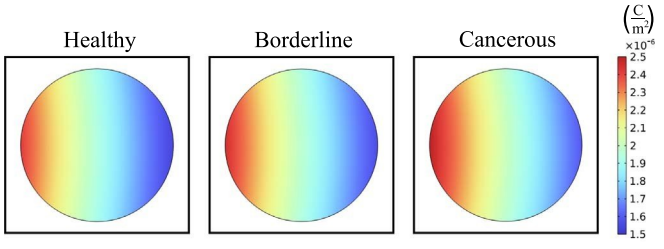


Fig. 11. Polarization in the healthy, borderline and cancerous exosomes next to the $CsPbI_3$-coated microresonator.

sensitivity, assuming ideal binding and precise placement at the sensing interface. To assess the impact of non-ideal placement, a complementary study on the same biosensor platform was conducted, in which the exosome position was systematically varied in both radial and angular directions (see [54]), assuming a 50 nm exosome with $n_{\text{eff}} = 1.3847$ [36]. The results show that, while radial displacement leads to a pronounced reduction in the sensing response for all configurations due to the rapid evanescent-field decay, the $CsPbI_3$-coated microresonator exhibits a markedly reduced dependence on small angular position variations compared to the uncoated and PS-coated cases, with an almost unchanged local electric field (variation $\approx 0.3\%$), as well as nearly invariant resonance wavelength shift and induced exosome polarization [54], as can be seen in Figure 12. This behavior is attributed to the more uniform near-surface field distribution associated with the hybrid photonic–excitonic mode supported by the perovskite coating.

From a practical perspective, the reduced angular dependence of the proposed $CsPbI_3$-coated microresonator, specifically designed for single-exosome detection, relaxes the requirement for precise exosome positioning at a single optimal location and supports more robust sensing under realistic conditions. At the same time, the results confirm that maintaining the exosome in close proximity to the resonator surface remains essential for maximizing sensitivity, highlighting the importance of surface-functionalization strategies for effective exosome trapping [54]. Scenarios involving multiple simultaneously adsorbed exosomes are beyond the scope of this study, as the sensing concept is based on controlled delivery and positioning of individual exosomes, for example through microfluidic-assisted transport and ligand-based surface functionalization.

7 Conclusion

In this work, we have proposed and numerically demonstrated an excitonic–photonic hybrid whispering-gallery-mode microresonator as an ultra-sensitive optical biosensor for single-exosome detection and health-state classification. A 30 μm silica microdisk with a 100 nm nanocoating was modeled in an aqueous environment and excited via a single-mode fiber using full-wave driven-mode computational electromagnetics simulations in Comsol Multiphysics. To the best of our knowledge, this represents the first full-field driven-mode study of coated WGM microresonators with explicit fiber coupling for biosensing applications. In parallel, we developed a physics-based framework to determine the effective refractive index of individual exosomes as a function of their protein and nucleic-acid content, combining the Barer relation for the lumen, a core–shell exosome geometry, and volume-averaged D/E retrieval from full-wave simulations. This yields dispersive effective refractive indices $n_{\text{eff}}^{(\text{H,B,C})}(\lambda)$ for healthy, borderline, and cancerous exosomes and provides, to the best of our knowledge, the first ERI-based description of exosomes explicitly parameterized by health state.

On the photonic side, we systematically compared uncoated, PS-coated, and $CsPbI_3$-coated microresonators. The PS nanocoating was shown to enhance the sensing performance by increasing the quality factor Q_{f} by $\sim$15%, reducing the mode volume V_{mode} by 25.8%, and amplifying the electric field at the sensing location by a factor of 1.75 relative to the uncoated case. Building on our previous eigenmode screening [36, 41], we then identified a configuration in which a $CsPbI_3$ perovskite nanocoating exhibits strong excitonic–photonic coupling with the silica WGM

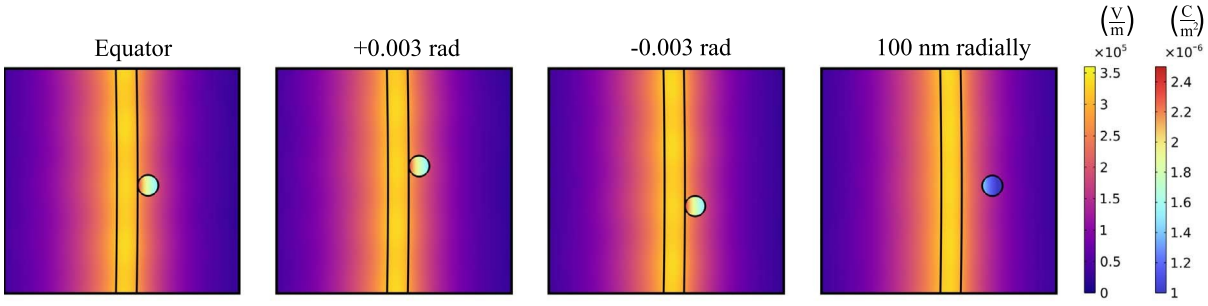


Fig. 12. Electric-field intensity distribution and the polarization magnitude $|\boldsymbol{P}|$ inside a single exosome for the $CsPbI_3$-coated microresonator, when the exosome position is varied in both angular ($\pm$ 0.003 rad) and radial (100 nm shifted) directions.

at $\lambda_{res,CsPbI_3} \approx 886.6$ nm. The resulting hybrid mode is largely confined within the perovskite nanocoating while maintaining a strong evanescent tail in water. Despite the intrinsic material loss of the perovskite, this configuration yields the highest Q_f (a factor $\sim$2.4 above the uncoated case), the smallest V_{mode} (reduced by $\sim$45%), and the strongest surface field enhancement (electric-field intensity increased by factors of 2.11 and 3.68 relative to the PS-coated and uncoated resonators, respectively), while also establishing a more spatially uniform near-surface field distribution at the sensing interface. To the best of our knowledge, this is the first report of an engineered excitonic–photonic mode coupling in a perovskite-coated WGM microresonator specifically tailored for biosensing.

Coupling the exosome ERI model to the hybrid microresonators, we quantified the resonance-wavelength shifts induced by single healthy, borderline, and cancerous exosomes placed at the sensing position. Both PS and $CsPbI_3$ nanocoatings substantially increase the absolute resonance shifts compared to the uncoated resonator, with the perovskite-coated device providing the largest enhancement. For example, the healthy-exosome shift increases from 0.75 pm (uncoated) to 1.08 pm (PS-coated) and 1.79 pm ($CsPbI_3$-coated), corresponding to enhancements of 44% and 139%. At the same time, the separation between health states is improved: the difference between cancerous and healthy exosomes, $\Delta\lambda_C - \Delta\lambda_H$, increases from 0.33 pm (uncoated) to 0.50 pm (PS-coated) and 0.59 pm ($CsPbI_3$-coated). These results show that the proposed architecture does not merely detect the presence of an exosome but also amplifies the contrast between healthy, borderline, and cancerous ERI values. Taken together, this constitutes, to the best of our knowledge, the first demonstration of an optical WGM biosensor that exploits excitonic–photonic hybridization to both detect and characterize the health state of single exosomes. Importantly, the hybrid-mode-induced field redistribution also reduces the dependence of the sensing response on angular exosome positioning, thereby enhancing robustness against non-ideal binding conditions [54].

The presented study establishes a complete, simulation-based design workflow that links biochemical composition (protein/NA content) to exosome ERI and ultimately to measurable resonance signatures in a realistic, fiber-coupled WGM biosensor. Future work will focus on translating these designs into fabricated devices, including the controlled deposition and stabilization of $CsPbI_3$ nanocoatings on silica microdisks, and on the experimental validation of the predicted single-exosome shifts in microfluidic environments. In this context, controlled delivery and isolation of individual exosomes, e.g., via microfluidic-assisted transport combined with ligand-based surface functionalization, will be essential to ensure precise positioning at the sensing interface and to mitigate multi-particle effects, which are beyond the scope of the present study. Beyond cancer diagnostics, the proposed framework can be extended to other EV subtypes and disease markers, and combined with statistical or machine-learning classifiers to enable robust ERI-based phenotyping of heterogeneous exosome populations. More broadly, our results highlight the potential of excitonic materials as functional nanocoatings for next-generation photonic biosensors operating at the single-particle level.

Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 4102100053004. We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Conflicts of interest

The authors declare no conflicts of interest associated to this work.

Data availability statement

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Author contribution statement

Modeling: M.M.G., B.B., M.J.; Simulation: M.M.G., M.J.; Investigation: M.M.G., M.J.; Conceptualization: M.J.; Funding acquisition: M.J., D.E.; Supervision: D.E., M.J.; Writing and editing: M.M.G., M.J., D.E.; Review: D.E., M.J.

References

1 Vollmer F, Braun D, Libchaber A, Khoshsima M, Teraoka I, Arnold S, Protein detection by optical shift of a resonant microcavity, *Appl. Phys. Lett.* **80**, 4057–4059 (2002). https://doi.org/10.1063/1.1482797.
2 Arnold S, Khoshsima M, Teraoka I, Holler S, Vollmer F, Shift of whispering-gallery modes in microspheres by protein

adsorption, *Opt. Lett.* **28**, 272–274 (2003). https://doi.org/10.1364/OL.28.000272.
3 Jiang X, Qavi AJ, Huang SH, Yang L, Whispering-gallery sensors, *Matter.* **3**, 371–392 (2020). https://doi.org/10.1016/j.matt.2020.07.008.
4 Loyez M, Adolphson M, Liao J, Yang L, From whispering gallery mode resonators to biochemical sensors, *ACS Sens.* **8**, 2440–2470 (2023). https://doi.org/10.1021/acssensors.2c02876.
5 Lopez-Yglesias X, Gamba JM, Flagan RC, The physics of extreme sensitivity in whispering gallery mode optical biosensors, *J. Appl. Phys.* **111**, 084701 (2012). https://doi.org/10.1063/1.3698319.
6 Foreman MR, Swaim JD, Vollmer F, Whispering gallery mode sensors, *Adv. Opt. Photonics.* **7**, 168–240 (2015). https://doi.org/10.1364/AOP.7.000168.
7 Su J, Label-free biological and chemical sensing using whispering gallery mode optical resonators: past, present, and future, *Sensors.* **17**, 540 (2017). https://doi.org/10.3390/s17030540.
8 Vitullo DLP, Zaki S, Jones DE, Sumetsky M, Brodsky M, Coupling between waveguides and microresonators: the local approach, *Opt. Express.* **28**, 25908–25914 (2020). https://doi.org/10.1364/OE.399978.
9 Wu X, Zhu C, Wang Y, Fan X, Micro/nano lasers for biomolecular sensing and cellular analysis, *Photonic Sens.* **15**, 250123 (2025). https://doi.org/10.1007/s13320-024-0711-7.
10 Toropov N, Cabello G, Serrano MP, Gutha RR, Rafti M, Vollmer F, Review of biosensing with whispering-gallery mode lasers, *Light Sci. Appl.* **10**, 42 (2021). https://doi.org/10.1038/s41377-021-00471-3.
11 Anwar AR, Mur M, Humar M, Microcavity- and microlaser-based optical barcoding: A review of encoding techniques and applications, *ACS Photon.* **10**, 1202–1224 (2023). https://doi.org/10.1021/acsphotonics.2c01611.
12 Baaske MD, Foreman MR, Vollmer F, Single-molecule nucleic acid interactions monitored on a label-free microcavity biosensor platform, *Nature Nanotech.* **9**, 933–939 (2014). https://doi.org/10.1038/nnano.2014.180.
13 Su J, Label-free single exosome detection using frequency-locked microtoroid optical resonators, *ACS Photon.* **2**, 1241–1245 (2015). https://doi.org/10.1021/acsphotonics.5b00142.
14 Giannetti A, Berneschi S, Baldini F, Cosi F, Conti GN, Soria S, Performance of eudragit coated whispering gallery mode resonator-based immunosensors, *Sensors* **12**, 14604–14611 (2012). https://doi.org/10.3390/s121114604.
15 Daaboul GG et al., Digital detection of exosomes by interferometric imaging, *Sci. Rep.* **6**, 37246 (2016). https://doi.org/10.1038/srep37246.
16 Dragovic RA et al., Sizing and phenotyping of cellular vesicles using nanoparticle tracking analysis, *Nanomedicine.* **7**, 780–788 (2011). https://doi.org/10.1016/j.nano.2011.04.003.
17 Imanbekova M, Suarasan S, Lu Y, Jurchuk S, Wachsmann-Hogiu S, Recent advances in optical label-free characterization of extracellular vesicles, *Nanophotonics.* **11**, 2827–2863 (2022). https://doi.org/10.1515/nanoph-2022-0057.
18 Kalluri R, LeBleu VS, The biology, function, and biomedical applications of exosomes, *Science* **367**, eaau6977 (2020). https://doi.org/10.1126/science.aau6977.
19 Chen Y, Luh F, Ho Y, Yen Y, Exosomes: a review of biologic function, diagnostic and targeted therapy applications, and clinical trials, *J. Biomed. Sci.* **31**, 67 (2024). https://doi.org/10.1186/s12929-024-01055-0.
20 Yu D et al., Exosomes as a new frontier of cancer liquid biopsy, *Mol. Cancer.* **21**, 56 (2022). https://doi.org/10.1186/s12943-022-01509-9.
21 Irmer B, Chandrabalan S, Maas L, Bleckmann A, Menck K, Extracellular vesicles in liquid biopsies as biomarkers for solid tumors, *Cancers* **15**, 1307 (2023). https://doi.org/10.3390/cancers15041307.
22 Tai Y, Chen K, Hsieh J, Shen T, Exosomes in cancer development and clinical applications, *Cancer Sci.* **109**, 2364–2374 (2018). https://doi.org/10.1111/cas.13697.
23 Hoshino A et al., Extracellular vesicle and particle biomarkers define multiple human cancers, *Cell* **182**, 1044–1061.e18 (2020). https://doi.org/10.1016/j.cell.2020.07.009.
24 Shen H et al., Exosomal proteins: new targets for early diagnosis and treatment of cancer, *Front. Immunol.* **16**, 1613494 (2025). https://doi.org/10.3389/fimmu.2025.1613494.
25 Rupert DLM et al., Effective refractive index and lipid content of extracellular vesicles revealed using optical waveguide scattering and fluorescence microscopy, *Langmuir.* **34**, 8522–8531 (2018). https://doi.org/10.1021/acs.langmuir.7b04214.
26 de Rond L et al., Refractive index to evaluate staining specificity of extracellular vesicles by flow cytometry, *J. Extracell. Vesicles.* **8**, 1643671 (2019). https://doi.org/10.1080/20013078.2019.1643671.
27 Welsh JA et al., Towards defining reference materials for measuring extracellular vesicle refractive index, epitope abundance, size and concentration, *J. Extracell. Vesicles.* **9**, 1816641 (2020). https://doi.org/10.1080/20013078.2020.1816641.
28 Shao L et al., Detection of single nanoparticles and lentiviruses using microcavity resonance broadening, *Adv. Mater.* **25**, 5616–5620 (2013). https://doi.org/10.1002/adma201302572.
29 Capocefalo A, Quintiero E, Conti C, Ghofraniha N, Viola I, Droplet lasers for smart photonic labels, *ACS Appl. Mater Interfaces.* **13**, 51485–51494 (2021). https://doi.org/10.1021/acsami.1c14972.
30 Thomson CA, Popczyk A, Schubert M, Gather MC, Biointegrated microlasers: technologies, applications, and emerging developments, *Optica* **12**, 1311–1326 (2025). https://doi.org/10.1364/OPTICA.563263.
31 Guo P et al., Room–temperature red–green–blue whispering–gallery mode lasing and white–light emission from cesium lead halide perovskite ($CsPbX_3$, X = Cl, Br, I) microstructures, *Adv. Optical Mater.* **6**, 1700993 (2018). https://doi.org/10.1002/adom.201700993.
32 Li Q et al. Lasing from reduced dimensional perovskite microplatelets: Fabry-pérot or whispering-gallery-mode?, *J. Chem. Phys.* **151**, 211101 (2019). https://doi.org/10.1063/1.5127946.
33 Zhang L, Li X, Song Y, Zou B, Ultrafast antisolvent growth of single-crystal $CsPbBr_3$ microcavity for whispering-gallery-mode lasing, *Nanomaterials* **13**, 2116 (2023). https://doi.org/10.3390/nano13142116.
34 Lanty G, Bréhier A, Parashkov R, Lauret JS, Deleporte E, Strong exciton–photon coupling at room temperature in microcavities containing two-dimensional layered perovskite compounds, *New J. Phys.* **10**, 065007 (2008). https://doi.org/10.1088/1367-2630/10/6/065007.
35 Polimeno L et al., Room temperature polariton condensation from whispering gallery modes in $CsPbBr_3$ microplatelets, *Adv. Mater.* **36**, e2312131 (2024). https://doi.org/10.1002/adma.202312131.

36 Jalali M, Erni D, Early stage, label-free detection of breast cancer based on exosome's protein content alteration, in *Proceedings SPIE 12139, Optical Sensing and Detection VII*, 2022, **121390G**. https://doi.org/10.1117/12.2621062.

37 Bessif B, Jalali M, Erni D, Investigating the efficient light coupling into a microdisk in presence and absence of a polymer coating, in Proceedings SPIE 13006, *Biomedical Spectroscopy, Microscopy, and Imaging III*, 2024, 130060Y. https://doi.org/10.1117/12.3022110.

38 Garcia MM, Bessif B, Jalali M, Erni D, Detection and characterization of exosomes via a polystyrene-coated silica microresonator, *EPJ Web Conf.* **335**, 04002 (2025). https://doi.org/10.1051/epjconf/202533504002.

39 Polyanskiy MN, Refractiveindex.info database of optical constants, *Sci. Data.* **11, 94** (2024). https://doi.org/10.1038/s41597-023-02898-2.

40 Ilican S, Zor M, Caglar Y, Caglar M, Optical characterization of the CdZn $(S_{1-x}\ Se_x)_2$ thin films deposited by spray pyrolysis method, *Optica Applicata.* **36**, 29–37 (2006).

41 Jalali M, Erni D, Enhancing the sensitivity of silica microresonator-based biosensors through adding a perovskite coating, in *International Conference on Single-Molecule, Sensors and Nano Systems (S^3IC)*, 2023, 61.

42 van der Pol E et al., Absolute sizing and label-free identification of extracellular vesicles by flow cytometry, *Nanomedicine.* **14**, 801–810 (2018). https://doi.org/10.1016/j.nano.2017.12.012.

43 Nagle JF, Tristram-Nagle S, Structure of lipid bilayers, *Bioch. Biophys. Acta.* **1469**, 159–195 (2000). https://doi.org/10.1016/S0304-4157(00)00016-2.

44 González CM, Pizarro-Guerra G, Droguett F, Sarabia M, Artificial biomembrane based on DPPC – investigation into phase transition and thermal behavior through ellipsometric techniques, *Bioch. Biophys. Acta.* **1848**, 2295–2307 (2015). https://doi.org/10.1016/j.bbamem.2015.07.001.

45 Parkkila P, Elderdfi M, Bunker A, Viitala T, Biophysical characterization of supported lipid bilayers using parallel dual-wavelength surface plasmon resonance and quartz crystal microbalance measurements, *Langmuir.* **34**, 8081–8091 (2018). https://doi.org/10.1021/acs.langmuir.8b01259.

46 Kienle DF, de Souza JV, Watkins EB, Kuhl TL, Thickness and refractive index of DPPC and DPPE monolayers by multiple-beam interferometry, *Anal. Bioanal. Chem.* **406**, 4725–4733 (2014). https://doi.org/10.1007/s00216-014-7866-9.

47 Mapar M, Sjöberg M, Zhdanov VP, Agnarsson B, Höök F, Label-free quantification of protein binding to lipid vesicles using transparent waveguide evanescent-field scattering microscopy with liquid control, *Biomed. Opt. Express.* **14**, 4003–4016 (2023). https://doi.org/10.1364/BOE.490051.

48 Daimon M, Masumura A, Measurement of the refractive index of distilled water from the near-infrared region to the ultraviolet region, *Appl. Opt.* **46**, 3811–3820 (2007). https://doi.org/10.1364/AO.46.003811.

49 Barer R, Refractometry and interferometry of living cells, *J. Opt. Soc. Am.* **47**, 545–556 (1957). https://doi.org/10.1364/JOSA.47.000545.

50 Zhao H, Brown PH, Schuck P, On the distribution of protein refractive index increments, *Biophys. J.* **100**, 2309–2317 (2011). https://doi.org/10.1016/j.bpj.2011.03.004.

51 Jia X et al., Enabling online determination of the size-dependent RNA content of lipid nanoparticle-based RNA formulations, *J. Chromatogr. B.* **1186**, 123015 (2021). https://doi.org/10.1016/j.jchromb.2021.123015.

52 Tatischeff I, Extracellular vesicle-DNA: the next liquid biopsy biomarker for early cancer diagnosis? *Cancers.* **15**, 1456 (2023). https://doi.org/10.3390/cancers15051456.

53 Pleet ML, et al. Extracellular vesicle refractive index derivation utilizing orthogonal characterization, *Nano Lett.* **23**, 9195–9202 (2023). https://doi.org/10.1021/acs.nanolett.3c00562.

54 Garcia MM, Bessif B, Jalali M, Erni D, Effect of sensing location on the sensing functionality of coated microresonators for exosome detection, submitted to *Proceedings SPIE 14086: Optical Sensing and Detection IX*, (2026).

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 29
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026030
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE

OPEN ACCESS

Evaluation of surface topography and microstructure of electroplated copper ultraprecision-machined with a single-crystal diamond tool

Hideo Takino[1,*], Hidenori Aizawa[2], and Masahiko Kanaoka[3]

[1] Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan
[2] Sendai Nikon Corporation, Natori, Miyagi 981-1221, Japan
[3] JTEC Corporation, Ibaraki, Osaka 567-0086, Japan

Received 3 December 2025 / Accepted 14 March 2026

Abstract. We investigated in detail the surface characteristics of ultraprecision-machined electroplated copper (Cu). Electroplated Cu, as well as electroless-plated nickel phosphorus (NiP) and oxygen-free copper (OFC) for comparison, was machined using an ultraprecision turning machine with a single-crystal diamond tool. The resulting surfaces were measured over an area of 140 μm × 105 μm by white-light interferometry, revealing that the electroplated Cu surface exhibited the lowest surface roughness of 0.83 nm root-mean-square (RMS) among the three materials. For a smaller area of 1 μm × 1 μm measured by atomic force microscopy, the electroplated Cu surface exhibited a sand-like texture with a surface roughness of 1.74 nm RMS. Power spectral density (PSD) analysis revealed that the electroplated Cu surface had a lower PSD than the OFC surface below a spatial frequency of 2×10^4 mm^{-1} and was comparable to the electroless-plated NiP surface below a spatial frequency of 1×10^3 mm^{-1}. Electron backscatter diffraction and X-ray diffraction analyses indicated that the electroplated Cu was composed of fine crystalline microstructures, which was probably the reason for the sand-like texture. Overall, the electroplated Cu surface exhibited excellent smoothness and machinability, superior to the OFC surface and close to electroless-plated NiP surface, demonstrating its suitability for ultraprecision optical components.

Keywords: Surface, Machining, Diamond turning, Electroplated Cu, Electroless-plated NiP, Oxygen-free copper.

1 Introduction

In optical fabrication, an ultraprecision machine tool with a single-crystal diamond tool is effectively used to obtain smooth and highly accurate metal surfaces for optical components, such as those of mirrors and injection molds [1, 2]. Electroless-plated nickel phosphorus (NiP) is a widely used material for single-crystal diamond machining to achieve a smooth surface [3–14]. However, the major issues in machining electroless-plated NiP are tool wear and chipping [4, 5, 7–10, 13, 14].

To overcome this problem, alternative materials that can achieve smooth optical surfaces with lower tool wear are desirable. Because copper is softer than electroless-plated NiP, machining copper may help reduce the wear and chipping of the diamond tool. Oxygen-free copper (OFC) is one candidate material because it can be machined with a single-crystal diamond tool [15–22]. However, the machined surface of OFC often exhibits irregularities owing to crystal grains, resulting in an increase in surface roughness [15, 16, 18, 19, 21]. Although OFC has high thermal conductivity and is suitable for optical components subjected to high thermal loads when adequate cooling is applied, these surface irregularities can affect the surface quality when extremely smooth optical surfaces are required. Therefore, it is important to investigate alternative copper-based materials that can maintain the advantages of OFC while enabling the formation of smoother optical surfaces by ultraprecision diamond machining.

Electroplated copper (Cu) is therefore expected to be a promising alternative material for optical components. Because electroplated Cu generally has finer grains than OFC, it may enable the formation of smoother optical surfaces. In addition, electroplated Cu is also softer than electroless-plated NiP and is therefore expected to reduce

* Corresponding author: takino.hideo@it-chiba.ac.jp, takino.hideo@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

tool wear and chipping during diamond machining. Furthermore, when a copper electroplating layer is applied to the surface of OFC, it may enable the fabrication of optical components that combine extremely smooth optical surfaces with high thermal conductivity.

In our previous studies [23, 24], we demonstrated the successful fabricating of mirror arrays with high shape accuracy using an ultraprecision milling machine with a single-crystal diamond tool. In the machining, a mirror surface was formed on a 1-mm-thick electroplated Cu layer on a block, allowing the mirror array to be machined into the thick copper layer. The reason for the use of electroplated Cu was that it is considered to cause less tool wear and chipping than electroless-plated NiP. Moreover, we expected that the grain size of electroplated Cu would likely be smaller than that of OFC, providing smooth surfaces without irregularities. In our previous studies, we focused on the shape accuracy of the electroplated Cu surface in ultraprecision milling, and we have not studied the roughness of the machined surface in detail.

For optical components, the surface roughness after machining is a critical factor because it directly affects optical performance. To the best of our knowledge, the surface characteristics of electroplated Cu after ultraprecision diamond machining have not yet been fully clarified [23–31]. Understanding these surface characteristics is important for the design and fabrication of high-performance optical components. Therefore, the objective of this study is to clarify the surface roughness characteristics of electroplated Cu after ultraprecision machining and to compare them with those of OFC and electroless-plated NiP.

To achieve this objective, the surfaces of electroplated Cu, OFC, and electroless-plated NiP were machined by ultraprecision turning with a single-crystal diamond tool, and the resulting surface roughnesses in areas of 140 μm × 105 μm and 1 μm × 1 μm were evaluated. Moreover, the microstructure of the machined surfaces was analyzed by electron backscatter diffraction (EBSD) and X-ray diffraction (XRD) to investigate the relationship between the microstructure and the surface topography. This study provides insight into the feasibility of electroplated Cu as an alternative material for ultraprecision optical fabrication.

2 Experimental methods

The electroplated Cu specimens and the OFC specimens were cylinders with a diameter of 20 mm and a height of 10 mm. The NiP-electroless-plated specimens were cylinders with a diameter of 20 mm and a height of 7 mm. Cu plating and NiP plating were carried out by Kyowa Sangyo Corporation, Japan. The thicknesses of the electroplated Cu and electroless-plated NiP were approximately 0.2 and 0.1 mm, respectively. The substrates for the electroplated Cu and the electroless-plated NiP were OFC and stainless-steel cylinders, respectively. Cu electroplating was carried out using a plating solution in which Cu sulfate served as the main constituent. Electroless-plated NiP used in this study had a phosphorus concentration of approximately 11%.

The face of the specimens was turned with an ultraprecision turning machine with a single-crystal diamond tool having a nose radius of 0.5 mm. First, all specimens were roughly turned. After replacing the rough-turning tool with a new one, finish turning was performed using the same tool for all materials. The finishing process was performed sequentially on the OFC, electroplated Cu, and electroless-plated NiP so that electroless-plated NiP, which was considered most likely to cause tool damage, was machined last. The depth of cut was set to 5 and 2 μm in the rough and finishing processes, respectively. The tool feed rate was 2 mm/min, and the spindle speed was 1000 rpm for each process. At these tool feed rate and spindle speed, the distance between tool marks is calculated to be 2 μm.

After turning, the surface topographies of the machined surfaces were measured using a white-light interferometer (WLI) (Zygo, NewView 7300) and an atomic force microscope (AFM) (Shimazu, SPM-9700HT). In the measurement using the WLI, an area of 140 μm × 105 μm was measured with a horizontal resolution of 220 nm. In the measurement using the AFM, an area of 1 μm × 1 μm was measured with a horizontal resolution of 4 nm. Both measurements were performed at approximately the midpoint in the radial direction of the workpiece, 5 mm from the center. Moreover, the microstructure of the machined surfaces was analyzed by XRD (Rigaku, SmartLab) and EBSD (Oxford Instruments, Symmetry). XRD analysis was performed using the in-plane method. This method enables the investigation of the surface structure by limiting X-ray penetration through grazing incidence.

3 Results and discussion

3.1 Machined surface topography

Figure 1 shows the topographies of the machined surfaces measured using the WLI. Figures 1a–1c correspond to the topographies of the OFC, electroplated Cu, and electroless-plated NiP surfaces, respectively. As shown in Figure 1a, the OFC surface shows steps, which probably result from crystal grains. In Figure 1a, the height of the step through the center of the measured area is approximately 5 nm, and measurements performed at several other locations also revealed step structures showing comparable overall morphology with minor variations in contour and height. The root-mean-square (RMS) roughness of the OFC surface was 1.78 nm. As shown in Figure 1b, the electroplated Cu surface shows no steps. The RMS roughness of the electroplated Cu surface was 0.87 nm, showing a smoother surface than the OFC surface. As shown in Figure 1c, the electroless-plated NiP surface also shows no step structures, but two hills can be observed with an amplitude of approximately 3 nm and an interval of approximately 90 μm. The RMS roughness of the electroless-plated NiP surface was 1.21 nm. Since similar hills were observed in measurements at several locations, it is considered that concentric waves with a pitch of approximately 90 μm were generated on the surface. Referring again to the electroplated Cu surface shown in Figure 1b, high regions can be observed at approximately 90 μm intervals,

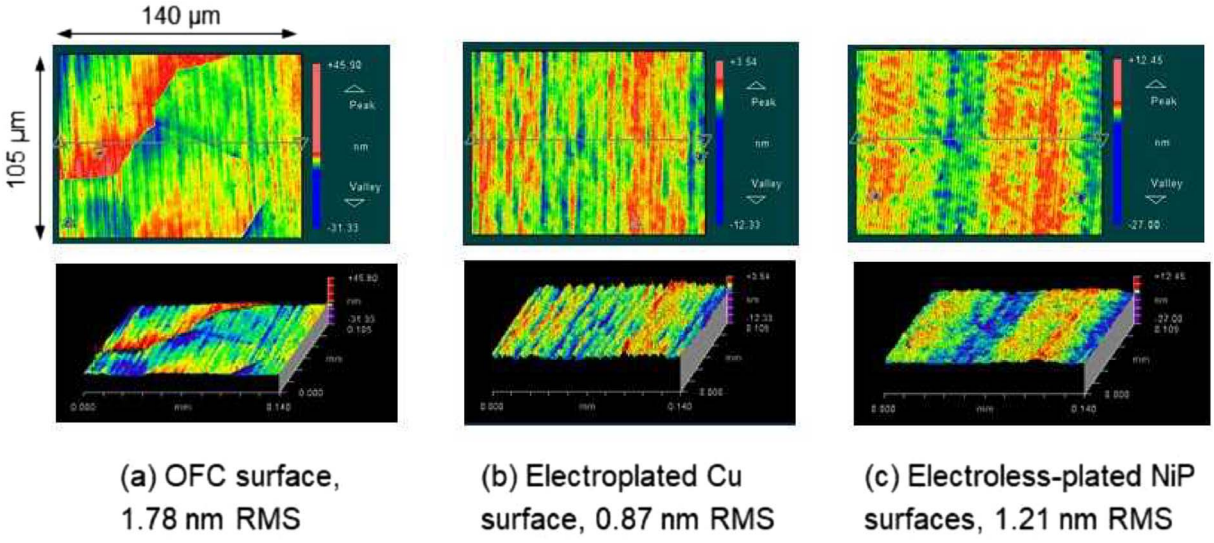


Fig. 1. Surface topographies measured with a WLI.

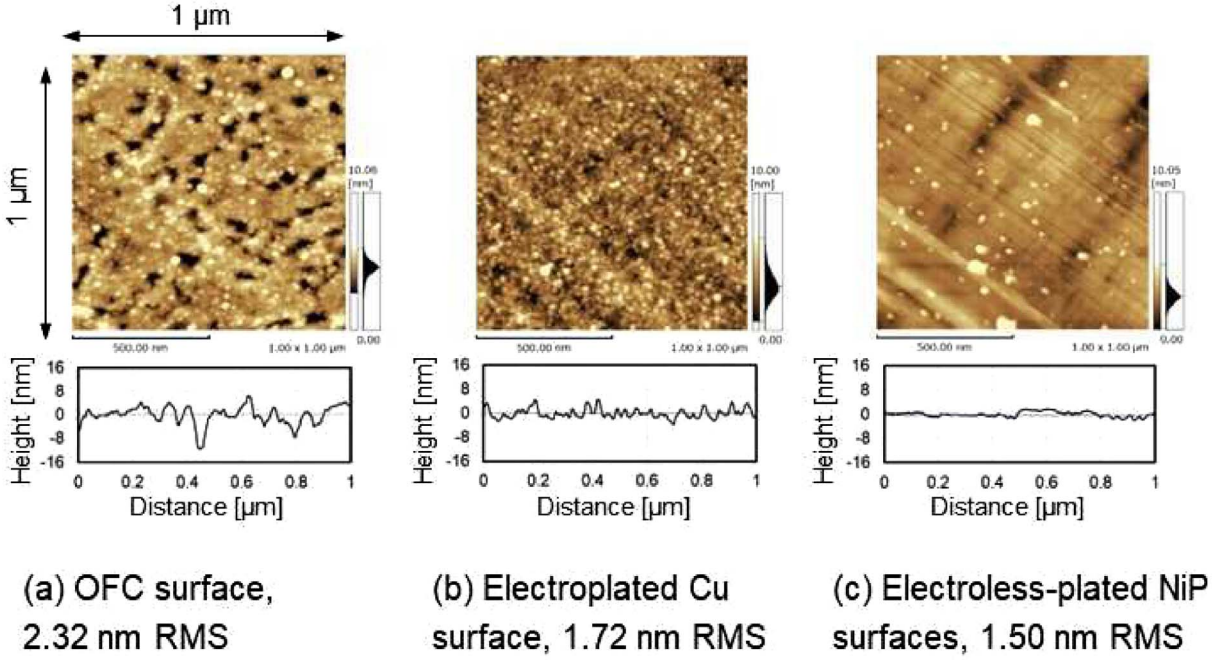


Fig. 2. Surface topographies measured with an AFM.

although they are less distinct than those on the electroless-plated NiP surface. This reduced clarity may be attributed to the softness of the electroplated Cu, which could have caused deformation during machining. For the OFC surface shown in Figure 1a, it is likely that the large step structures make it difficult to observe the hills. Although not shown here, the hills can be observed on the OFC surface when the step height is small. From these results, it is inferred that these hills, i.e., the wavinesses, were not caused by the material properties but by the motion error of the ultra-precision turning machine.

In addition, measurements were performed at three different locations near the radial midpoint of each workpiece, and the average, maximum, and minimum RMS roughness were obtained. The average, maximum, and minimum RMS roughness were 1.65, 1.78, and 1.52 nm for OFC; 0.83, 0.87, and 0.77 nm for electroplated Cu; and 1.26, 1.30, and 1.21 nm for electroless-plated NiP, respectively. The relatively large range between the maximum and minimum values for OFC is attributed to variations in the contour and height of the crystal grain boundary steps. Thus, the relative magnitudes of the average RMS roughness

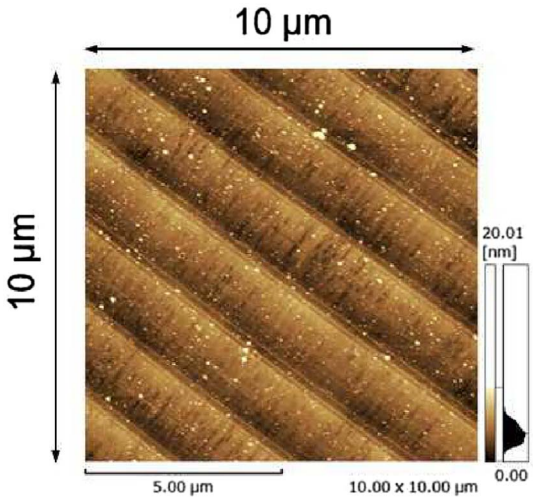


Fig. 3. Surface topography measured with an AFM: measurement result for a 10 µm × 10 µm area of the electroless-plated NiP cut surface.

values obtained from multiple measurements were consistent with those shown in Figure 1.

Next, the surface topographies within an area of 1 µm × 1 µm were measured with the AFM. Figure 2 shows the measurement results, including the surface topographies together with cross-sectional profiles approximately through their centers. The OFC surface had tiny holes, as shown in Figure 2a, whereas the electroplated Cu surface had a sand-textured finish, as shown in Figure 2b. The RMS roughness of the electroplated Cu surface was smaller than that of the OFC surface, that is, 1.72 nm and 2.32 nm, respectively. Compared with these surfaces, the electroless-plated NiP surface was smoother, as shown in Figure 2c. For the electroless-plated NiP surface, the RMS roughness was 1.50 nm. In Figure 2c, small elevations (white dots) are observed scattered on the surface. These are considered to be fine chips generated during turning adhered to the surface. After turning, the surface was only rinsed with an organic solvent using a wash bottle without wiping, in order to preserve the surface morphology as much as possible. As a result, some fine chips likely remained on the machined surface. Thus, although each machined surface exhibited a different surface appearance, the electroplated Cu surface was not as smooth as the electroless-plated NiP surface but was smoother than the OFC surface when uniformly compared in terms of RMS roughness.

Figure 3 shows the AFM measurement result for a 10 µm × 10 µm area of the electroless-plated NiP surface machined last. Regular tool marks are observed, and each tool mark exhibits a profile corresponding to the arc of the tool tip. These observations indicate that the original geometry of the cutting edge was almost maintained during the finishing process, suggesting that no significant tool wear or chipping occurred.

In addition, measurements were performed at three different locations near the radial midpoint for each material. The average, maximum, and minimum RMS roughness

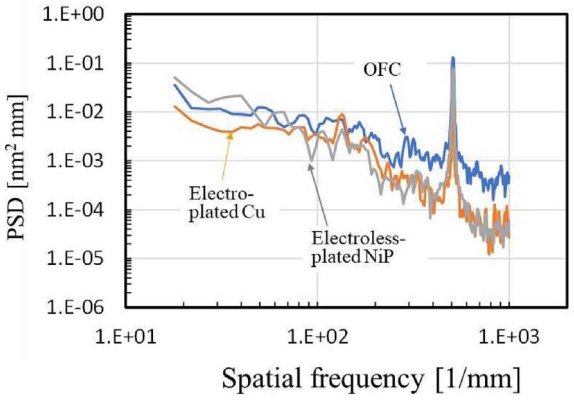


Fig. 4. PSD curves calculated from topographies of various machined surfaces measured with a WLI.

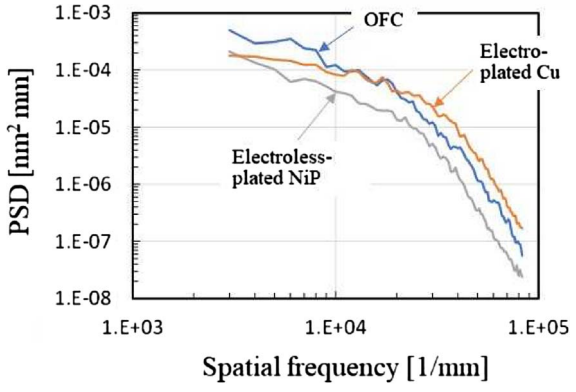


Fig. 5. PSD curves calculated from topographies of various machined surfaces measured with an AFM.

were 2.29, 2.32, and 2.27 nm for OFC; 1.74, 1.77, and 1.72 nm for electroplated Cu; and 1.44, 1.50, and 1.36 nm for electroless-plated NiP, respectively. Thus, the relative magnitudes of the average RMS roughness values obtained from multiple measurements were consistent with those shown in Figure 2.

3.2 Power spectral density (PSD) analysis

We performed the PSD analysis of the obtained surface profiles. Figure 4 shows the PSD curves calculated from the surface profiles measured with the WLI. Specifically, the one-dimensional (1D) PSD curve obtained by WLI was calculated using the software built into the measurement system. The PSD was calculated from a line profile extracted from the center of the measured surface, and the analysis was performed over the full evaluation length of the profile. The PSD curve shown represents the average of three measurements for each material.

As shown in Figure 3, the PSD of the OFC surface tends to be the highest among the three materials, whereas those of the electroplated Cu and electroless-plated NiP surfaces

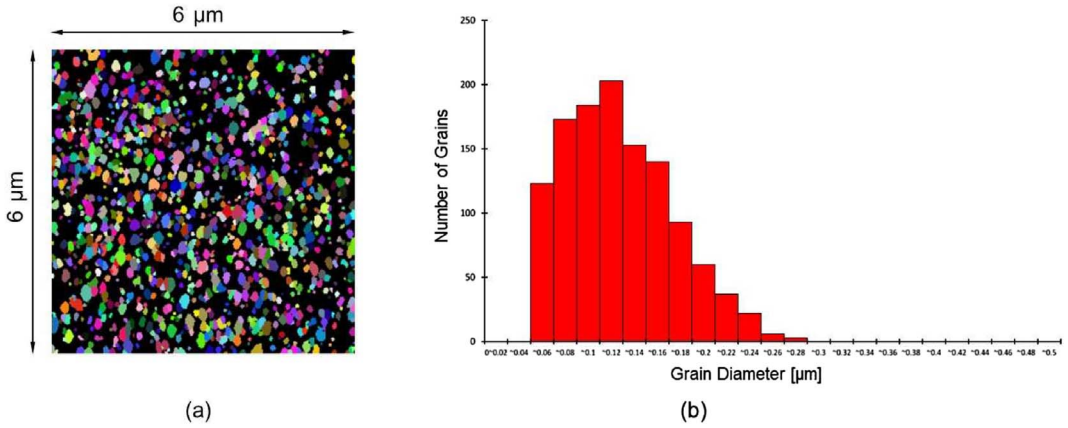


Fig. 6. Results of EBSD analysis of the machined electroplated Cu surface. (a) Grain map and (b) grain size distribution histogram.

are similarly low. Thus, the electroplated Cu and electroless-plated NiP surfaces exhibited comparable smoothness and were smoother than the OFC surface. Furthermore, the electroplated Cu and electroless-plated NiP surfaces exhibited similar spatial frequency characteristics over the entire frequency range.

In the previous section, we described that the electroplated Cu and electroless-plated NiP surfaces exhibited hills at intervals of approximately 90 µm. Although the peaks corresponding to these hills exist, the peaks do not appear in the PSD curves because they must be located in the low-frequency region outside the plotted range. However, it is possible that the increase at the left end of the PSD curve of the electroless-plated NiP surface is related to the presence of these hills. Moreover, Figure 4 shows that each PSD curve exhibits a similar sharp peak at a spatial frequency of approximately 5.0×10^3 mm^{-1}. This corresponds to a tool-mark interval of about 2 µm attributed to turning.

Figure 5 shows the PSD curves calculated from the surface profiles measured with the AFM. Similarly, the one-dimensional (1D) PSD curves obtained from the AFM measurements were calculated using in-house software. The software calculates the PSD for all cross-sectional profiles across the measured surface, and the analysis was performed over the full evaluation length of each profile. The PSD curves shown in the figure represent the average of these profiles and correspond to the surface data of the measured area shown in Figure 2. The PSD of the OFC surface was the highest among all materials at spatial frequencies below 1×10^4 mm^{-1}, likely attributable to the presence of the tiny holes on the surface. For the OFC surface, the high PSD in the low-frequency region below 1×10^4 mm^{-1} is considered to result in the largest RMS value among the three materials. On the other hand, the PSD of the electroplated Cu surface was the highest among all materials at spatial frequencies above 2×10^4 mm^{-1}, which is considered to result from the sand-like texture appearing on the surface.

Overall, from the PSD analysis results shown in Figures 4 and 5, it can be concluded that the PSD of the electroplated Cu surface is lower than that of the OFC surface at spatial frequencies below 2×10^4 mm^{-1}. Furthermore, at spatial frequencies below 1×10^3 mm^{-1}, the PSD of the electroplated Cu surface is as low as that of the electroless-plated NiP surface, indicating comparable surface quality.

3.3 Material microstructural analysis

The microstructure of the machined electroplated Cu was analyzed by EBSD. This analysis was carried out on the machined surfaces without post-treatments such as polishing and etching, because the machined electroplated Cu was smooth enough for analysis. Figure 6 shows the results of the EBSD analysis, in which (a) the grain map over an area of 6 µm × 6 µm and (b) its grain size distribution histogram are shown. EBSD data were analyzed by defining misorientations exceeding 5° as grain boundaries. The measurements were conducted by scanning an electron beam at a pitch of 20 nm, and the lower detection limit of the EBSD instrument was 40 nm. Thus, fine grains with sizes ranging from 40 to 280 nm were detected on the electroplated Cu. The black areas in Figure 6a are considered to correspond to either micro-grains smaller than 40 nm or amorphous regions.

Next, the microstructure of the machined electroplated Cu surface was analyzed by XRD. The electroless-plated NiP surface was also analyzed as a reference for comparison; in general, the electroless-plated NiP with a phosphorus concentration of 11% is amorphous [32, 33]. The XRD analysis was also carried out for the machined surfaces without post-treatments. Figures 7a and 7b show the XRD profiles of the electroless-plated NiP and electroplated Cu surfaces, respectively. In Figure 7a, the XRD profile of the electroless-plated NiP surface is broad, which indicates that the electroless-plated NiP is amorphous, as generally known. In contrast, the XRD profile of the electroplated Cu surface in Figure 7b is not broad. Instead, multiple peaks attributed to Cu were detected. This indicates that there was no amorphous region on the electroplated Cu surface, and the electroplated Cu is crystallized and exists in a polycrystalline state. Table 1 shows the peak fitting results for the major Cu-related reflections shown in

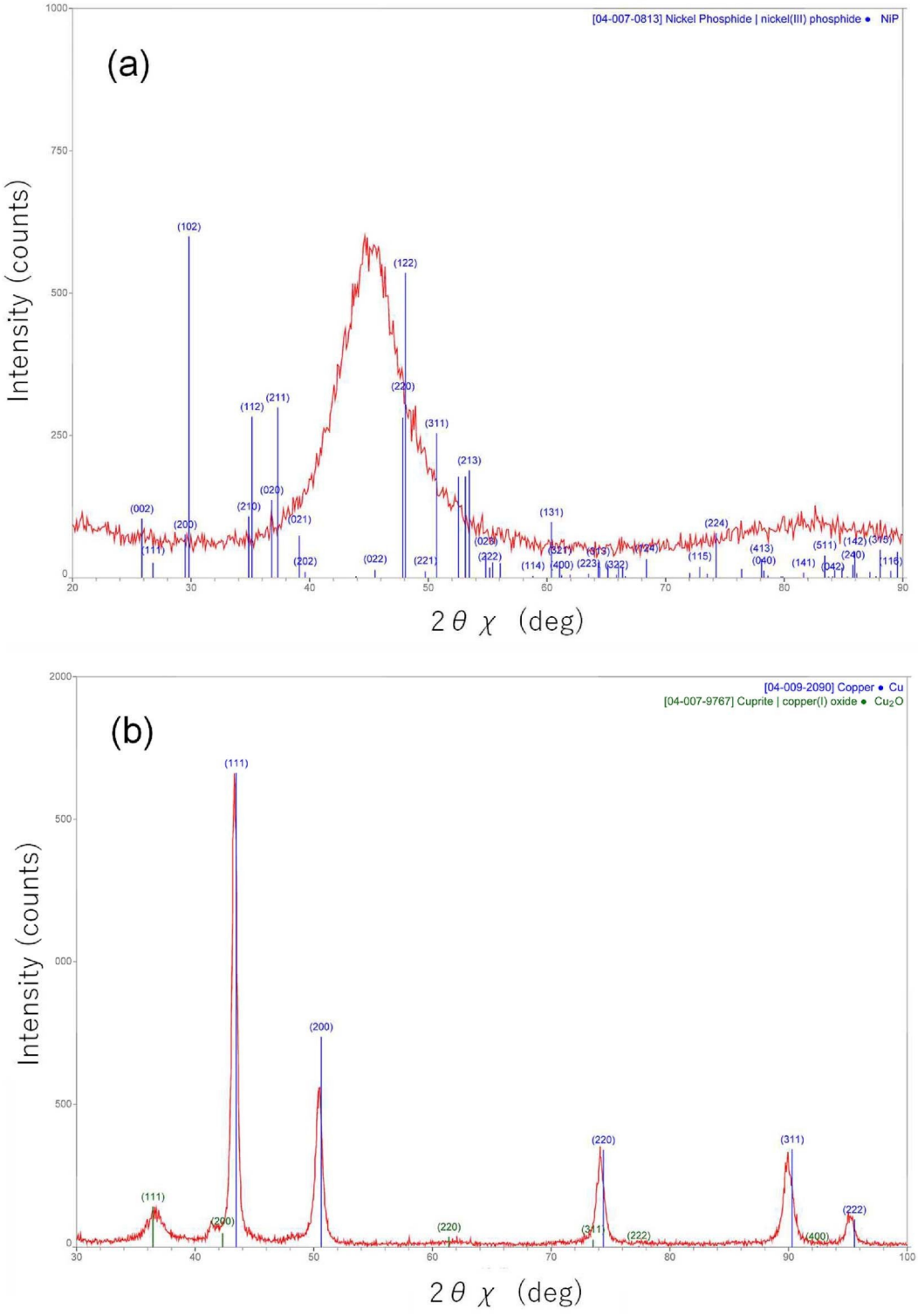


Fig. 7. XRD profiles of the machined (a) electroless-plated NiP and (b) electroplated Cu surfaces.

Figure 7b. Table 1 lists the relative intensities of other reflections normalized to the (111) reflection and the average crystallite sizes calculated from the full width at half maximum (FWHM) of each peak. The average crystallite size for each reflection ranged from 20 to 47 nm. Thus, the absence of an amorphous phase on the electroplated Cu indicates that the black regions in the EBSD grain map in Figure 6 are not amorphous, but they are micrograins smaller than 40 nm, which is below the measurement limit.

Table 1. Peak fitting results of the obtained XRD profile of electroplated Cu surface.

(h k l)	$2\theta\chi$ (°)	d (Å)	Height	Area	Area (%)	FWHM (°)	Crystalline size (nm)
(1 1 1)	43.279	2.0888	1156	15563	100.0	0.468	47
(2 0 0)	50.831	1.8098	370	6875	44.2	0.646	23
(2 2 0)	74.076	1.2788	210	4250	27.3	0.695	23
(3 1 1)	89.877	1.0905	207	4771	30.7	0.822	20
(2 2 2)	95.099	1.0440	71	1208	7.8	0.703	26

The difference between the EBSD and XRD results arises from both their measurement principles and the nature of what they detect. EBSD identifies grains on the basis of orientation differences and merges slightly misoriented crystallites into a single grain if the misorientation is below the threshold angle. In contrast, XRD estimates the size of crystallites, i.e., coherently diffracting domains within a grain that correspond to subdomains of a crystal. Consequently, the crystallite sizes obtained by XRD analysis tend to be smaller, whereas the grain sizes measured by EBSD analysis appear larger. In any case, the electroplated Cu surface exhibits fine crystalline microstructures; thus, the sand-like texture observed on the machined electroplated Cu surface, as shown in Figure 2b, is considered to result from these fine microstructures. Moreover, a possible reason for the absence of the step structures on the machined electroplated Cu surface, as shown in Figure 1b, is that its crystal grains are too small to significantly affect the machining process.

4 Conclusions

We investigated the roughness and microstructures of the electroplated Cu machined by ultraprecision diamond turning, as well as those of the OFC and electroless-plated NiP for comparison.

The average RMS surface roughnesses measured over a large area of 140 μm × 105 μm using a WLI in ascending order were 0.83 nm for electroplated Cu, 1.26 nm for electroless-plated NiP, and 1.65 nm for OFC. Step structures were observed on the OFC surface, probably caused by crystal grains, leading to larger RMS roughnesses. Concentric waves were observed on the NiP plating; however, they are considered to be caused not by the material itself but by the motion error of the ultraprecision turning machine. The electroplated Cu surface appeared the smoothest with no step structures.

For a smaller area of 1 μm × 1 μm measured using an AFM, the average RMS surface roughnesses in ascending order were 1.44 nm for electroless-plated NiP, 1.74 nm for electroplated Cu, and 2.29 nm for OFC. The OFC surface showed tiny holes, whereas the electroplated Cu surface showed a sand-textured finish. In comparison, the electroless-plated NiP surface was smoother than the other two.

Focusing on the electroplated Cu, the results of PSD analysis can be summarized as follows. At spatial frequencies below 2×10^4 mm^{-1}, the PSD of the electroplated Cu surface is lower than that of the OFC surface. When the range is narrowed to spatial frequencies below 1×10^3 mm^{-1}, the PSD of the electroplated Cu surface becomes comparable to that of the electroless-plated NiP surface and remains lower than that of the OFC surface.

Microstructural analysis by EBSD revealed fine crystal grains on the electroplated Cu surface ranging from 40 to 280 nm, whereas XRD indicated smaller crystallite sizes of 20–47 nm, confirming that the electroplated Cu is fully crystallized and polycrystalline. The discrepancy between the EBSD and XRD results arises from differences in their measurement principles. In any case, the electroplated Cu was found to be composed of microcrystalline grains, as confirmed by both EBSD and XRD analyses. Therefore, the sand-like texture observed on the machined electroplated Cu surface is likely due to the presence of these microcrystalline grains. Moreover, a possible reason for the absence of the step structures on the machined electroplated Cu surface is that its crystal grains are too small to significantly affect the machining process.

Overall, the findings demonstrate that electroplated Cu provides superior surface smoothness and machinability compared with OFC, although it is slightly inferior to electroless-plated NiP at high spatial frequencies, highlighting its suitability for the ultraprecision machining of optical components, where both high shape accuracy and small surface roughness are essential.

Acknowledgments

We thank Hideyuki Aizawa of Sendai Nikon Corporation for turning the specimens. We also thank Hideki Munakata of Chiba Institute of Technology for measuring the surface topographies.

Funding

This research received no external funding.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Data availability statement

The data that support the findings of this study, including those not shown in this article, are available from the corresponding author upon reasonable request.

Author contribution statement

HT conceptualized and supervised the study, participated in all experiments, interpreted all the data, and wrote the manuscript.

HA supervised and managed the machining experiments, and provided expert advice. MK supervised and managed the measurements, interpreted the data, and provided expert advice. All authors have read and approved the final manuscript.

References

1 Rhorer R, Evans C, Fabrication of optics by diamond turning, in *Handbook of Optics*, 3rd edn. **Vol. II**: Design, Fabrication and Testing, Sources and Detectors, Radiometry and Photometry, edited by Bass M, DeCusatis C, Enoch J, Lakshminarayanan V, Li G, Macdonald C, Mahajan V, Stryland EV, (McGraw-Hill, 2009), Chap. 10.

2 Lucca DA, Klopfstein MJ, Riemer O, Ultra-precision machining: cutting with diamond tool, *ASME. J. Manuf. Sci. Eng.* **142**, 11 (2020). https://doi.org/10.1115/1.4048194.

3 Syn C, Saito T, Taylor J, Donaldson R, Materials properties influence on smoothness of diamond-turned electroless nickel, in *Optical Fabrication and Testing Workshop, Technical Digest Series* Optica Publishing Group, 1985), paper ThBB6. https://doi.org/10.1364/OFT.1985.ThBB6.

4 Taylor JS, Syn CK, Saito TT, Donaldson RR, Surface finish measurements of diamond-turned electroless-nickel-plated mirrors, *Opt. Eng.* **25**, 9 (1986). https://doi.org/10.1117/12.7973947.

5 Syn CK, Taylor JS, Donaldson RR, Diamond tool wear vs. cutting distance on electroless nickel mirrors, in *Proc. SPIE 0676, Ultraprecision Machining and Automated Fabrication of Optics*, (1987). 10.1117/12.939527.

6 Eda H, Kishi K, Ueno H, Masubuchi S, Ultra precision cutting of electroless nickel plating layer, *J. Jpn. Soc. Precis. Eng.* **53,** 5 (1987). (in Japanese) https://doi.org/10.2493/jjspe.53.814.

7 Eda H, Kishi K, Ueno H, Oyatsu M, For a long life of natural single crystal diamond tool—the investigation of ultra precision cutting for electroless Ni-P, *J. Jpn. Soc. Precis. Eng.* **55,** 1 (1989). (in Japanese) https://doi.org/10.2493/jjspe.55.161.

8 Pramanik A, Neo KS, Rahman M, Li XP, Sawa M, Maeda Y, Cutting performance of diamond tools during ultra-precision turning of electroless-nickel plated die materials, *J. Mater. Process. Technol.* **140**, 1–3 (2003). https://doi.org/10.1016/S0924-0136(03)00751-9.

9 Chon KS, Namba Y, Single-point diamond turning of electroless nickel for flat X-ray mirror, *J Mech Sci Technol.* **24**, 8 (2010). https://doi.org/10.1007/s12206-010-0512-3.

10 Pramanik A, Neo KS, Rahman M, Li, XP, Sawa M, Maeda Y, Ultra-precision turning of electroless-nickel: Effect of phosphorus contents, depth-of-cut and rake angle, *J. Mater. Process. Technol.* **208**, 1–3 (2008). https://doi.org/10.1016/j.jmatprotec.2008.01.006.

11 Yan J, Oowada T, Zhou T, Kuriyagawa T, Precision machining of microstructures on electroless-plated NiP surface for molding glass components, *J. Mater. Process. Technol.* **209** (2009). https://doi.org/10.1016/j.jmatprotec.2008.12.008.

12 Brinksmeier E, Gläbe R, Schönemann L, Diamond micro chiseling of large-scale retroreflective arrays, *Precis. Eng.* **36**, 4 (2012). https://doi.org/10.1016/j.precisioneng.2012.06.001.

13 Chon KS, Takahashi H, Namba Y, Wear inspection of a single-crystal diamond tool used in electroless nickel turning, *Opt. Eng.* **53**, 3 (2014). https://doi.org/10.1117/1.OE.53.3.034102.

14 Milan N, Sorgato M, Parenti P, Annoni M, Lucchetta G, Effects of micromilled NiP mold surface topography on the optical characteristics of injection molded prismatic retroreflectors, *Precis. Eng.* **61** (2020). https://doi.org/10.1016/j.precisioneng.2019.10.006.

15 Decker DL, Bennett JM, Soileau MJ, Porteus JO, Bennett HE, Surface and optical studies of diamond-turned and other metal mirrors, *Opt. Eng.* **17**, 2 (1978). https://doi.org/10.1117/12.7972204.

16 Moriwaki T, Okuda K, Machinability of copper in ultra-precision micro diamond cutting, *Ann. CIRP*, **38**, 1 (1989). https://doi.org/10.1016/S0007-8506(07)62664-X.

17 Sugano T, Takeuchi K, Yoshida Y, Diamond turning of oxygen-free copper for mirrors, *JSME Int. J. Ser. C*, **36**, 4 (1993). https://doi.org/10.1299/jsmec1993.36.549.

18 Zhang H, Zhang X, Factors affecting surface quality in diamond turning of oxygen-free high-conductance copper, *Appl. Opt.* **33**, 10 (1994). https://doi.org/10.1364/AO.33.002039.

19 Zhang X, Zhang Y, Study on the surface quality of a diamond-turned oxygen-free high-conductance copper reflector used in a high-power CO2 laser, *Opt. Eng.* **36**, 3 (1997). https://doi.org/10.1117/1.601277.

20 Tanaka H, Shimada S, Higuchi M, Yamaguchi T, Kaneeda T, Obata K, Mechanism of cutting edge chipping and its suppression in diamond turning of copper, *Ann. CIRP*, **54**, 1 (2005). https://doi.org/10.1016/S0007-8506(07)60047-X.

21 Ding X, Jarfors AEW, Lim GC, Shaw KC, Liu YC, Tang LJ, A study of the cutting performance of poly-crystalline oxygen free copper with single crystalline diamond micro-tools, *Precis. Eng.* **36**, 1 (2012). https://doi.org/10.1016/j.precisioneng.2011.09.001.

22 Mishra V, Biswas AK, Kumar N, Kukreja LM, Sarepaka RV, Fabrication of $\lambda/2$ phase step mirror for CO_2 laser resonator using diamond turning, *Opt. Eng.* **53**, 3 (2014). https://doi.org/10.1117/1.OE.53.3.036107.

23 Takino H, Kawai T, Takeuchi Y, 5-axis ultraprecision machining of complex-shaped mirrors for extreme ultraviolet lithography system, *Ann. CIRP*, **56** 1 (2007). https://doi.org/10.1016/j.cirp.2007.05.031.

24 Takino H, Takeuchi Y, Machining of smooth optical surfaces by ultraprecision milling with compensated feeding mechanisms, *Int. J. Autom. Technol.* **13**, 2 (2019). https://doi.org/10.20965/ijat.2019.p0185.

25 Saito TT, Simmons LB, Performance characteristics of single point diamond machined metal mirrors for infrared laser applications, *Appl. Opt.* **13**, 11 (1974). https://doi.org/10.1364/AO.13.002647.

26 Waldrop F, Bezik M, Tewes W, Waldrop R, Electroplating application to the fabrication of optics, *Appl. Opt.* **14**, 8 (1975). https://doi.org/10.1364/AO.14.001783.

27 Sollid JE, Sladky RE, Reichelt WH, Singer S, Single-point diamond-turned copper mirrors: figure evaluation, *Appl. Opt.* **15**, 7 (1976). https://doi.org/10.1364/AO.15.001656.

28 Kimura WD, Saito TT, Glancing incidence measurements of diamond turned copper mirrors, *Appl. Opt.* **26**, 4 (1987). https://doi.org/10.1364/AO.26.000723.

29 Frost F, Takino H, Fechner R, Schindler A, Ohi N, Nomura K, Smoothing of diamond-turned copper surfaces using ion beams with aid of planarizing film, *Jpn. J. Appl. Phys.* **46**, 9A (2007). https://doi.org/10.1143/JJAP.46.6071.

30 Ebihara K, Yamamoto A, Kawai T, Takeuchi Y, Study on high efficiency machining with reaction-free shuttle unit,

J. Jpn. Soc. Precis. Eng. **73,** 2 (2007). (in Japanese) https://doi.org/10.2493/jjspe.73.220.

31 Yoshikawa T, Kyoi M, Tanaka H, Onozuka H, Maeda Y, Taya M, Ultra precision cutting of roll die with micro lens array for plastic film, *JSME Int. J. Ser. C*, **77**, 777 (2011). https://doi.org/10.1299/kikaic.77.2146.

32 Mai QX, Daniels RD, Harpalani HB, Structural changes induced by heating in electroless nickel-phosphorus alloys, *Thin Solid Films*, **166**, 1 (1988). https://doi.org/10.1016/0040-6090(88)90384-7.

33 Bredael E, Blanpain B, Celis JP, Roos JR, On the amorphous and crystalline state of electrodeposited nickel-phosphorus coatings, *J. Electrochem. Soc.* **141**, 1 (1994). https://doi.org/10.1149/1.2054703.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 11
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026007
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

Comparison of seismic records obtained by distributed acoustic sensing and ocean bottom seismometers

Orlando Frazão[1,*], Susana Silva[1], Carlos Corela[2], Afonso Loureiro[2,3], Susana Gonçalves[2], Paulo Robalinho[1], Ricardo Sousa[1], Hugo F. Martins[4], Fernando Carrilho[5], Rachid Omira[5], Manfred Niehus[6,7], and Luis Matias[2]

[1] INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, Rua do Campo Alegre, 687, 4150-179 Porto, Portugal
[2] Instituto Dom Luiz, University of Lisbon, 1749-016 Lisbon, Portugal
[3] Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação, Funchal, Portugal
[4] Instituto de Óptica "Daza de Valdés" IO-CSIC, C/Serrano 121, 28006 Madrid, Spain
[5] Instituto Português do Mar e da Atmosfera, Rua C ao aeroporto, 1749-077 Lisbon, Portugal
[6] IT Instituto de Telecomunicações Lisboa, Campus Alameda, Torre Norte - Piso 10, Av. Rovisco Pais, 1, 1049 - 001 Lisboa, Portugal
[7] ISEL Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emídio Navarro 1, 1959-007 Lisboa, Portugal

Received 3 December 2025 / Accepted 23 January 2026

Abstract. This work presents an experimental framework for offshore seismic monitoring that combines Distributed Acoustic Sensing (DAS) with ocean-bottom seismometers (OBS). The study was conducted in the Azores region – Faial, where an HDAS interrogator prototype was connected to dark fiber submarine fiber-optic cable, complemented by the installation of two Ocean Bottom Seismometers (OBS) for calibration and validation of DAS technology. The main objective is to demonstrate that seismic observations obtained by DAS from seafloor cables can provide essential information similar to OBS and particularly in areas where land-based monitoring stations are limited.

Keywords: Distributed acoustic sensing, Submarine cable, Ocean-bottom seismometers, Sismology.

1 Introduction

Distributed Acoustic Sensing (DAS) enables fiber-optic cables to be used as dense, distributed arrays of acoustic sensors via a simple OTDR-based approach, offering clear advantages in marine environments where conventional instrumentation is difficult to deploy [1]. Because the fiber acts simultaneously as the sensing element and the transmission medium, and its optical properties respond to external disturbances, careful validation and calibration are required for quantitative applications such as earthquake magnitude estimation [2]. DAS has been successfully used for detecting and locating earthquakes, imaging subsurface structure, and monitoring natural hazards, with recent studies extending its use to hypocenter estimation, magnitude proxies, and rupture characterization from strain/strain-rate data [3]. Typical sensing ranges lie between ∼30–40 km and ∼100 km, limited by fiber attenuation and Rayleigh backscatter degradation [4]. Distributed Raman amplification can extend the range beyond ∼100 km [5], and with the use of pulse compression it can reach distances of >150 km [6]. Ocean Bottom Seismometers (OBS) are autonomous platforms that record ground motion on the seafloor, enabling offshore seismicity monitoring and lithospheric studies in areas lacking land networks. They typically integrate a 3-component broadband seismometer and a hydrophone, improving the separation of seismic, hydrodynamic, and acoustic signals [7]. In this work, we describe the experimental framework for monitoring using both DAS and OBS. We then summarize the main properties of the DAS and OBS datasets and discuss preliminary earthquake analyses as recorded by both systems. Finally, we demonstrate that in regions such as the Azores – Faial where seismic monitoring is largely restricted to land stations – DAS observations from offshore cables can provide valuable information for real-time operations and for deferred-time investigations of offshore seismicity.

2 Installation of the HDAS and OBS on the submarine cable

The submarine cable used in this study comprises three fiber pairs of standard single-mode optical fiber (SMF-28).

* Corresponding author: orlando.frazao@inesctec.pt

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

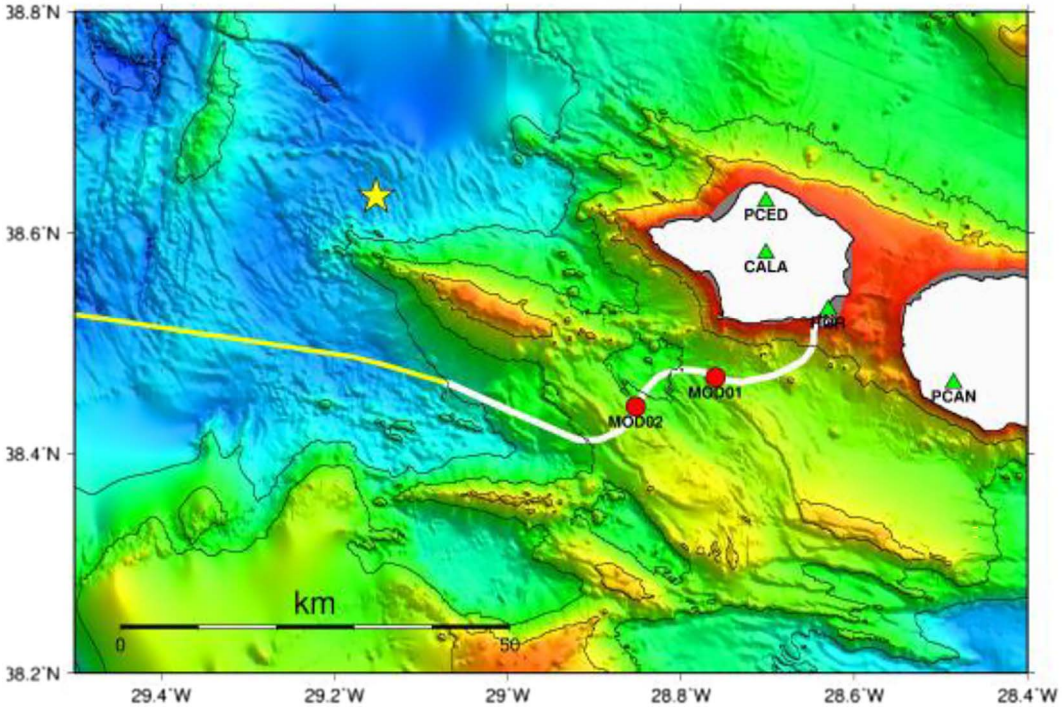


Fig. 1. Experiment setting in Faial, Azores. The monitored cable by HDAS is shown in thick white. The two OBS are shown in red. The land stations are shown by green triangles. The epicenter location is highlighted with a yellow star.

The specified attenuation is 0.12 dB/km at 1550 nm. The cable is 170 km long, unpowered, and contains no repeaters. It was laid mostly on the ocean floor, with protection levels that vary with seafloor depth; the depth increases rapidly away from the coastline. In Figure 1, the monitored cable by HDAS is shown as a thick white line. The two OBS are shown as red dots. The three land stations are shown as green triangles. The epicentre location is highlighted with a yellow star. DAS data was collected on a dark fiber from an unused telecom fiber pair. At the ALTICE cable-landing station, we connected a HiFi Distributed Acoustic Sensor (HDAS) to the Fibroglobal terminal equipment. The HDAS is a chirped-pulse ϕ-OTDR interrogator prototype developed by UAH and CSIC [8]. Figure 1 presents the Experimental setting in Faial Azores. The monitored cable by HDAS is shown in thick white. The two OBS are shown in red. The land stations are shown by green triangles.

An FC/APC connector was used to minimize back-reflections at the interface with the submarine cable, preventing signal degradation and potential damage to the HDAS photodetector. The interrogator is identical to the CP-DAS system described in [9], based on ϕ-OTDR using chirped optical pulses, a patented technology. A patented Raman-amplification range-extension module increases the sensing reach from 35 to 40 km to up to 70 km per fiber access. The manufacturer specifies a strain resolution better than 5 *ne* (nanostrain). In this experiment, the HDAS interrogated ∼50 km of cable with a 10 m gauge length and 10 m spatial sampling, yielding 4,992 channels, with an acoustic sampling frequency of 1 kHz. The strain time series were later downsampled to 50 Hz, which was sufficient to resolve the seismic signals of interest while reducing data volume. For calibration and validation of the HDAS recordings, the MODAS project deployed two OBS units developed at IDL, advanced seabed instruments designed for long-term offshore monitoring. Each OBS records ground motion with a broadband seismometer and seismic–acoustic signals in the water column with a hydrophone. The deployment sites (Fig. 1) lie approximately 500 m from the submarine cable, allowing direct comparison between OBS and DAS measurements. Accurate timing for joint analyses with HDAS was ensured by synchronizing the OBS internal clock at the start and end of the experiment. The seismometers recorded leveled ground-velocity data on the three components (N, E, Z) at 250 Hz, while the hydrophone channel was sampled at 100 Hz. Each OBS includes an HTI-04-PCA/ULF hydrophone (100 s–8 kHz), a Güralp Certis broadband seismometer with a 120-s corner period, and a Güralp Certimus data logger. Accurate timing for joint analyses with HDAS was ensured by synchronizing the OBS internal clock at the start and end of the experiment.

3 Preliminary analysis of earthquake data

The DAS and OBS recorded several earthquakes over a period of five months. Figure 2 shows one of these earthquakes, recorded at around 12:36 a.m. on September 24, 2024, with a magnitude of 2.7 on the Richter scale and an epicenter located 20 km from the submarine cable, according to information provided by *Instituto Português do Mar e Atmosfera* (IPMA). However, we can also observe

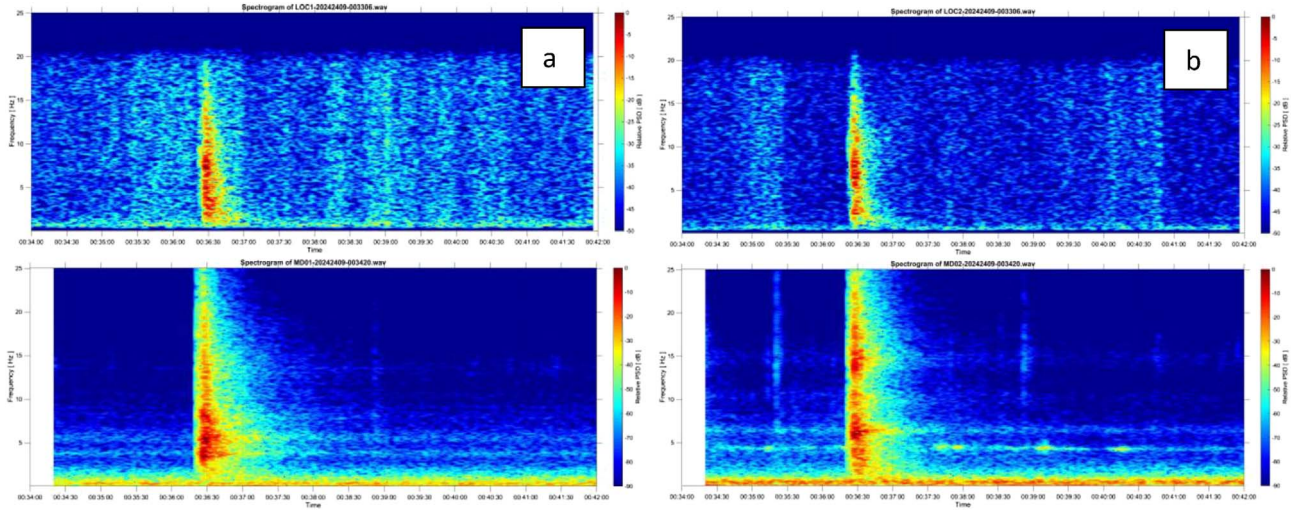


Fig. 2. Spectrogram comparison results of an earthquake with a magnitude of 2.7 read in a DAS (top) and on the two OBS (bottom) in LOC1 and MAD01 (a) and LOC2 and MAD02 (b).

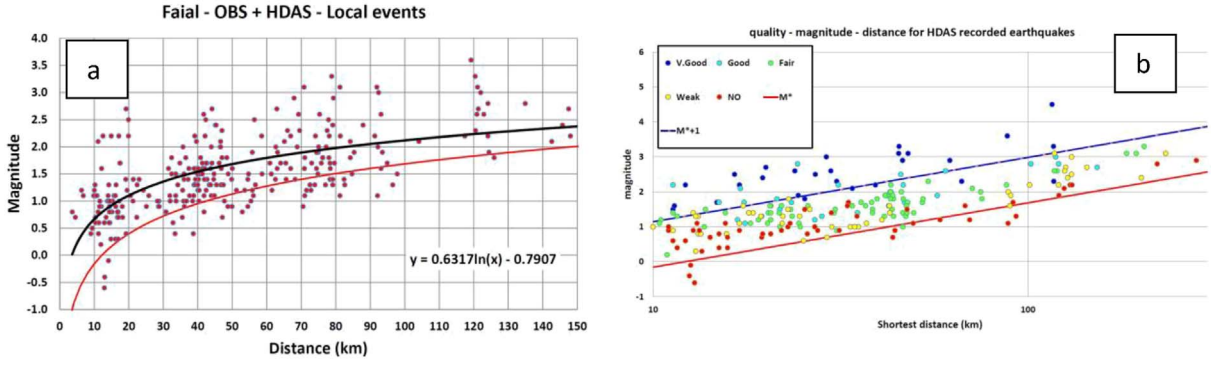


Fig. 3. a) Earthquake magnitude versus distance for events detected by land stations during the HDAS–OBS overlap. The black line shows the best-fit logarithmic curve, and the red line indicates the sensitivity threshold. b) Magnitude versus distance for the 229 earthquakes examined for HDAS recordings. Colors represent waveform quality (see text). The red line shows the proposed logarithmic reference threshold law M*. In blue we show M*+1 in magnitude, which is close to the threshold for good-quality P-waves recorded by HDAS.

other energy signals that have not yet been identified. By comparing the two OBS locations, we see that their responses are similar, with a time delay that is expected due to the relative positioning of the OBS units. During the five-month interval in which the HDAS and the OBS were recording simultaneously, the IPMA network – comprising land-based stations on the neighboring islands (see Fig. 1 for locations) – detected 229 local and regional earthquakes. To quantitatively assess the network's sensitivity and detection threshold in this area, we examined the relationship between event magnitude and the minimum distance to the submarine cable. Specifically, by plotting magnitude as a function of the closest distance to the cable (see Fig. 3), we can evaluate how sensitivity decreases with increasing source–receiver separation. This allowed us to delineate the magnitude–distance envelope below which events are no longer consistently recorded. This provides a useful baseline for comparing the performance of DAS/HDAS and OBS observations with that of the land network, particularly for offshore seismicity where terrestrial instrumental coverage is inherently limited.

A sensitivity-threshold function for the land network was derived (red curve in Fig. 3a), empirically outlining the magnitude–distance range in which onshore stations reliably detect earthquakes; events falling below this curve are increasingly likely to go undetected. In parallel, HDAS recordings were examined for 229 events near the OBS sites (LOC1 near MAD01 and LOC2 near MAD02), where cross-method comparison is strongest. Each event's HDAS signal quality was classified into five categories: Very good (clear P and S waves on both cable sections), Good (at least one identifiable P wave plus clear S waves), Fair (well-defined S waves but no clear P wave), Poor (weak, poorly defined S waves), and None (no recognizable seismic

signal).This classification converts the qualitative inspection of the recordings into a metric that is comparable across events, highlighting the influence of source–receiver distance and magnitude on HDAS detectability. The results are displayed in Figure 3b as a function of distance and magnitude and are compared with the proposed reference threshold law. This comparison is important because it shows the extent to which HDAS matches (or exceeds) the sensitivity of the land network, particularly for offshore seismicity, where DAS is expected to offer clear advantages due to the proximity of the cable to the seismic sources.

4 Conclusion

Over five months, the DAS and OBS recorded several earthquakes, including a magnitude 2.7 event that fell on Faial Island. In addition to these earthquakes, other unidentified energy signals were detected. The two OBS stations show similar responses, with a time delay expected due to their spatial separation. During the same period, the IPMA network detected 229 local and regional earthquakes. The relationship between event magnitude and the minimum distance to the submarine cable was analyzed to assess the sensitivity and detection threshold in the area. This analysis defines the magnitude–distance limit beyond which events are no longer consistently recorded, providing a reference for comparing the performance of DAS/HDAS and OBS observations with that of the land-based network, especially for offshore seismicity.

Funding

This work is supported by the Portuguese Fundação para a Ciência e Tecnologia, FCT, I.P./MCTES through national funds (PIDDAC): UID/50019/2025 and LA/P/0068/2020 (https://doi.org/10.54499/LA/P/0068/2020), by the MODAS project 2022.02359.PTDC, and by EC project SUBMERSE project HORIZON-INFRA-2022-TECH-01-101095055. HFM acknowledges financial support from the MCIN/AEI/10.13039/5011000 11033 and European Union «NextGenerationEU» /PRTR under grants RYC2021-035009-I, CPP2022-009772, PID2024-162301OA-C22, CPP2024-011822. This work is co-funded by Component 5 – Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 – 2026, within project OBSERVA, with reference 2024.07610.IACDC (https://doi.org/10.54499/2024.07610.IACDC).

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

Data are available upon reasonable request at the indicated e-mail.

Author contribution statement

Conceptualization, L. Matias and O. Frazão; Methodology, C. Corela and F. Carrilho; Software, A. Loureiro and R. Sousa; Validation, L. Matias and M. Niehus; Formal Analysis, H. Martins; Investigation, C. Corela, A. Loureiro, P. Robalinho, F. Carrilho and R. Omira; Writing – Original Draft Preparation, O. Frazão and S. Silva; Writing – Review & Editing, L. Matias, M. Niehus and O. Frazão; Visualization, H. Martins; Supervision, L. Matias; Project Administration, L. Matias.

References

1 Juarez JC, Maier EW, Choi KN, Taylor HF, Distributed fiber-optic intrusion sensor system using phase-sensitive OTDR (ϕ-OTDR), *J. Lightwave Technol.* **23**(6), 2081–2087 (2005). https://doi.org/10.1109/JLT.2005.849924.

2 Matthaiou I, Masoudi A, Araki E, Kodaira S, Modafferi S, Brambilla G, Classification of images derived from submarine fibre optic sensing: detecting broadband seismic activity from hydroacoustic signals, *Geophys. J. Int.* **240**(1), 483–501 (2025). https://doi.org/10.1093/gji/ggae400.

3 Taha BA, Addie AJ, Haider AJ, Osman SA, Ramli MZ, Arsad N, A review of seismic detection using fiber optic distributed acoustic sensing: from telecommunication cables to earthquake sensors, *Nat. Hazards* **121**, 13927–13959 (2025). https://doi.org/10.1007/s11069-025-07370-5.

4 Zinsou R, Liu X, Wang Y, Zhang J, Wang Y, Jin B, Recent progress in the performance enhancement of phase-sensitive OTDR vibration sensing systems, *Sensors* **19**(7), 1709 (2019). https://doi.org/10.3390/s19071709.

5 Martins HF, Martín-López S, Corredera P, Filograno ML, Frazão O, Gonzalez-Herráez M, Phase-sensitive optical time domain reflectometer assisted by first-order Raman amplification for distributed vibration sensing over >100 km, *J. Lightwave Technol.* **32**(8), 1510–1518 (2014). https://doi.org/10.1109/JLT.2014.2308354.

6 Waagaard OH, Rønnekleiv E, Haukanes A, Stabo-Eeg F, Thingbø D, Forbord S, Aasen SE, Brenne JK, "Real-time low noise distributed acoustic sensing in 171 km low loss fiber," *OSA Continuum* **4**(2), 688–701 (2021).

7 Mànuel A, Roset X, Rio JD, Toma DM, Carreras N, Panahi SS, Garcia-Benadí A, Owen, T, Cadena J, Ocean bottom seismometer: design and test of a measurement system for marine seismology, *Sensors* **12**(3), 3693–3719 (2012). https://doi.org/10.3390/s120303693.

8 Pastor-Graells J, Martins HF, Garcia-Ruiz A, Martin-Lopez S, Gonzalez-Herraez M, Single-shot distributed temperature and strain tracking using direct detection phase-sensitive OTDR with chirped pulses, *Opt. Express* **24**(12), 13121–13133 (2016). https://doi.org/10.1364/OE.24.013121.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 13
© The Author(s), published by EDP Sciences, 2026
https://doi.org/ 10.1051/jeos/2026008
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

Modeling of surface vessels using distributed acoustic sensing data and physics-based optimization

Pedro Martins[1,2,*], Ilmer Van Golde[3], Susana Silva[1,2], Orlando Frazão[1,2], and Ricardo Sousa[1,2]

[1] INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[2] Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
[3] Instituto Hidrográfico, Rua das Trinas 49, 1249-093 Lisboa, Portugal

Received 15 December 2025 / Accepted 25 January 2026

Abstract. Technological advances in global communications depend significantly on robust and efficient long-distance infrastructures. One notable example is the submarine cable network. Installed on the ocean floor, these cables use fiber optic technology to transmit large volumes of data at high speed and low latency between continents. Beyond their primary communication function, recent innovations allow these cables to serve as Distributed Acoustic Sensing (DAS) systems, effectively converting tens of kilometers of passive fiber into massive, coherent arrays of vibration sensors. The primary objective of this project is to enhance maritime surveillance capabilities by integrating DAS technology with advanced kinematic modeling. This paper establishes a rigorous physical and mathematical framework for interpreting the acoustic signatures of surface vessels detected by bottom-mounted fibers. We derive the complete opto-acoustic transfer function, formulate the hyperbolic moveout equations based on a moving point-source solution to the wave equation, and implement a stochastic inversion scheme using Differential Evolution. By optimizing a correlation-based loss function, we demonstrate the ability to recover vessel trajectory, speed, and depth from complex interferometric patterns with speed estimation errors consistently below 1%. This approach allows for the extraction of quantitative physical parameters from raw optical data, bridging the gap between photonics and hydroacoustics.

Keywords: Distributed acoustic sensing, Maritime surveillance, ϕ-OTDR, Hyperbolic moveout, Differential evolution, Optimization.

1 Introduction

Distributed Acoustic Sensing (DAS) has revolutionized the field of optical monitoring by enabling the conversion of standard fiber optic cables into large-scale, ultra-sensitive vibration sensor arrays [1, 2]. Originally developed for vertical seismic profiling and pipeline monitoring within the oil and gas industry, DAS technology is now finding critical applications in environmental monitoring and critical infrastructure security. Specifically, submarine telecommunication cables can be repurposed as distributed seismic or acoustic sensors, providing meter-scale spatial resolution and continuous coverage over tens of kilometers of the ocean floor [3, 4]. Recent studies have demonstrated that DAS is capable of detecting underwater hydroacoustic signals, including the mechanical noise generated by surface vessels, with performance metrics comparable to conventional hydrophone arrays, although the Signal-to-Noise Ratio (SNR) may vary across frequency bands [5].

The primary challenge in transforming a DAS system into a reliable vessel tracking platform lies in the complexity of the signal processing required. As a ship moves through the water, it generates continuous acoustic noise resulting from engine vibrations and propeller cavitation. Because the ship moves relative to the stationary fiber on the seabed, these acoustic wavefronts reach different sections of the fiber at different times, producing distinct, geometry-dependent patterns in the recorded data. This paper presents a comprehensive mathematical formulation of these hyperbolic moveout patterns and introduces an inverse problem framework designed to extract precise vessel movement and physical characteristics directly from raw DAS waterfall plots.

Unlike prior studies, which have primarily focused on the simple detection or qualitative visualization of vessel signatures, our approach enables the quantitative inference of vessel dynamics and physical parameters from DAS measurements. Global maritime security currently relies heavily on Automatic Identification System (AIS) transponders [6]. However, AIS data is inherently vulnerable to spoofing, can

* Corresponding author: pedro.l.martins@inesctec.pt

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

be deliberately disabled by "dark vessels," or may be unavailable in remote ocean areas due to sparse receiver coverage. Consequently, there is a critical need for passive, non-cooperative surveillance systems that do not rely on the target's participation. Submarine cables, which already crisscross the world's oceans, offer an ideal, covert platform for this purpose [7]. By connecting a DAS interrogator to dark fiber strands, real-time acoustic monitoring can be achieved over vast distances with high resolution. This work presents a rigorous physical–mathematical framework for the acoustic detection of surface vessels using bottom-mounted optical fibers, deriving the opto-acoustic transfer function and hyperbolic moveout equations from the wave equation for a moving point source, and solving the inversion using a Differential Evolution–based stochastic optimization strategy.

2 Physical principles of phase-sensitive OTDR

To interpret the acquired data, we model the Phase-sensitive Optical Time Domain Reflectometry (ϕ-OTDR) process [2]. The fiber acts as a one-dimensional waveguide with random scattering centers. As a coherent laser pulse propagates, the backscattered electric field is formed by the superposition of contributions from scatterers within the pulse width.

When an external acoustic pressure wave $P(\mathbf{r}, t)$ interacts with the cable, it induces a mechanical strain on the fiber core. In DAS, we measure the differential phase shift $\Delta\phi$ between two sections of fiber separated by a gauge length L_g. Assuming the acoustic wavelength is larger than the gauge length, the relationship between the recorded optical phase and the axial dynamic strain ϵ_{zz} is linear:

$$\Delta\phi(z,t) \approx \frac{4\pi n \xi L_g}{\lambda} \epsilon_{zz}(z,t), \tag{1}$$

where $\xi \approx 0.78$ is the photo-elastic scaling factor and n is the refractive index. This establishes that the recorded optical phase is a direct proxy for the strain induced by the vessel's acoustic emission.

3 Kinematic and acoustic wavefield modeling

We model the ocean as a semi-infinite homogeneous half-space and the vessel as a moving point source emitting a broadband signal $S(t)$.

3.1 Hyperbolic moveout and directional sensitivity

For a vessel moving with velocity v_s along a linear trajectory at the surface, the acoustic arrival time at the fiber creates a hyperbola:

$$t_{\text{arr}}^2(x) = t_{\text{CPA}}^2 + \frac{(x - x_{\text{CPA}})^2}{V_{\text{app}}^2}, \tag{2}$$

where t_{CPA} is the time at the Closest Point of Approach (CPA) and V_{app} is the moveout velocity. This relationship allows us to invert the vessel speed and impact parameter from the data curvature. Additionally, the fiber acts as a dipole sensor; its sensitivity to axial strain results in a characteristic amplitude "null" when the acoustic wavefronts arrive perpendicular to the cable (at CPA), facilitating precise localization [8].

Algorithm 1. Pseudocode: Forward model simulation

Input: Grid (x, t), Vessel Params $\mathbf{m}$
Output: Synthesized Data Matrix $D(x, t)$
Initialize $D \leftarrow 0$
for each time step t_j **do**
  Calculate vessel position $\mathbf{r}_s(t_j)$ using v_s, θ_0
  for each sensor channel x_i **do**
    Compute range $R_{ij} = \|\mathbf{r}_s(t_j) - \mathbf{r}_{fib}(x_i)\|$
    Compute Doppler factor f_{Dop} and Arrival Time t_{arr}
    Spectral Source: $S_{raw} \leftarrow \sum \sin(2\pi f_{Dop} t_{arr})$
    Interference: Apply Lloyd's Mirror modulation
    Hull Window: Apply spatial envelope $H(l)$
    Spreading: Scale by $R_{ij}^{-\gamma}$
    Accumulate strain: $D(i,j) \leftarrow \epsilon_{zz}(S_{raw}, R_{ij})$
  end for
end for
Sensor Model: Convolve D with Gauge Length L_g
Return D

3.2 Spectral interference and Doppler effects

Beyond kinematics, the signal contains spectral features derived from the source depth and motion. The interaction between the direct acoustic path and the surface reflection (Lloyd's Mirror effect) creates frequency notches dependent on the source depth z_s. Simultaneously, the broadband noise undergoes a Doppler shift due to the relative radial velocity, resulting in a sigmoidal frequency transition. These physical constraints – interference patterns and Doppler shifts – are critical for validating estimated velocities and depths [9].

3.3 Numerical implementation

To validate the model, we implemented a numerical forward model that synthesizes the DAS waterfall plot $D(x, t)$. The simulation procedure is summarized in Algorithm 1. It generates a parametric source $S(t)$ composed of a fundamental frequency and harmonics, modulated by Lloyd's Mirror interference. The source is spatially windowed by a geometric hull function to simulate the acoustic wake, scaled by geometrical spreading, and finally convolved with the gauge length operator to replicate the physical measurement process.

4 Inverse problem and optimization strategy

Inverting kinematic parameters from DAS data is a nonlinear optimization problem complicated by oscillatory

Table 1. Inverted kinematic parameters and validation against AIS (Mean ± SD over 10 independent runs).

	Case A (Mestre Horácio)		Case B (Wilson Hirtshals)	
Parameter	Value	Error (%)	Value	Error (%)
Speed (v_s) [m/s]	-3.23 ± 0.12	3.29	4.97 ± 0.03	0.61
Trajectory angle [deg]	33.88 ± 1.36	–	41.10 ± 0.80	–
Source depth (z_s) [m]	0.60 ± 0.13	0.45	5.47 ± 0.19	2.32
Dom. frequency (f_0) [Hz]	44.94 ± 7.50	–	13.43 ± 0.58	–
Vessel length (L_s) [m]	20.77 ± 1.02	3.85	85.69 ± 3.46	4.79
Vessel width (W_s) [m]	4.96 ± 0.54	9.81	14.50 ± 2.08	4.60
Spreading factor (γ)	0.71 ± 0.13	–	0.50 ± 0.00	–

interference patterns and the hyperbolic nature of the moveouts. To avoid the local minima that are inherent to gradient-based methods in such rugged landscapes, we employed Differential Evolution (DE) [10], a robust stochastic global optimization algorithm. The vessel state is defined by the parameter vector $\mathbf{m} = [v_s, \theta_0, z_s, f_0, L_s, W_{s,\gamma}]^{\mathrm{T}}$, representing the speed, angular offset, source depth, fundamental frequency, vessel length, vessel width, and spreading coefficient, respectively. The search space is bounded by physically constrained limits (e.g., $v_s \in [1, 20]$ m/s) to ensure the algorithm converges to realistic solutions.

Using a classic *DE/rand/1/bin* strategy, the algorithm evolves a population of candidate vectors over successive generations. A mutant vector $\mathbf{v}_{i,g}$ is generated for each target vector $\mathbf{m}_{i,g}$ by mixing three random population members:

$$\mathbf{v}_{i,g} = \mathbf{m}_{r1,g} + F \cdot \left(\mathbf{m}_{r2,g} - \mathbf{m}_{r3,g}\right), \tag{3}$$

where F is the differential scaling factor. Binomial crossover is then performed to create a trial vector, which replaces the target vector only if it yields an improvement in the objective function.

To address the lack of absolute amplitude calibration in DAS data (which can vary due to fiber coupling and signal fading), we chose to optimize for structural similarity rather than absolute intensity. We minimize a correlation-based loss function $J(\mathbf{m})$:

$$J(\mathbf{m}) = 1 - \frac{\mathrm{Cov}(S(\mathbf{m}), R)}{\sigma_S \sigma_R}, \tag{4}$$

where $S(\mathbf{m})$ is the simulated field generated by the candidate parameters, R is the observed experimental data, and σ denotes the standard deviation. This scale-invariant metric ensures the model locks onto geometric features – such as the hyperbolic curvature and interference nulls – independent of the overall signal magnitude.

5 Material and methods

5.1 Data acquisition

The experimental data was acquired using a standard commercial DAS interrogator unit connected to a bottom-mounted submarine telecommunications cable. The interrogator's pulse repetition rate yielded a temporal sampling frequency f_s sufficient to capture the acoustic spectrum of vessel noise, while the optical gauge length L_g provided a spatial sampling resolution of approximately 9.8 m.

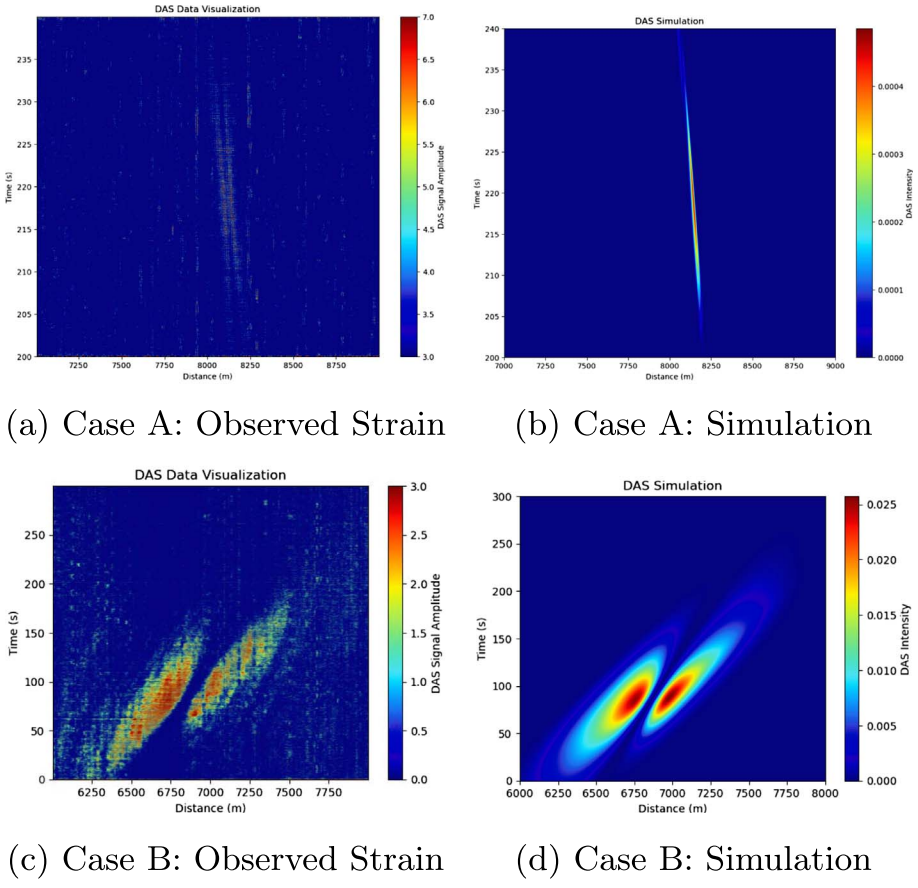
5.2 Signal preprocessing pipeline

Before the optimization process, the raw optical phase data undergoes a rigorous, multi-stage preprocessing pipeline designed to maximize the Signal-to-Noise Ratio (SNR) and ensure statistical consistency with the forward model. Raw DAS data is typically dominated by high-amplitude, low-frequency environmental noise, such as ocean swell and hydrostatic pressure variations (<1 Hz). To isolate the specific mechanical signature of the vessel, we applied an 8th-order Butterworth bandpass filter with a passband of [2, 120] Hz. This range was selected to capture the fundamental engine harmonics and propeller blade pass frequencies while effectively rejecting microseisms and high-frequency system noise.

Following filtration, we addressed the computational constraints imposed by the high sampling rates of DAS systems. We performed temporal decimation immediately following the anti-aliasing filter to reduce the computational load of the fitness evaluation function. For datasets with high SNR (Case B), the raw coherent strain was used directly. However, for lower SNR scenarios (Case A), we computed the Hilbert envelope of the signal. This transformation focuses the optimizer on the energy distribution of the acoustic arrival rather than the rapidly oscillating phase carrier, significantly improving convergence stability in noise-limited regimes. Finally, because the correlation-based loss function requires that the simulated and experimental data share comparable statistical properties, we applied a Z-score standardization to the preprocessed data matrix D, subtracting the global mean μ_D and dividing by the standard deviation σ_D. To prevent transient artifacts (e.g., system glitches or dropped channels) from skewing the correlation metric, the data was robustly clipped to the 99th percentile. This ensures the optimization surface remains smooth and gradients are driven by the persistent vessel wake geometry.

6 Results

The efficacy of the proposed model was assessed via wavefield reconstruction and validation against independent AIS



(a) Case A: Observed Strain (b) Case A: Simulation

(c) Case B: Observed Strain (d) Case B: Simulation

Figure 1. Comparative validation of the forward model against experimental DAS data. Top row (a, b): Case A (Mestre Horácio), showing strong, coherent hyperbolic arrivals. Bottom row (c, d): Case B (Wilson Hirtshals), where the simulation accurately reconstructs the fainter interference lobules (Lloyd's Mirror effect), providing visual confirmation of the source depth estimation.

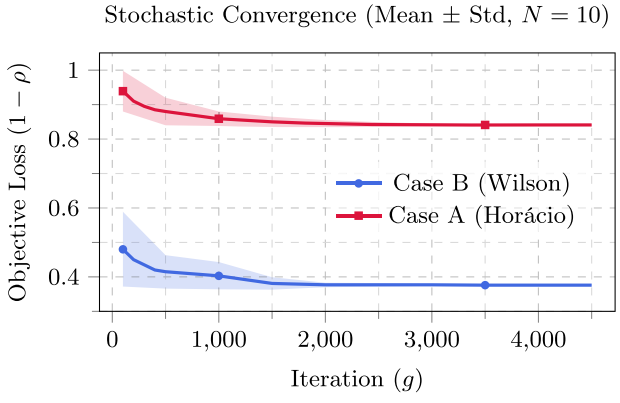


Figure 2. Stochastic convergence of the Differential Evolution algorithm ($N = 10$ runs). The solid lines represent the mean best loss $(1 - \rho)$ as a function of generation, and the shaded areas indicate ± 1 standard deviation. Case B (blue) converges to a significantly lower residual (≈ 0.38) with high stability, while Case A (red) stabilizes at a higher noise floor (≈ 0.84).

ground-truth data (Table 1). Figure 1 demonstrates the high fidelity of the model in reproducing the signal topology for both test cases (Fig. 2).

7 Conclusion

This work has successfully established a comprehensive physics-based framework for vessel tracking using Distributed Acoustic Sensing (DAS) combined with Differential Evolution optimization. We have demonstrated the ability to recover critical kinematic parameters (speed, trajectory, depth) and vessel dimensions from raw interferometric data, confirming the viability of repurposing existing single-mode submarine cables as massive, distributed hydrophone arrays. The proposed optimization scheme proved to be robust against varying levels of environmental noise, accurately resolving complex hyperbolic moveout features and interference patterns. Specifically, we achieved velocity estimation errors below 1% in high-SNR conditions. Future work will focus on integrating these kinematic priors into deep learning architectures, aiming to develop fully automated, real-time maritime surveillance systems.

Funding

This work is supported by Component 5 – Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU),

framed in the Next Generation EU, for the period 2021 – 2026, within project OBSERVA, with reference 2024.07610.IACDC (https://doi.org/10.54499/2024.07610.IACDC). This work is co-funded by the Portuguese Fundação para a Ciência e Tecnologia, FCT, I.P./MCTES through national funds (PIDDAC): UID/50019/2025 and LA/P/0068/2020, by the MODAS project 2022.02359.PTDC, and by EC project SUBMERSE project HORIZON-INFRA-2022-TECH-01-101095055.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability statement

Data is not available for this article.

Author contribution statement

Pedro Martins: Conceptualization, Methodology, Software, Investigation, Writing – original draft.
Ilmer van Golde: Writing – review & editing.
Susana Silva: Writing – review & editing.
Orlando Frazão: Writing – review & editing.
Ricardo Sousa: Conceptualization, Resources, review, Supervision.

References

1 Masoudi A, Newson TP, Distributed optical fibre dynamic strain sensing, *Opt. Lett.* **24**(10), 10941–10954 (2016).
2 Hartog AH, *An Introduction to Distributed Optical Fibre Sensors* (CRC Press, Boca Raton, 2017).
3 Zhao Y, Yin B, Sang G, et al. 2025. Applications of optical fiber sensors in marine observation: a review. *Intell. Mar. Technol. Syst.* **3**(1), 38 (2025).
4 Zhong Z, Liu K, Han X, Lin J, Review of fiber-optic distributed acoustic sensing technology, *Instrumentation* **6**(4), 1–14 (2025).
5 Markom M, Saharudin S, Hisham MH, Systematic review of fiber-optic distributed acoustic sensing: advancements, applications, and challenges, *Optical Fiber Technol.* **94**, 104293 (2025).
6 Guo Z, et al. Anomaly detection for AIS trajectories based on kinematic interpolation, *Ocean Eng.* **236**, 109432 (2021).
7 Lindsey NJ, Dawe TC, Ajo-Franklin JB, Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing, *Science* **366**(6469), 1103–1107 (2019).
8 Yilmaz Ö, *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*, (Society of Exploration Geophysicists, 2001).
9 Urick RJ, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983).
10 Storn R, Price K, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *J. Global Optim.* **11**, 341–359 (1997).

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 40
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026019
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

RESEARCH ARTICLE OPEN ACCESS

Eye-safe non-line-of-sight localization using compact nanosecond laser diodes and single-photon-avalanche-diode arrays

Konstantin Albert[1,2,*], Julian Klein[1], Manuel Ligges[1], and Anton Grabmaier[1,2]

[1] Fraunhofer Institute for Microelectronic Circuits and Systems, 47057 Duisburg, Germany
[2] University of Duisburg-Essen, 47057 Duisburg, Germany

Received 26 November 2025 / Accepted 24 February 2026

Abstract. Active optical non-line-of-sight (NLOS) imaging faces severe photon loss and stringent timing demand, often motivating bulky, non-eye-safe systems with femto- or picosecond lasers and spatial scanning schemes. We present an eye-safe, compact NLOS localization approach that uses a single-photon-avalanche-diode array with on-chip timing and inexpensive and compact nanosecond pulsed laser diodes. To circumvent first-photon induced saturation in soft-gated detectors, the scene is illuminated from two positions placed outside the detector's field of view. The detector observes the relay wall in parallel, yielding a photon-efficient, non-confocal measurement. Transient simulation of multiple scattered laser pulses, that returned to the relay wall after an object interaction, are used with back projection reconstruction to assess illumination schemes. It is shown that superimposing reconstructions from two non-central illumination positions reduces the pulse-width-induced uncertainty of the determined target position, which is validated by measurements. Further compensation of the extended pulse width is analyzed with temporal and spatial filters. Matched filtering in the temporal domain outperforms spatial edge detection in the reconstruction volume, so that localization resolution becomes primarily limited by the detector's temporal resolution and measurement geometries rather than temporal pulse width. We outline a hybrid LiDAR-NLOS system for direct relay-wall calibration, pointing toward practical, eye-safe NLOS localization, with compact solid-state hardware.

Keywords: Non-line-of-sight, Eye-safe, Localization, Single-photon-avalanche-diode.

1 Introduction

Most environmental sensing systems, i.e. radar, light detection and ranging (LiDAR), ultrasonic sensors or conventional cameras are designed to observe a scene in direct line of sight. The measured area or volume of interest varies substantially, spanning from single-point intensity measurements to spatially resolved, three-dimensional information. When the observing systems cannot be positioned to view the scene of interest directly, systems based on non-line-of-sight (NLOS) approaches might be used. Potential applications for such approaches include collision avoidance in automotive and robotic contexts, locating people behind doors or rubble for public safety and rescue, and industrial inspection of occluded defects. Although NLOS systems can employ various remote sensing modalities [1–4], time-correlated photon-based approaches have been the main focus of research to date [5–10]. They exploit diffuse reflections from everyday surfaces and afford discrete illumination positions and high temporal resolution, which are crucial for high-resolution reconstruction of hidden scenes [9].

Active optical NLOS imaging refers to measuring multiple scattered laser pulses to extract information about scene elements that are not directly visible to the observer. The system typically comprises a collimated pulsed laser and a detector with high temporal resolution, which together illuminate and observe multiple positions on a relay wall positioned in the direct line of sight. The pulse scatters off the relay wall into the NLOS scene, where it undergoes secondary scattering from objects hidden in the scene and the rebounds back to the relay wall, where it can be observed directly by an optical detector (see Fig. 1a). The resulting multiply scattered transients on the relay wall are measured with spatial and temporal resolution. Such optical NLOS systems are differentiated by the arrangement of illumination and detection points. Confocal arrangements [8, 10–12] scan the relay wall and measure transients only at the current illumination point, ignoring most returning reflections and thus being inherently photon-inefficient. In contrast, non-confocal arrangements, with arbitrary numbers and positions of illumination and

* Corresponding author: konstantin.albert@stud.uni-due.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

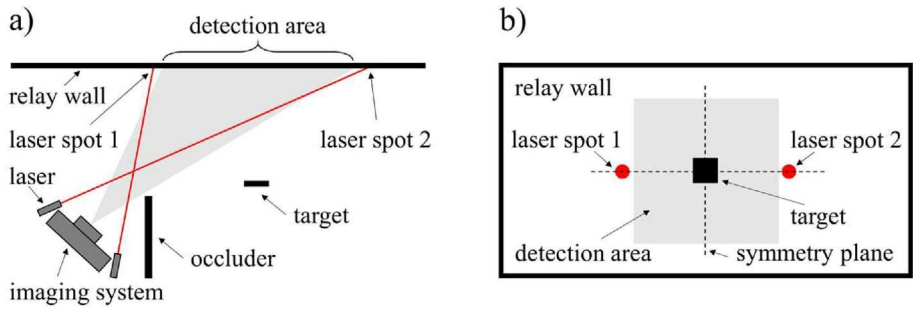


Figure 1. a) Top-view schematic of the NLOS setup. b) Front-view of the relay wall. The black square represents the projected target position.

detection points, can be designed in a more photon-efficient manner.

Both confocal and non-confocal arrangements share fundamental hardware components comparable to those of direct time-of-flight (ToF) LiDAR systems used for three-dimensional imaging in the direct field of view (FoV). However, in NLOS systems the laser pulse paths are more complex than those encountered when measuring distances in direct line of sight. First, at least three reflections before detection drastically reduce the receivable photon flux compared with single reflection in direct ToF LiDAR. Second, the achievable lateral resolution of reconstructed NLOS scenes cannot be controlled solely via detector pixel resolution or the number of illumination points, as in LiDAR. Instead, not the number of pixels but the temporal resolution combined with the spatial extend of the illumination and detection scheme on the relay wall define the achievable depth and lateral resolution. These requirements place heavy demands on the measurement setup. Existing lab-grade systems are often expensive and far from compact, relying on pico- and femtosecond laser sources, single-photon-sensitive detectors, and complex time-correlated singe-photon counting electronics. Moreover, free-space propagation loss scales with the inverse fourth power of the object distance to the relay wall [7], compared with the inverse-square loss in LiDAR. This is often compensated by higher laser power, which comprises eye-safety and limits applicability to laboratory environments.

In this work, an eye-safe NLOS localization system is presented that addresses the constraint of restricted laser emission through a high-efficiency photon-detection design and achieves inexpensive, compact implementation using nanosecond-pulsed laser diodes. Specifically, a two-dimensional single-photon-avalanche-diode (SPAD) array with integrated timing capabilities and camera optics is developed and used to observe, in parallel, the detection region on the relay wall. A single illumination point combined with the dense grid of detection points meets the requirements for a NLOS measurement. Scan-free NLOS imaging [13, 14] has been reported previously with this single illumination system design using picosecond pulsed laser sources. Extending this prior work, we examine the performance gains for localization using additional illumination position. Simulated time-of-arrival photon statistics, validated by measurements with the implemented system, are used for the investigation. Despite the longer pulse width compared to comparable NLOS systems, three-dimensional localization of a target in a NLOS scenario is achieved at a resolution determined solely by detector parameters.

2 Material and methods

2.1 Measurement setup

The measurement setup comprises two pulsed laser diodes and a time-resolved multi-pixel detector [15] with camera optics that image the relay wall. Each diode operates at the accessible emission limit specified for eye-safe operation, meaning that neither the repetition rate nor the pulse energy can be increased without reducing the other in order to increase the system's performance without losing civilian applicability. Nevertheless, the spatial separation and individual emission angles allow the use of more than one diode in the system, as only one laser can be imaged onto a hypothetical retina at a time. Therefore, spatially separated sources are used to comply with laser safety standards, even though the total emission into the scene is multiplied. The dual illumination scheme corresponds to the simplest case of multiple illumination and provides two symmetry planes for the following analysis. Although more sources could be implemented, illuminating arbitrary positions using a mirror system with a single laser source for instance, circumvents the exploitation of multiplied total emission. Depending on the reflecting angles of the mirror, a nearby retina could receive the laser power from multiple emission angles. Moreover, the laser diodes used are cost-efficient and solid-state-based compared to galvo mirror designs.

The arrangement is a common NLOS imaging scenario with a flat relay wall, an occluder, an observing system, and an NLOS scene containing a single target placed in 1.5 m distance to the relay wall (Fig. 1a).

The SPAD-array detector [15] has 24 × 32 pixels with individual time-to-digital converters (TDCs) that generate ToF histograms under a first-photon detection scheme. With each laser pulse, each pixel yields at most one timestamp. This time of arrival corresponds to the first induced avalanche in the SPAD within one measurement cycle of $1/f_{\mathrm{rep}} = 80\ \mu\mathrm{s}$ duration, where f_{rep} denotes the laser pulse repetition frequency of 12.5 kHz. Because signal photons, ambient light, and dark counts can all trigger these avalanches, histogramming of the time stamps is required to

gather sufficient statistical information and recover the required transients for NLOS reconstruction. The pixel-specific bin widths of these histograms are approximately 400 ps, as given by the integrated TDC circuits. The detector has a microlens-enhanced fill factor of 26.8% and a quantum efficiency of 2% at 905 nm [16]. The FoV covers a 2 m × 2 m. rectangular region on the relay wall. The soft-gated detector precludes direct illumination within the FoV because the direct reflection of the laser pulse would saturate the detector. To mitigate the resulting illumination inhomogeneity, two illumination points are placed at opposite edges of the FoV, as shown in Figure 1b.

The laser sources are pulsed laser diodes operating at 905 nm, this mitigating eye-safety conditions and optical pollution due to their non-visibility to the human eye. They generate pulses with 2.4 ns temporal width (full width at half maximum) and a nearly top-hat temporal beam profile. Each diode generates 30 W peak optical, resulting in an average power below 1 mW at the repetition rate of 12.5 kHz. Their pulse-on-demand capabilities are used to emit laser pulses synchronized with the detector. Because each pulse emission can only generate one timestamp per pixel, the measurement statistics scale with the number of emitted pulses. Given the fixed repetition rate, the achievable frame rate is inversely proportional to the number of pulses per acquisition. The alternating emission of the laser sources is necessary to assign the measured timestamps to an illumination position, which is essential for valid reconstruction. With a doubled detector frame rate, the two alternating 12.5 kHz emissions can be captured, resulting in the same measurement frame rate for the single and dual illumination scheme.

2.2 Simulation and calibration

For a controlled, calibration-independent assessment of system performance, simulated measurements are used in addition to proof-of-concept measurements with the physical setup. However, certain experimental conditions, such as specific pulse lengths and a central illumination scheme, cannot be achieved with the physical setup that has been developed. Spatially and temporally resolved transients of scattering light at the relay wall are therefore simulated while incorporating first-photon detector behavior. Because a photon can be detected only if the SPAD has not already fired, the first-detection probability decays exponentially with time at a rate set by the background light level r_B. Impinging photon rates r_L of the three-bounce laser pulses are calculated using view factors, which yield the fractions of Lambertian scattering into a specified solid angle. These rates and their corresponding ToF define the photon detection probability density function (PDF) of each detector pixel:

$$\mathrm{PDF}(t) = \begin{cases} r_B \cdot e^{-r_B \cdot t}, & 0 < t < t_{ToF} \\ (r_B + r_L) \cdot e^{-r_B \cdot t} \cdot e^{-r_L \cdot (t - t_{ToF})}, & t_{ToF} \le t < t_{ToF} + t_P \\ r_B \cdot e^{-r_B \cdot t} \cdot e^{-r_L \cdot t_P}, & t_{ToF} + t_P \le t \end{cases} \tag{1}$$

Here, t_P is the temporal width of a rectangular laser pulse and t_{ToF} is the time of flight determined by the distances between the illumination position, target position and the imaged pixel position on the relay wall. Synthetic histograms are generated using the inverse transfer method by sampling arrival times from the PDF (1). Subtracting the exponentially decaying background component induced by a constant background rate r_B mitigates intensity gradients in the reconstruction.

The simulation assumes only Lambertian surfaces. Specular and retroreflective targets would naturally lead to alternative distributions of the receivable light, as described in Section 3.1. Nevertheless, the extensive area of detection increases the probability of receiving specular components of the target reflections, especially given the changing angle of incidence of the target illumination due to the dual illumination scheme. Retroreflective targets, which are best detected by confocal measurements, can also be detected, because the illumination positions are placed close to the edge of the detection area.

While the simulation covers the entire light transfer function from laser emission to detection via triple reflection, the original data of the measurement setup must be further calibrated in order to extract the relevant parts of the transient histograms. The line-of-sight (LOS) propagation of the laser pulse from the system to the relay wall and the LOS observation of the relay wall must be subtracted to obtain the light transmission relevant for NLOS reconstruction. Theoretically, the propagation times between the detector and the detection point on the relay wall could be extracted from a LiDAR measurement of the relay wall. To rule out errors from such additional measurements, the spatial alignment between the system and the relay wall is measured manually. For this purpose, the intensity-based photon-counting mode of the system is used to map the relay wall equipped with retroreflector landmarks. The virtual start time of the pulse propagation originating from the relay wall is calibrated using the high-intensity peak of the primary reflection of the laser pulse on the relay wall within the measured histograms. The pixel-dependent propagation times between the detection point on the relay wall and actual detection at the system are compensated using the geometrical model of the measurement setup, yielding the required light transfer function between relay wall and target for the reconstruction procedure. Pixel-dependent temporal bin widths and periodic bin width variations are compensated prior to reconstruction using individual TDC calibration data provided by the system. This interpolates all measured histograms to a uniform temporal grid with 400 ps steps.

Multipath clutter of light paths with more than three bounces are ignored, as signal intensities decrease drastically with each diffusive reflection. Nevertheless, vertical walls adjacent to the relay wall have a significant impact on the transient data, requiring subtraction of an otherwise empty scene to extract the target positions as described below. The simulations therefore omit additional walls, but the physical measurement includes a parallel surface (the floor) at a distance of 3 m from the relay wall. Its additional NLOS contribution within the transient data is

subtracted but requires increased measurement statistics to detect targets near this interfering wall, as the signal strength of such large surfaces exceeds the target signal.

2.3 Reconstruction and filters

For the reconstruction of the NLOS scene based on the transient data, that corresponds to individual ToF per target position and detection position, several methods have been reported. Individual requirements on illumination and detection positions, being uniformly distributed [17] or in confocal arrangements [8, 18], prevents their applicability to the non-confocal system under study without error inducing transformation of the measurement data [19]. In contrast, the filtered back-projection (FBP) [20–23] used here is an ellipsoid-based reconstruction method that can handle arbitrary measurement schemes and gives insights into pulse width induced errors directly. The formation of the transients used for the reconstruction is described by the three-bounce path of the laser pulse. The time-of-flight (ToF) of the laser pulse from its scatter on the relay wall to its return to the wall after target reflection is measured (Fig. 2). This ToF is proportional to the sum of the two paths a and b indicated in Figure 2a. Possible target positions that could have produced the measured ToF are all located on an ellipsoid, or on an ellipse in two dimensions, that satisfies $a + b = a' + b'$. The foci of these ellipsoids are the illumination position and the corresponding detection point on the relay wall. By introducing multiple detection points or illumination points, the intersection of the superimposed ellipsoids indicates the target position (Fig. 2b).

For a resolution estimation of the non-confocal setup under study, the signal contribution of finite pulse widths within back projection reconstruction is examined. In Figure 2c, three relevant ellipses are shown with an effective width of approximately half the pulse width times the speed of light. The target is centered at p_T. The central ellipse II describes the signal contribution of a laser position and the detection point at the opposite border of the detection area, for example illumination position 1 and detection point N. Because this arrangement is symmetric, both laser position yield the same ellipsoid. The ellipse I results from laser position 1 and its nearest detection point 1 and is approximated by a circle around the laser position. The possible ellipses of all other pairs of laser position 1 and the remaining detection points are not shown but they all contribute to the regime hatched in grey, which is defined by the overlap of these two specific ellipsoids. The circle III belonging to illumination position 2 defines the dark grey region where all ellipsoids of the laser position 1 and 2 contribute. The lateral width l_{single} (5) is halved (6) by restricting the feasible target area from illumination position 1 through superposition with the reconstruction from illumination position 2. This is observed by the lateral distance of the intersections at points p_1 (2) and p_2 (3) for a target at position p_T compared with the boundary p_3 (4) of the more confined area (dark grey, Fig. 2c).

$$p_{1,\text{lateral}} = \frac{t_p \cdot t_{ToF}}{4 \cdot w} \cdot c^2 + \frac{{t_p}^2}{4 \cdot w} \cdot c^2, \tag{2}$$

$$p_{2,\text{lateral}} = -\frac{t_p \cdot t_{ToF}}{4 \cdot w} \cdot c^2, \tag{3}$$

$$p_{3,\text{lateral}} = \frac{t_p \cdot t_{ToF}}{8 \cdot w} \cdot c^2 + \frac{{t_p}^2}{16 \cdot w} \cdot c^2, \tag{4}$$

$$l_{\text{single}} = p_{1,\text{lateral}} - p_{2,\text{lateral}} = \frac{t_p \cdot t_{ToF}}{2 \cdot w} \cdot c^2 + \frac{{t_p}^2}{4 \cdot w} \cdot c^2, \tag{5}$$

$$l_{\text{double}} = 2 \cdot p_{3,\text{lateral}} = \frac{t_p \cdot t_{ToF}}{4 \cdot w} \cdot c^2 + \frac{{t_p}^2}{8 \cdot w} \cdot c^2 = \frac{1}{2} \cdot l_{\text{single}}. \tag{6}$$

Here, c denotes the speed of light. These expressions follow from a geometrical model with ellipse foci at $\pm w$, which correspond to the illumination offset to the scene center. The semi-major axis α of the ellipse corresponding to II in Figure 2c and the radius r of the intersecting circle (7) centered at the focus point are defined by the target ToF for the rising edges (9) with additional dependence on the pulse width for the falling edge (10). The lateral intersection position of two circles at p_3 (4) with a spacing of $2w$ are given in (8) used for calculation of p_3.

$$p_{\text{ellipse, circle}} = \pm \frac{\alpha \cdot (\alpha - r)}{w}, \tag{7}$$

$$p_{\text{circle, circle}} = \pm \frac{{r_1}^2 - {r_2}^2}{4 \cdot w}, \tag{8}$$

$$d_{\text{rising}} = \frac{t_{ToF}}{2} \cdot c, \tag{9}$$

$$d_{\text{falling}} = \frac{t_{ToF} + t_P}{2} \cdot c. \tag{10}$$

According to the geometric model, the lateral spread of feasible target positions is expected to be halved along the axis defined by the illumination points. The resulting reduction in three-dimensional localization error due to this superposition is examined in the Results section.

The filtering applied to the reconstruction result and the underlying temporal domain data discussed is not intended to enable general imaging capabilities but to compensate the effects caused by the finite temporal laser pulse width, which is several times larger than the detector's temporal resolution. The extracted target position from the reconstruction result corresponds to the maximum-valued voxel in the reconstruction volume. To use this position extraction, only a single target may be simulated or measured at a time to avoid multiple local maxima. Moreover, partial occlusion by additional targets and phantom targets resulting from signal overlap of individual targets would complicate the extraction of target positions. Partial occlusion of the relay wall by the target is also omitted, so that all points of detection are accessible to the system and the reconstruction scenario discussed above remains valid. A *Laplacian-of-*

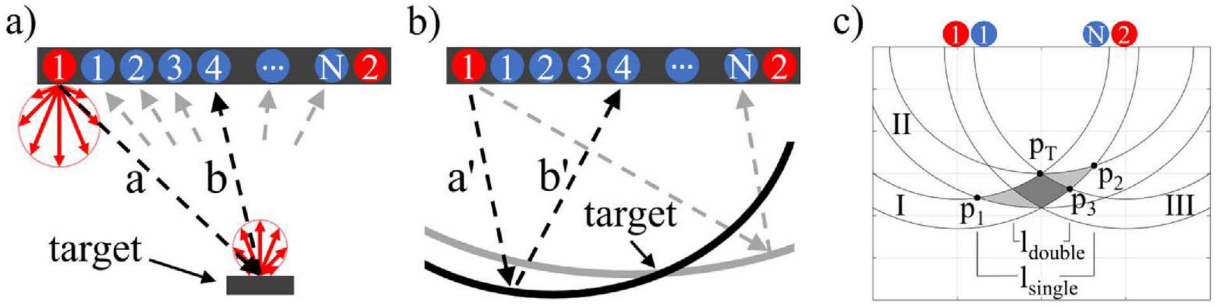


Figure 2. a) NLOS measurement process with two Lambertian scatterings indicated with red arrows. Dashed arrows indicate pulse propagation. b) Resulting ellipses of possible target reflections that could have produced the measured data of detection positions 4 and N (blue circles) from illumination position 1 (red circle). The intersection indicates the actual target position. c) Overlapping ellipses for finite pulse width of a centrally positioned target. Ellipses I and II originate from laser position 1 (red circle) and detection position 1 and N respectively. Ellipse III originates from laser position 2 and detection position N, confining the light grey area to the dark grey area.

Gaussian (LoG) filter which is typical for FBP [21, 24] can optionally be applied to the reconstruction, with the *Gaussian* width matched to the voxel size to detect the rising edge of the broad overlapping ellipsoids. The temporal data can optionally be filtered before reconstruction with a *Gaussian* filter with a sigma matched to the pulse width. A *matched* filter with an appropriate temporal shift can also be applied to concentrate the received pulse form on the rising edge of the original signal to mimic a discrete ToF besides the finite pulse width.

3 Results and discussion

3.1 Illumination schemes

A central illumination position within the detection area on the relay wall may seem obvious in terms of scene illumination but strongly reduces the sensitivity for multiply scattered NLOS signals due to detector saturation by direct scattering in the employed first photon detection scheme. Hence, illumination positions outside the detection area are examined. Possible illumination positions are placed on an exemplary axis across the center of the detection area as shown in Figure 1b along which a target will be placed with an offset to the relay wall. A plane of interest is defined that is perpendicular to the relay wall and intersects with the aforementioned illumination axis. Within this plane, the distribution of receivable signal is computed using view factors that describe the radiation from a surface into a specific solid angle. The receivable signal quantifies the photon flux that reaches the detection area on the relay wall given by the photons that impinge on the target. The latter depends on both the illumination and the target position. Since the receivable signal from a target is directly linked to the signal-to-noise ratio (SNR) of a NLOS measurement it is a figure of merit for achievable reconstruction performance.

For a non-central illumination located at the left edge of the detection area the distribution of receivable signal is highly inhomogeneous across the scene as shown in Figure 3a. An additional illumination position shown in Figure 3b is used to compensate for the disadvantaged right side of the scene. At the cost of an additional laser source and alternating measurements of each illumination position with doubled detection rate, the photon flux into the NLOS scene can be doubled while complying with eye safety restrictions. All illumination schemes (Figs. 3a–3c) show that the receivable signal is maximized for target positions in front of the illumination positions. For larger target distances from the relay wall, the receivable signal decreases due to free-space propagation losses. Non-central targets benefit from non-central illumination, as indicated by higher values at lateral target positions for the dual illumination scheme compared to the central scheme where the receivable light decreases visibly towards the side edges (cf. Figs. 3b, 3c). Conversely, the advantage of central illumination (Fig. 3c) for central target positions is confined to the near field of the relay wall, mostly covered by the direct line-of-sight and thus of little relevance for the NLOS discussion. Fig. 3d) shows the regions where central illumination outperforms dual illumination, presenting the ratio of receivable light. At all positions colored white, the receivable light under dual illumination equals or surpasses that of the central scheme with single illumination. Consequently, non-central illumination can be used to illuminate the NLOS scene with minimal loss of receivable light, preventing detector saturation due to directly scattered laser light.

While the receivable signal is only an indicator for the relative localization performance across the NLOS scene, simulated transients are used to further assess the non-confocal reconstruction method. This provides a qualitative insight into the reconstruction performance under the dual-illumination scheme. A NLOS scene with a single target at a distance of 1.5 m from the relay wall is considered. The target is either centered within the detection area, and thus symmetric with respect to both illumination positions, or laterally displaced toward illumination position 2, as depicted in Figure 4a. The aforementioned plane of interest of the three-dimensional back-projection result is shown in Figures 4b and 4c for a centered target and each

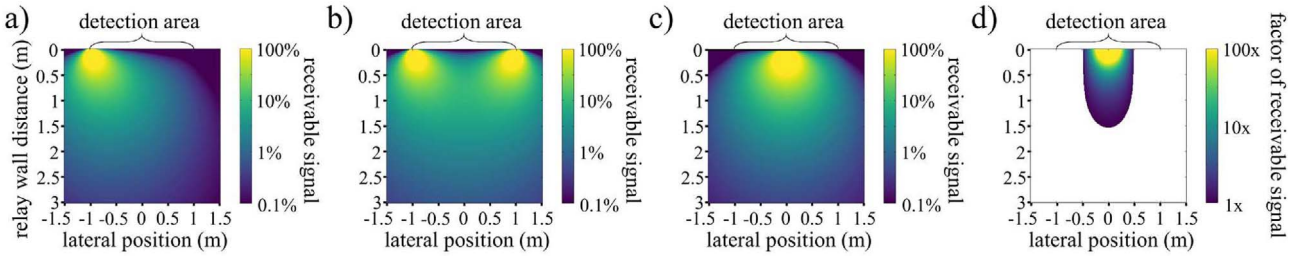


Figure 3. Receivable signal on the detection area per target position a) for one illumination position to the left of the detection area, b) for two illumination positions around the detection area, c) for a central illumination. d) Target positions where the central illumination yields more signal than the dual illumination scheme.

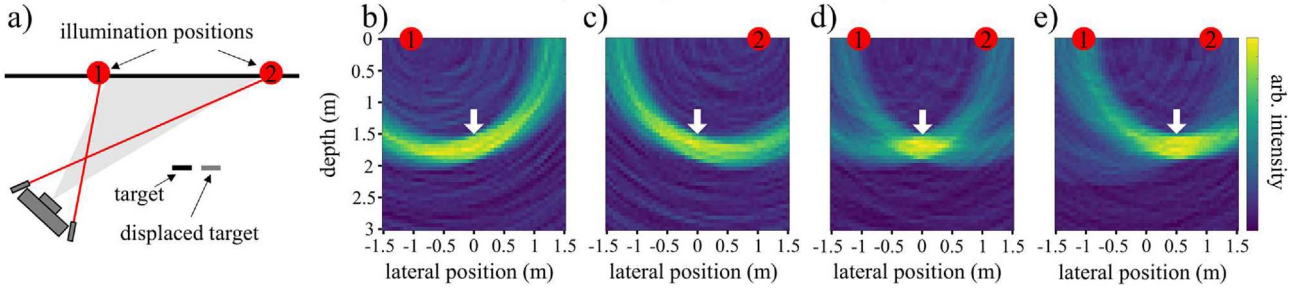


Figure 4. a) Top-view schematic of the NLOS setup with indicated target and illumination positions. b–d) 2D slices of the simulated unfiltered ellipsoid based reconstruction volume for laser position 1 and 2 and for the superposition of both, e) for a target off the symmetry axis by 0.5 m to the right. White arrows indicate the target position.

illumination position. In Figure 4b, with the relay wall at the top and a single laser positions, the overlapping ellipsoids produce a spatially extended high-intensity region that indicates the target position. Because of the finite pulse width, this intersection of ellipsoids does not reduce to a single point. Moreover, due to limited statistics and added noise in the simulated transients, the overlap is even more extended than predicted by the geometric approximation introduced earlier. The alternative illumination position in Figure 4c yields the same result, except for a horizontal mirror due to the shift of the fixed focus point to the new illumination position. The sum of these reconstructions (Fig. 4d) produces a more confined, higher-contrast region at the target position. For a single illumination position, all ellipsoids share one focus and are therefore similar, whereas the tangents of the ellipsoids at the target position differ more strongly for individual illumination positions. For targets off the symmetry axis, this geometric confinement by superposed reconstructions persists, but the resulting high-intensity region is less pronounced (Fig. 4e). This reduced confinement is attributed to unequal ellipsoid intensities in each reconstruction caused by different distances to the illumination position. Under the same simulated-measurement parameters of the transient simulation, but for a lateral target displacement of 0.5 m toward illumination position 2, the ellipsoids of illumination point 1 are barely visible. The visualized reconstruction results indicate, that the dual-illumination provides geometric benefits beyond more uniform scene illumination.

A root-mean-squared-error (RMSE) analysis is performed to quantitively assess localization errors and the confinement gain by superimposed reconstructions from individual illumination positions. The confined high-intensity region corresponds to feasible target positions, which are extracted from the unfiltered reconstruction result by maximum detection. Despite the two-dimensional depiction in Figure 4, the simulation covers a three-dimensional volume, so that the RMSE refers to the three-dimensional offset between the extracted target position and the ground truth.

For a central target at a distance of 1.5 m from the relay wall, the localization error decreases with an increasing number of measurements per laser position, as expected (Fig. 5a). A localization error of more than one meter for the lowest count of 125 simulated measurements per reconstruction is not considered a successful localization. Nevertheless, the residual RMSE for the largest statistic considered reflects the uncertainty in the feasible target area induced by the pulse width and is therefore consistent with a valid localization within the constraints given by the assumed system parameters. As indicated in Figure 4, superposition of the two reconstructions reduces the localization error compared with using only one illumination position. This enhancement cannot be explained simply by the increased number of receivable signal photons by the additional measurement, which would only follow the general trend of decreased errors with increased statistics. This is evident because the superposed result at 1,250 measurements outperforms single illumination results that use 10× more measurements, see Figure 5a. The central illumination scheme yields the maximum receivable signal among single illumination positions, yet it is still outperformed by

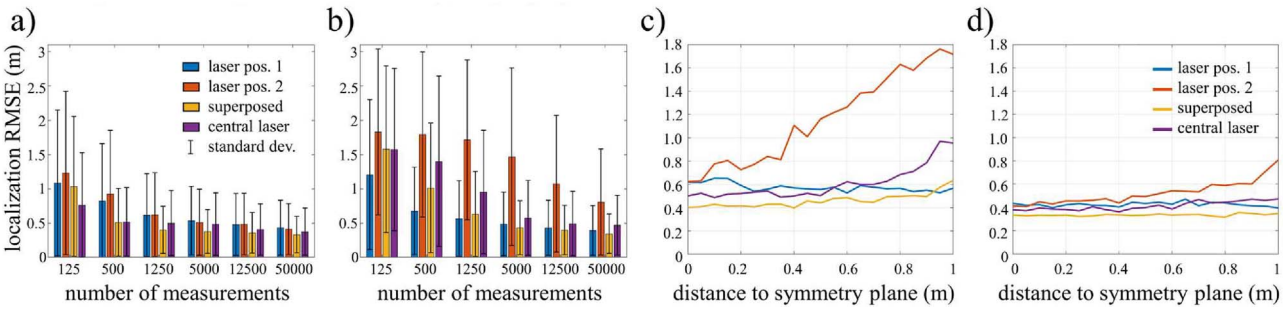


Figure 5. a) Simulated illumination dependent localization errors per number of measurements on the symmetry plane, b) near laser position 1. c) Localization errors for lateral target displacement at 1,250 measurements, and d) at 50,000 measurements.

the superposed reconstruction for practical measurement counts.

For other target positions the benefit of superimposed reconstructions varies and can even degrade localization performance. For example, a target at the same relay wall distance but outside the uniformly illuminated center is best localized using only the nearest laser position when the number of measurements is small, see Figure 5b. This lack of improvement from superposition is caused by the imbalance in signal to noise ratio (SNR) between the two reconstructions, as visualized in Figure 4e. For large numbers of measurements and resulting higher SNR the confined target region is again reduced geometrically by superposition, compare Figure 2c. The lower localization error for the nearest illumination compared with central illumination is due to the higher receivable signal caused by minimized free-space propagation losses.

Figure 5c shows how the localization error evolves with lateral target displacement from the centered position and reveals the performance difference between central illumination and off-center illumination. A centered target benefits from central illumination, while a target displacement reduces the receivable signal in two ways. First, a longer free space path and the angular dependence of Lambertian illumination by the laser spot reduces the signal reaching the target. Second, the view factor from the target to the detection area decreases at non-central positions. For a centered target, illumination by a non-central laser position is comparable to central illumination except for the slight reduction in signal reaching the target. With lateral target displacement, this loss decreases for the approached laser position and increases for the other position. The increased signal at the target for the closer laser position counteracts the reduced view factor towards the detection area, which results in a more stable localization error across the scene. This is indicated by the nearly constant RMSE for illumination position 1 (blue curve, Fig. 5c). Note that although the distant laser position 2 does not provide reliable localization, superposition is nevertheless advantageous for limited displacements (yellow curve, Fig. 5c). The general geometric confinement benefit of superposition is valid for high measurement counts as illustrated in Figure 5d since it resembles the model of Figure 2c. However, the reduction of the localization error for low statistics is more important regarding the overall systems performance.

Summarizing, the multi-spot illumination scheme improves the localization system not only through spatially more balanced scene illumination but also through geometrical confinement of the feasible target region, even when individual illumination positions show limited standalone performance. While the confinement gain was analyzed with two illumination positions for an exemplary axis, the use of four positions at the middle of each edge of the rectangular field of view naturally leads to confinement along two orthogonal axes. Further confinement could be achieved with more illumination points, although a finite number of illuminations is desirable due to limited detector frame rates and the requirement that only one illumination position can be used in every measurement cycle.

3.2 Measurement validation

The localization benefit of superposed reconstructions is validated with measurement data shown in Figure 6. The designed measurement system is compact and eye-safe. It uses a SPAD-array detector with integrated timing and two laser diodes with nanosecond pulses as sketched in the Measurement Setup section. The system is placed on the floor of a room with a height of 3 m imaging the flat ceiling at an angle of 15° in order to localize a target that is not directly visible to the system. An occluding wall that prevents direct imaging of the target (Fig. 1a) would potentially contribute an additional NLOS signal, so it is omitted. A 20 × 20 cm^2 flat target positioned in parallel to the ceiling is used, in contrast to the point-like scatterer assumed in the simulations. Consequently, not only the pulse width but also its spatial extend influences the localization error by introducing multiple ToF per target and detection point. The target is placed near the center of the scene at a distance of 1.5 m from the relay wall (ceiling).

As predicted by the simulations, the localization error decreases as the number of measurements increases (Fig. 6a). When the number of measurements rises from 1,250 to 5,000, reducing the possible repetition rate of a measurement from 10 frames per second (fps) to 4 fps, the localization error decreases from 1 m to 0.74 m and 0.68 m for laser positions 1 and 2, respectively. With more measurements, the localization error for laser position 2 decreases stronger than for position 1; at 12,500 measurements (1 fps), the error is 0.32 m for position 1 and

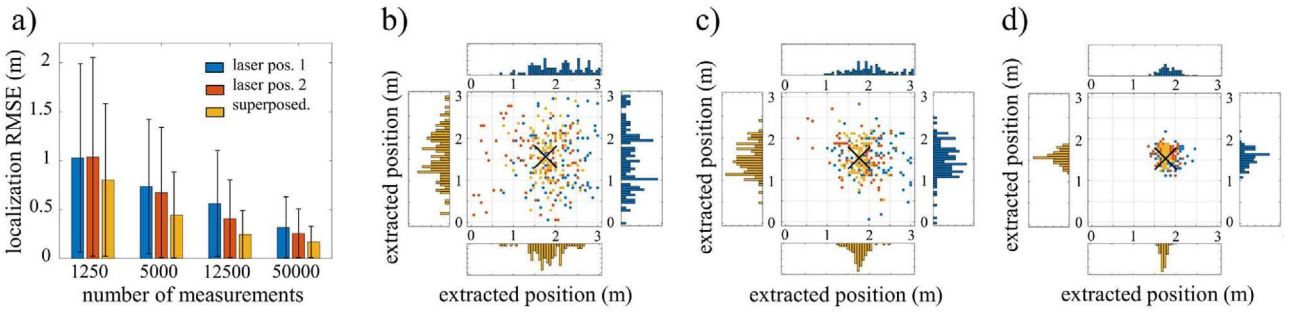


Figure 6. a) Measured localization errors for different measurement statistics. b–d) Extracted lateral target positions of 100 repetitions with 1,250 (b), 5,000 (c) and 50,000 (d) measurements per reconstruction. Histograms show the distribution of extracted target positions for a single illumination (top and right, blue histograms) and superposed reconstructions (bottom and left, yellow histograms) of each lateral axis. The black cross indicates the actual target position.

0.26 m for position 2. This difference is attributed to imperfect Lambertian reflection at the primary reflection of the laser pulse on the ceiling. While the specular component for laser position 1 can geometrically contribute to the target illumination, the direct reflection from position 2 is directed away from the NLOS scene, to the right in Figure 1a. At 50,000 measurements, which corresponds to four seconds of measurement time, the localization error and the standard deviation of the extracted target locations fall even below the values predicted by the simulations. Independent of the individual laser positions, the combination of both illumination points shows the smallest error for all measurement statistics. Therefore, the simulations and the experimental measurements agree concerning the performance of the illumination concept proposed in this work: to achieve a small error in NLOS localization measurements, it is necessary to use more than one illumination points. This is especially important for low measurement statistics or fast measurements, but even at high statistics, using two illumination points can further reduce the localization error.

Nevertheless, there are discrepancies between the simulation and the measurement concerning the generation of the feasible target region that is assessed by the localization error. In contrast to a single-valued plateau approached with increasing measurement statistic for the simulation, the physical setup causes a global maximum to be formed within the laser pulse width induced feasible target region. On the one hand, the spatial extent of the physical target generates confined phantom objects through the overlap of signals from different reflection points on the target. On the other hand, system dependent factors lead to a global maximum of intersections in the reconstruction, which supports convergent localization. The imperfect top hat laser profile affects the signal distribution along the width of the ellipsoid and prevents the formation of an equal valued volume behind the actual target position as discussed above. Furthermore, pixel specific dark count rates and sensitivities influence the signal contribution of each ellipsoid orientation depending on the detection position relative to the illumination position. This artificially confines the overlapping signal volume by favoring reconstruction voxels along individual ellipsoids.

Since the distance from the relay wall of the target position is measured with little error even for low measurement statistics, a two-dimensional map of extracted lateral target positions is used to illustrate the three-dimensional localization in Figures 6b–6d. For each measurement statistic, the target position is measured 100 times and its extracted position is plotted. This naturally results in multiple occupancy of the discrete reconstruction voxel space. To better illustrate convergent localization, the distribution of determined target positions is also indicated by histograms along both lateral axes. Opposite histograms represent the results of superposition (yellow dots) and laser position 1 (blue dots) respectively and are equally scaled. In agreement with the simulations, the reduced localization error is mainly observed along the illumination axis defined by the line connecting the separated illumination points. This can be observed by the narrower distribution in the horizontal histograms compared with the vertical histograms of the super positioned results (yellow).

The compact measurement setup locates a NLOS target successfully under eye-safe conditions and validates the simulated benefits of superposed reconstructions of individual illumination points. The deviation from the homogenous feasible target region predicted by the simulation, resembles the formation of a global maximum in the actual measurement result. Although this effectively decreases the standard deviation of extracted positions, it results in a systematic error depending on the pulse form and target position. This indicates the need for filtering to achieve localization that is independent of pulse width.

3.3 Spatial and temporal filtering

Our demonstrated eye-safe NLOS localization exhibits an extended feasible target region that leads to localization errors larger than the temporal detector resolution and the finite reconstruction grid presets, caused mainly by the temporal laser pulse width. The achievable localization performance, limited by the laser pulse width and measurement statistics, has been addressed through scene illumination schemes and their superposition. To directly address the extended pulse width, associated with the compact laser diodes, filtering of the reconstruction result and the

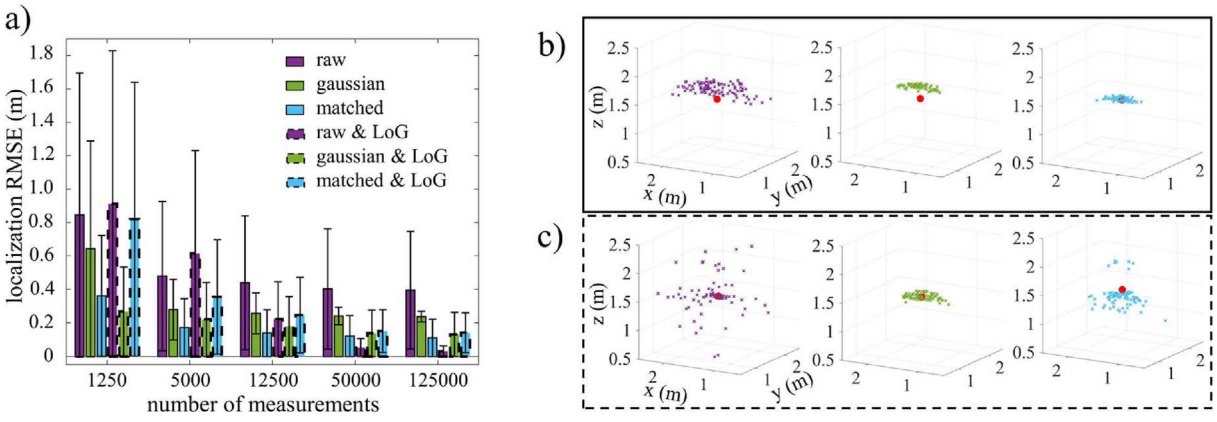


Figure 7. a) Simulated localization errors for filter combinations of no temporal filtering, *Gaussian* filtering and *matched* filtering with an optional spatial *LoG* edge detection. b) Corresponding resulting target positions per temporal filter at 12,500 measurements per reconstruction and a laser pulse width six times the detector's temporal resolution and c) with additional *LoG* edge detection in the reconstruction volume. The red dots denote the nominal target position.

temporal domain data is examined. Filtering of the reconstruction is commonly used in ellipsoid based back-projection methods to increase the visibility of features in complex scenes in NLOS FBP imaging. There, the edge sensitive filters are applied to the reconstruction volume to compensate for uncorrelated ellipsoids from other parts of the scene that overlap with the target signal of interest. Although no other target is present in the localization scenario discussed here, edge filtering of the reconstruction result can still be used to compensate for noise and the pulse width induced extension of the target signal. By design, only the rising-edge of the pulse that interacts with the NLOS scene corresponds to the actual ToF of the three-bounce scattering path. Therefore, filtering in the temporal domain of the measured transients is also investigated that preserves the rising-edge information.

To determine the influence of different filters on the localization result and error, further simulations were carried out. The simulations are based on a data set of 100 repetitions per measurement statistic and filter combination. The filter combination consists of filters used on the temporal domain data and spatial filtering of the reconstruction volume. The temporal filters are given by a *Gaussian* filter aiming for denoising and a *matched* filter for promoting the rising-edge of the received pulse from. The spatial filter is a *Laplacian of Gaussian* (*LoG*) edge sensitive filter to promote the edge of the ellipsoid intersection at the target position. The analysis of the resulting localization error after filtering is performed for different number of measurements per reconstruction (Fig. 7a), ranging from 1,250 to 125,000 measurements. Error bars denote the standard deviation of the determined target position and allow the distinction between systematic localization errors and the variation of extracted positions. Figures 7b and 7c visualize the extracted target positions for a statistic of 12,500 measurements per reconstruction, corresponding to 1 fps of data acquisition. For better visibility, the relay wall and thus the illumination positions are located at the bottom of each point cloud representation. Figure 7b corresponds to reconstruction with temporal filtering only, and Figure 7c shows results with additional *LoG* filtering of the reconstruction volume prior to the target position extraction.

Maximum detection on an unfiltered reconstruction yields the largest localization error for high statistics because the reconstruction forms an extended high signal volume behind the target position as discussed above (purple, solid outline, Fig. 7a). With *Gaussian* filtering in the temporal domain (green, solid outline, Fig. 7a), the standard deviation of the extracted position decreases at high statistics while the RMSE persists. This indicates the formation of a global maximum behind the actual target position at an increased distance from the relay wall (green dots, Fig. 7b). This maximum is formed because the *Gaussian* filter deforms the temporal top hat laser profile. Using a smaller width of the *Gaussian* filter can minimize this deformation and prevent a global maximum, but this resembles a transition to the unfiltered reconstruction and is therefore not pursued further. The *matched* filter (blue, solid outline, Fig. 7a), shows the best performance for results without spatial edge filtering. The filter is designed to generate a maximum at the target ToF in the detection point wise histograms representing the transients on the relay wall. Consequently, the overlapping ellipsoids will form a global maximum in the reconstruction as well (blue, solid outline, Fig. 7b). The position of this maximum corresponds to the actual target position, in contrast to the displaced maximum formed by *Gaussian* filtering. Remaining errors are due to temporal discretization and the finite voxel grid. It should be noted that the localization error still decreases more strongly along the axis defined by the illumination positions, where the geometric benefit of superposition comes into effect (y-axis in Figs. 7b and 7c).

Applying the edge filter to a reconstruction without temporal filtering yields the smallest systematic error at high statistics but performs worst at lower statistics (purple, dashed outline, Fig. 7a). For off-centered targets not shown explicitly, a position dependent systematic error is introduced due to shape sensitive edge detection of the high intensity area. With *Gaussian* filtering in the temporal domain beforehand, the rising edge of the pulse profile is

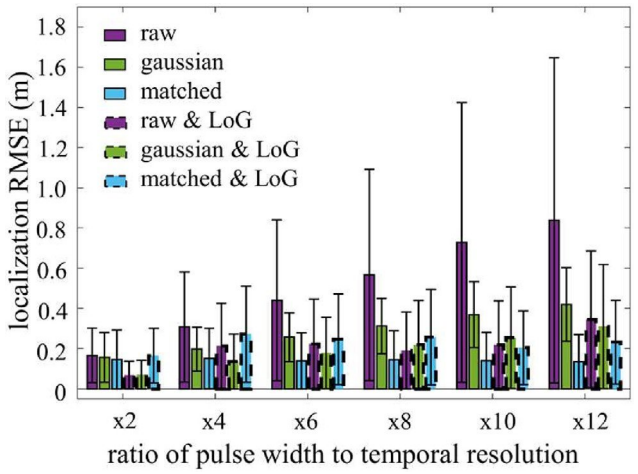


Figure 8. Simulated localization errors for different laser pulse widths at 12,500 measurements using different filter combinations.

flattened, which disadvantages precise edge detection, although the increased SNR enables coarse localization at lower statistics (green, dashed outline, Fig. 7a). For the statistics of 12,500 measurements used in Figure 7c, *Gaussian* filtering (green dots) promotes edge detection near the actual target position, while the temporally unfiltered reconstruction appears to perform worse (purple dots). In fact, the temporally unfiltered result has a comparable RMSE but shows larger offsets for outliers. This bare edge detection without temporal filtering (purple dots Fig. 7c) shows binary localization behavior and yields either the correct target position or an uncorrelated false position. *Matched* filtering combined with edge detection results in a systematic error, since the temporal filter is not designed to retain rising edges for edge detection but is included for completeness. The extracted positions are thus shifted towards the relay wall (blue dots, Fig. 7c).

The filter analysis shows, that a combination of temporal and spatial filtering can decrease the localization error best for the smallest statistics of 1,250 measurements (green, dashed outline, Fig. 7a), by denoising the rising-edge in the temporal regime and thus increasing the contrast of the feasible target region for better spatial edge detection. For higher statistics, the exclusive use of a temporal *matched* filter performs better, that is then outperformed by an exclusive spatial edge filter for the highest statistics. Nevertheless, the latter relies on very steep rising-edges of the pulse form, that is guaranteed for the simulation but not necessarily given by a physical setup due to temporal jitters and limited pulse form qualities.

For the determination of the best filter combination for extended pulse widths in general, the pulse-width-dependent localization performance is examined for a fixed measurement statistic of 12,500 measurements, corresponding to reasonable data acquisition of 1 fps. While shorter pulse widths generally allow for higher laser peak powers, this would increase the SNR in a way that is comparable to increasing the measurement statistics. In order to compare filter performance only, a fixed laser power is used in Figure 8 for each pulse width ranging from twice the value of the temporal detector resolution to a 12-fold pulse width. The direct link between pulse width and high intensity area in the reconstruction is evident for maximum detection of the unfiltered reconstruction, represented by the increasing localization error with increasing pulse width (purple, solid outline, Fig. 8). Besides *matched* temporal filtering without edge detection, all filter combinations show pulse width dependent localization errors. These either originate from detecting different edges of the enlarged overlapping area or from the formation of a displaced global maximum due to the reshaped laser profile. This maximum formation causes a decreased standard deviation of the localization result and introduces a systematic offset that depends on the laser pulse width (green, solid outline, Fig. 8). The exclusive spatial edge detection in the reconstruction volume (purple, dashed outline) shows best performance for the shortest pulse width. That stands to reason, since it is comparable to the standard FBP scenario where this filtering is commonly used. There the pulse width, the temporal detector resolution and the reconstruction resolution are matched, eliminating uncertainties based on temporal pulse width. For any larger pulse widths, the spatial edge detection faces the extended feasible target region that bares the risk of detecting any part of its edge, increasing the error with increasing feasible target region.

In summary, the target location using long laser pulses is best determined using edge detection as is common in FBP imaging. The condition for this is sufficient measurement statistics and a steep rising edge of the pulse form so that the reconstruction resembles the model given in Figure 2c. Nevertheless, systematic but position dependent localization errors occur due to edge detections along the pulse width induced feasibly target region. *Matched* filtering offers pulse width independent performance for localization of single targets in a NLOS scenarios. By incorporating all photons received in the pulse form used, this filtering is particularly suitable for eye-safe systems with limited emission and repetition rates.

4 Conclusion and outlook

Aiming for a compact NLOS localization system, applicable to scenarios outside the laboratory, illumination schemes for limited laser emission and compensation for the compact but long-pulse laser diodes have been examined. Since the reconstruction resolution of NLOS systems depends heavily on temporal detection resolution, the demonstrator presented, which is based on LiDAR hardware with comparatively low-resolution, cannot compete with state-of-the-art systems. However, this platform offers a photon-efficient and eye-safe solution to localization problems outside the direct field of view when exact scene replication is not required.

It was shown that multiple illumination positions can improve NLOS localization performance in two ways of systems that compile with eye-safety restrictions. First, separated laser sources allow for multiplication of emitted signal that can be distributed more equally across the

NLOS scene. This reduces performance variations that depend on the target position in comparison to single illumination schemes. Second, increased variation of illumination and detection pairs further confine the feasible target position as explicitly elaborated for the ellipsoid based back-projection reconstruction.

For single target localization scenarios, *LoG* based edge filtering of the reconstruction as commonly done in FBP NLOS imaging does not adequately counteract pulse-width-induced temporal smearing that enlarges the feasible target region. Preprocessing the temporal measurement data by explicitly managing pulse widths outperforms filtering of the reconstruction result. Matched filtering was used to incorporate all signal photons received, posing an additional advantage over spatial edge detection that relies only on the first part of the pulse form. This fits in with the endeavor to develop a photon-efficient NLOS system, to address the limited emissions in eye-safe operation.

The presented system can be developed further into a hybrid LiDAR-NLOS system with an additional spatially broadened laser source that illuminates the detector's field of view, known as solid-state Flash-LiDAR. A direct ToF measurement can determine the orientation and the system's position relative to the relay wall, which is necessary to process the NLOS measurement. One application of this eye-safe compact hybrid system could be the localization of persons beyond the line of sight on an upper floor through a window, using the ceiling as the relay wall.

Funding

This research received no external funding.

Conflicts of interest

The authors have nothing to disclose.

Data availability statement

The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Author contribution statement

Conceptualization, K.A. and M.L.; Methodology, K.A. and J.K.; Software, K.A. and J.K.; Validation K.A., J.K. and M.L.; Investigation, K.A.; Writing – Original Draft Preparation, K.A. and J.K.; Writing – Review & Editing, J.K., M.L. and A.G; Visualization, K.A. and J.K.; Supervision, M.L. and A.G.

All authors have read and agreed to the submitted version of the manuscript.

References

1 Klein J, Peters C, Martín J, Laurenzis M, Hullin MB, Tracking objects outside the line of sight using 2D intensity images, *Sci. Rep.* **6**, 32491 (2016). https://doi.org/10.1038/srep32491.

2 Musarra G, Lyons A, Conca E, Villa F, Zappa F, Altmann Y, Faccio D, in *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)* (OSA, Washington, D.C.), IM2B.5. https://doi.org/10.1364/ISA.2019.IM2B.5.

3 Lindell DB, Wetzstein G, Koltun V, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2019), p. 6773. https://doi.org/10.1109/CVPR.2019.00694.

4 Maeda T, Wang Y, Raskar R, Kadambi A, in *2019 IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2019), p. 1. https://doi.org/10.1109/ICCPHOT.2019.8747343.

5 Wu C, Liu J, Huang X, Li ZPC, Yu JT, Ye J, Zhang Q, Zhang X, Dou VK, Goyal F, Xu JWP, Non-line-of-sight imaging over 1.43 km, *Proc. Natl. Acad. Sci. USA* **118** (2021). https://doi.org/10.1073/pnas.2024468118.

6 Zhang W, Zhu S, Chen L, Bai L, Lam EY, Guo E, Han J, High-resolution and real-time non-line-of-sight imaging based on spatial correlation, *Opt. Lasers Eng.* **193**, 109100 (2025). https://doi.org/10.1016/j.optlaseng.2025.109100.

7 Nam JH, Brandt E, Bauer S, Liu X, Renna M, Tosi A, Sifakis E, Velten A, Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second, *Nat. Commun.* **12**, 6526 (2021). https://doi.org/10.1038/s41467-021-26721-x.

8 O'Toole M, Lindell DB, Wetzstein G, Confocal non-line-of-sight imaging based on the light-cone transform, *Nature* **555**, 338 (2018). https://doi.org/10.1038/nature25489.

9 Faccio D, Velten A, Wetzstein G, Non-line-of-sight imaging, *Nat. Rev. Phys.* **2**, 318 (2020). https://doi.org/10.1038/s42254-020-0174-8.

10 Liu X, Wang J, Xiao L, Shi Z, Fu X, Qiu L, Non-line-of-sight imaging with arbitrary illumination and detection pattern, *Nat. Commun.* **14**, 3230 (2023). https://doi.org/10.1038/s41467-023-38898-4.

11 Wang K, Zhang M, Shi Y, in *Fourth International Computational Imaging Conference (CITA 2024)*, edited by X. Shao (SPIE, 2024), p. 40. https://doi.org/10.1117/12.3055473.

12 Zhao Z, Zhang Q, Li X, Guo Y, Pu M, Zhang F, Guo H, Wang Z, Fan Y, Xu M, Luo X, High-resolution non-line-of-sight imaging based on liquid crystal planar optical elements, *Nanophotonics* (Berlin, Germany) **13**, 2161 (2024). https://doi.org/10.1515/nanoph-2023-0655.

13 Jin C, Tang M, Jia L, Tian X, Yang J, Qiao K, Zhang S, Scannerless non-line-of-sight three dimensional imaging with a 32x32 SPAD array, arXiv preprint arXiv:2011.05122, 2020.

14 Zhang W, Guo E, Zhu S, Huang C, Chen L, Liu L, Bai L, Lam EY, Han J, Real-time scan-free non-line-of-sight imaging, *APL Photon.* **9** (2024). https://doi.org/10.1063/5.0235687.

15 Haase JF, Grollius S, Grosse S, Buchner A, Ligges M, in *Photonic Instrumentation Engineering VIII*, edited by Y. Soskind, L.E. Busse (SPIE, 2021), p. 20. https://doi.org/10.1117/12.2578775.

16 Leitel R, Dannberg P, Seise B, Brüning R, Grosse S, Ligges M, in *Optical Sensing and Detection VIII*, edited by F. Berghmans, I. Zergioti (SPIE, 2024), p. 71. https://doi.org/10.1117/12.3023124.

17 Liu X, Bauer S, Velten A, Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems, *Nat. Commun.* **11**, 1645 (2020). https://doi.org/10.1038/s41467-020-15157-4.

18 Lindell DB, Wetzstein G, O'Toole M, Wave-based non-line-of-sight imaging using fast f-k migration, *ACM Trans. Graph.* **38**, 1 (2019). https://doi.org/10.1145/3306346.3322937.

19 Zheng Y, Wang W, Zhang C, Zhang Y, Zhang Q, Li L, Converting non-confocal measurements into semi-confocal

ones with timing-accuracy improving for non-line-of-sight imaging, *Opt. Lasers Eng.* **176**, 108067 (2024). https://doi.org/10.1016/j.optlaseng.2024.108067.

20 Laurenzis M, Velten A, Feature selection and back-projection algorithms for nonline-of-sight laser–gated viewing, *J. Electron. Imaging* **23**, 63003 (2014). https://doi.org/10.1117/1.JEI.23.6.063003.

21 Feng X, Gao L, Improving non-line-of-sight image reconstruction with weighting factors, *Opt. Lett.* **45**, 3921 (2020). https://doi.org/10.1364/OL.394742.

22 Velten A, Willwacher T, Gupta O, Veeraraghavan A, Bawendi MG, Raskar R, Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging, *Nat. Commun.* **3**, 745 (2012). https://doi.org/10.1038/ncomms1747.

23 Gu C, Sultan T, Masumnia-Bisheh K, Waller L, Velten A, in *2023 IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2023), p. 1. https://doi.org/10.1109/ICCP56744.2023.10233262.

24 Liu X, Wang J, Li Z, Shi Z, Fu X, Qiu L, Non-line-of-sight reconstruction with signal-object collaborative regularization, *Light Sci. Appl.* **10**, 198 (2021). https://doi.org/10.1038/s41377-021-00633-3.

J. Eur. Opt. Society-Rapid Publ. 2026, **22**, 52
© The Author(s), published by EDP Sciences, 2026
https://doi.org/10.1051/jeos/2026045
Available online at: https://jeos.edpsciences.org

EOSAM 2025
Guest editors: Omar El Gawhary, Stefan Witte, Ignacio Moreno

Journal of the European Optical Society-Rapid Publications

SHORT COMMUNICATION OPEN ACCESS

In-situ monitoring of a flashed solar simulator spectral characteristics for multijunction solar cells performance assessment

Mario V. Imperatore*, Franco Trespidi, Alessandro Minuto, and Gianluca Timò

Ricerca sul Sistema Energetico – RSE SpA, Via V. Callegari 21, Piacenza 29122, Italy

Received 30 January 2026 / Accepted 7 May 2026

Abstract. In the quest for decarbonization, industry and academia are increasing their efforts in research on innovative photovoltaic devices, and proper indoor characterization of devices in an essential element of this research, to correlate manufacturing parameters with electrical properties. For this purpose, it is essential to properly characterize artificial illumination from a solar simulator, as it might directly impact the measured quantities. To fill a literature gap, this work implements triggered high-speed spectrometers for a fast and in-situ characterization of a flashed solar simulator. A commercially available high-speed spectrometer, with a custom-made triggering circuit, was used to evaluate the impact of supply voltage on the spectral components of the solar simulator flash during its temporal evolution. This allowed to identify some spectral lines that decreased far more slowly and never turned off during the measurement window. Supply voltage, surprisingly, introduced large effects when it's varied to adjust irradiance. Spectral Mismatch Ratio was also determined as function of time and wavelength to provide an useful parameter when using artificial radiation. These artifacts impact the measured current–voltage curves, as evaluated on three devices. This fast spectral characterization enriched the indoor characterization procedure, allowing to identify how the extinction spectra vary with time, and to identify unexpected features introduced by changes in supply voltage.

Keywords: Solar simulator, Photovoltaics, Indoor, Characterization.

1 Introduction

In response to the concerning climate changes, such as the steep increase of temperature anomaly w.r.t. 1880 that is approaching 1.5 °C in 2025 [1], actors from both Industry and Academia have increasingly fuelled research on innovative photovoltaic (PV) devices. As a result, global PV capacity is increasing and approaching 2 TW [2].

Research on innovative photovoltaic devices is actively working to increase this value, by making photovoltaic devices more efficient [3, 4] and their power generation more continuous [5]. This research strongly relies on the indoor characterization of PV devices, for the proper assessment under standard, repeatable experimental conditions, to effectively correlate measured electrical properties with device manufacturing steps. The scientific community agreed on Standard Testing Conditions (STCs) for the indoor characterization of PV devices [6, 7], that define the PV cell to be characterized with its temperature kept at 25 °C, under an irradiance of 1000 W/m^2, in 1.5 air mass condition (AM 1.5, corresponding to a zenith angle of 48.2°). International standards such as IEC 60904 and ISO 9845-1 define procedures for indoor characterization of devices, and provide reference spectra that should be used, that correspond to the irradiance received on an equator-facing 37° tilted PV cell [8]. Indoor Characterization in STCs is carried out through solar simulators that accurately measure current–voltage (I–V) characteristics of the device under test (DUT), under artificial illumination with controlled spectrum and intensity. Solar simulator sources can be either pulsed or continuous-wave and are equipped with measurement circuits to record DUT electrical quantities such open-circuit voltage, short-circuit current, power conversion efficiency, and fill factor. During the measurement, the cell temperature is usually kept constant by means of a (e.g., thermoelectric) cooler.

Pulsed solar simulators are usually required when large currents are to be measured, as for multijunction (MJ) concentrating photovoltaics (CPV) devices, that could heat the experimental device up. Spectral matching to the standard spectrum is ensured through proper absorption filter, and any mismatch is evaluated in the spectral mismatch ratio (SMR) according to the standard IEC 60904-7:2008-11 [9]. A triggered measurement circuit is used to perform fast

* Corresponding author: mariovincenzo.imperatore@rse-web.it

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

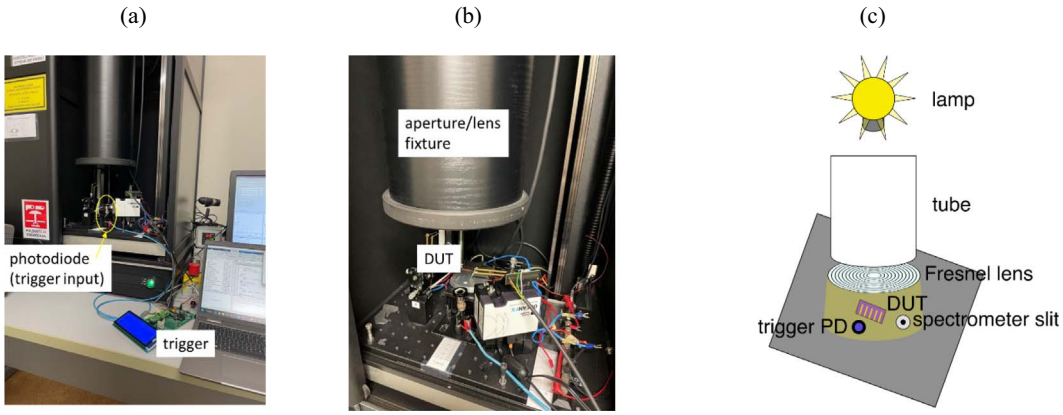


Fig. 1. (a) Measurement setup with a triggered fast spectrometer next to the DUT; (b) 1-sun setup, with radiation from one lamp illuminating the DUT through an aperture/lens fixture; (c) schematic of the setup shown in (b), displaying the trigger photodiode and the spectrometer slit both next to the DUT, under the light from the Fresnel lens.

I-V sweeps (either voltage or current controlled) during the constant region of the flash extinction.

So far, most of literature [10, 11] lacks fast spectral characterization of pulsed solar simulators, and those who employed fast spectrometer setups [12, 13] reported important effects that could not be measured with steady-state spectrometers, such as the illumination spectra shifting from blue rich to red rich during the flash extinction.

This work focused on the spectral characterization of a pulsed solar simulator, which makes use of xenon lamps discharge to produce fast (few milliseconds) flashes with good spectral and intensity uniformity over the illuminated area of the DUT. Solar simulator's pulsed flashes were extensively characterized in time and wavelength, to evaluate the impact of lamp's supply voltage on the flash extinction, and therefore on the measured I–V curves of three DUTs: a top single junction (SJ), a bottom SJ, and a silicon photodiode. The measurement setup can be of low impact, as it can be placed next to the DUT without any relevant shadowing, to allow the in-situ characterization of the exact light hitting the device during that specific measurement.

2 Material and methods

In-situ fast spectral characterization was carried out through the setup displayed in Figure 1. We employed a Technoexan Ltd (an innovation company of the Ioffe Institute in St. Petersburg, Russia) CLASS A (in 400–1200 nm) pulsed solar simulator, designed to work with either semi-collimated 1-sun radiation, with the aperture/lens fixture visible in Figure 1b, or in concentration, up to 5000 suns.

The simulator is designed to ensure a 5% spatial light uniformity across a 2 cm × 2 cm area, a 2% flash-to-flash stability, and can emit a pulse of light once every 15 s.

Current–voltage sweeps are carried out through fast voltage discharges over the 800 µs constant region of the flash extinction, to record light I–V characteristics in the range −5 V/+10 V and up to 10 A.

The solar simulator is equipped with four xenon arc lamps, allegedly identical, each mounted with an absorption filter to include atmospheric absorption and emulate the standard ASTM G173-03 AM1.5 1000 W/m^2 irradiance spectrum. The measured curves should be independent of lamp, from manufacturer's specifications.

To characterize the flash occurring during the I–V sweep, which lasts around 800 µs, a microprocessor-based custom-made circuit was built and set to trigger a measurement once input from a photodiode exceeded a set threshold. Spectra were recorded by the commercially available Ocean Optics FX-XR1-ES CCD spectrometer, used without further modification, in free-space, coupled directly with the semi-collimated light of the simulator. This spectrometer was chosen for its speed, since it's capable to perform one measurement every 221 µs with an integration time of 10 µs, and for its measurement range: 200–1025 nm, with a 0.4 nm resolution (typical). To characterize the whole flash extinction, therefore, 27 consecutive measurements were performed over a 6 ms duration, each containing 2136 points between 200 and 1025 nm. However, only the 350–950 nm range is reported, as it's the most relevant.

As visible in Figure 1, spectrometer slit and trigger photodiode were put directly in the simulator's chamber, next to the DUT.

Together with fast spectral characterization of flashes, I–V curves were recorded by the instrument itself, through its custom source-meter unit that performs voltage–current sweeps during the 800 µs measurement window. Three different photovoltaic devices were measured for this purpose: a top SJ manufactured by Fraunhofer and sensitive in the 300–700 nm range ("SJ top"), a bottom SJ manufactured by Fraunhofer and sensitive in the 850–1800 nm ("SJ bot"), and a silicon photodiode manufactured by THORLABS (p/n FDS1010) and sensitive in the 350–1100 nm range ("Si PD").

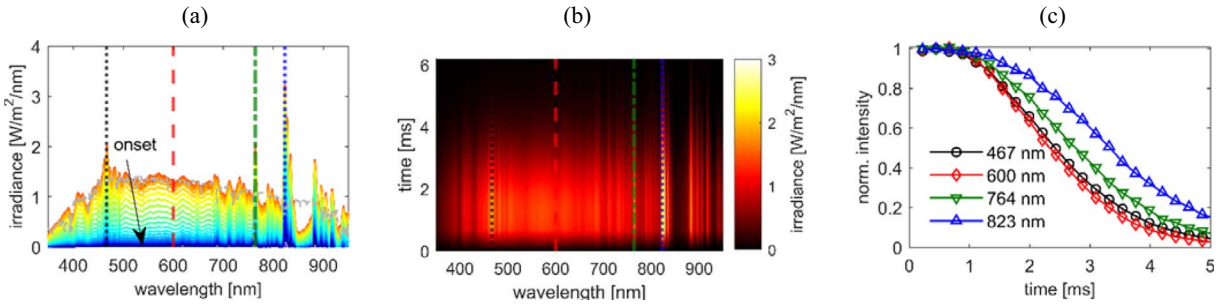


Fig. 2. Waterfall plot and surface plot (b) for the flash extinction from one xenon arc lamp, with one spectrum every 221 µs; the grey curve in (a) displays the ASTMG173 1000 W/m2 reference spectrum; the black arrow indicates the direction of extinction, i.e. from red to blue curves; (c) time series for selected wavelengths from data in (a) and (b), normalized over each curve's maximum.

3 Results and discussion

3.1 Characterization of flash extinction

Figure 2a depicts a typical flash from one lamp at a given supply voltage, compared to the ASTMG173 1000 W/m^2 reference spectrum, that the simulator tries to emulate with a ±25% maximum mismatch (class A). Each spectrum is formed by a broad blackbody radiation that peaks around 570 nm and quickly decreases to 50% after 2.5 ms, superimposed with narrow Xe discharge spectral lines (e.g., 764 and 822 nm) that decrease more slowly, and never turn off completely during the 6 ms window. Since different wavelengths exhibit different dynamics, the temporal variation should be considered at each wavelength, such as those visible in the plots in Figures 2b–2c.

3.2 Impact of supply voltage on lamp's flash

Supply voltage is varied to adjust the simulated irradiance until the 1-sun short-circuit current density (J_{sc}) of the DUT matches that of a calibrated reference cell, regardless of the differences in spectral response of the two devices [1], which is taken into account by the spectral mismatch factor (SMM, or M). When adjusting the supply voltage, one expects the whole spectrum to vary coherently, with every wavelength showing the same increase/decrease with increasing/decreasing voltage. If this was the case, spectral irradiance at any wavelength would increase linearly with supply voltage, integrated irradiance would increase linearly with supply voltage, and measured J_{sc} would increase linearly as well.

Results in Figure 3a, however, report that this occurs only for the blackbody radiation (around 600 nm), and that the wavelengths corresponding to the Xe lines (e.g., 764 and 823 nm) exhibit a slower increase with voltage, marked by the dips visible in Figure 3b. Irradiance does not increase coherently with supply voltage, as visible in Figure 3c, but lines and 764 and 823 nm exhibit slower increases with voltage. This non-ideal behaviour might introduce measurement artifacts, as spectra with higher supply voltage lacks some IR spectral components.

The effect can thus be estimated through the top/middle Spectral Mismatch Ratio plotted in Figure 3d, defined as $SMR_{top/mid} = \frac{\int E_{mistop}(\lambda)d\lambda}{\int E_{mismid}(\lambda)d\lambda} \frac{\int E_{refmid}(\lambda)d\lambda}{\int E_{reftop}(\lambda)d\lambda}$, being $E(\lambda)$ the measured/reference spectral irradiance in W/m^{2}nm for the InGaP-top (350–650 nm) or InGaAs-middle (650–950 nm) range. Consistent with results in Figure 2, and with results reported in [2], the $SMR_{top/mid}$ within the first 5 ms increases with applied voltage and decreases with time as the spectra become richer with IR, since these components decrease faster during the flash extinction.

3.3 Impact of supply voltage on measured I–V curves

Results reported in Section 3.2 suggest that the spectrum at one supply voltage (e.g. 650 V) does not increase coherently with voltage, but discharge at higher voltages (e.g. 800 V) lack of some spectral components, such as 467, 764, 823 nm, etc. Varying supply voltage without evaluation of its impact might lead to an underestimation of the simulator's irradiance and to an underestimation of the measured J_{sc}.

In order to prove this claim, we measured the I-V curves of three different devices, each sensitive to a different spectral range, thus each including a different number of the dips shown in Figure 3b. We expect the more dips are included in the DUT's range, the stronger underestimation of the J_{sc}.

For this purpose, Figure 4 shows, for each device, the J_{sc} at different supply voltage (650, 700, 750, 800 V), normalized over its value at 650 V. Each data point is the result of 5 consecutive measurements.

4 Discussion

Results in Section 3.1 reveal that a typical xenon lamp's intensity does not decrease coherently during its extinction, but it's formed by spectral lines that decrease far more slowly than others. This dynamics can be considered coherent enough during the first ~1 ms of the extinction, where voltage–current sweeps are carried out, as shown in the SMR reported in Figure 3d. This assumption was possible thanks to speed of the spectrometer that allowed to characterize the flash dynamics with good temporal resolution.

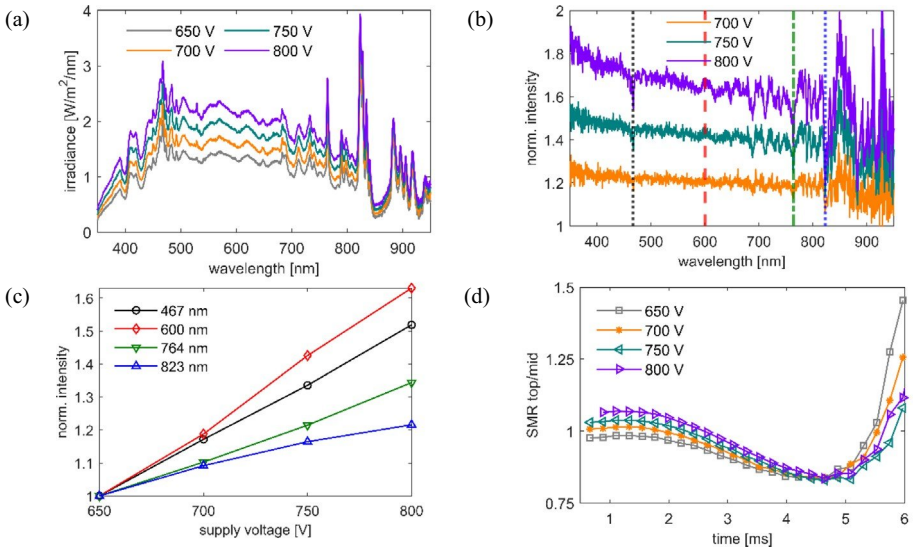


Fig. 3. (a) Irradiance maxima recorded for the same lamp at four different supply voltages; (b) normalized curves from (a) over the 650 V curve (black line); (c) values at specific wavelengths from (b) plotted vs. supply voltage; (d) $SMR_{top/mid}$ plotted vs. time, for four different supply voltages.

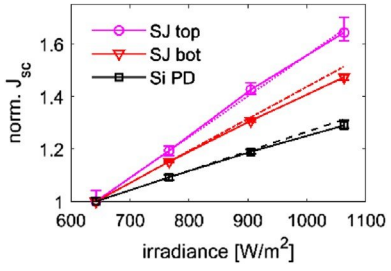


Fig. 4. Short-circuit current density, normalized over its initial value, versus the integrated irradiance (350–950 nm). For each device, the solid curve represents the measured data, and the dotted curve represents the linear regression of the first two points of the curve.

Results in Section 3.2 report that when supply voltage is varied to adjust irradiance, spectra do not change, but those wavelengths appearing as dips in Figure 3b increase more slowly, and the higher-voltage spectra appear poorer of these spectral components. SMR@1 ms in Figure 3d decreases from 1.069 to 0.985 as supply voltage is decreased from 800 to 650 V, which gives a clear numerical value to quantify the effect.

The impact of supply voltage on measured I–V curves is depicted in Figure 4. These results compare the ideal behaviour where spectra increase coherently with supply voltage (dotted lines), and the real behaviour (solid lines) constituted by the actual measurements. The magnitude of this discrepancy is quantified by the "Max nonlinearity" in Table 1, which is high for the "Si PD" device since it includes a large number of dips in its large conversion range 850–1800 nm, and highest for the "SJ bot" device, which conversion range 850–1800 nm extends further into the IR, beyond the spectrometer's range. These results suggest that further into IR, although not measurable, even more dips might appear and strongly affect the "SJ bot" current. Contrarily, being most of these dips in the 650–850 nm range, the "SJ top" device is almost unaffected, as its sensitivity range 300–700 nm barely includes any dip.

Table 1. For each device, maximum difference between measured norm. J_{sc} and its linear regression (max. difference between solid and dotted line), and $\Delta J_{sc} = J_{sc}$@800 V − fit@800 V over J_{sc}@800 V.

Device	Max nonlinearity	$\Delta J_{sc}/J_{sc}$ [%]
SJ top	−0.0117	−0.71
SJ bot	−0.0406	−2.75
Si PD	−0.0226	−1.75

Comparison between experimental results and those the device would have in ideal conditions, namely if all the spectral components would increase coherently with voltage, leads to the outcome that J_{sc} is always underestimated, up to −2.75% ("SJ bot").

5 Conclusion

Results in this work highlighted measurements artifacts that can impact indoor characterization, if not identified and considered. Temporal variation of spectral lines is not the same for every wavelength, but the discrepancy is acceptable for the first ∼1 ms of the discharge, where the fast I–V sweeps occur. Supply voltage, surprisingly, introduces 10% inhomogeneity in lamp's spectral components when increased by just 50 V, and 20% inhomogeneity when increased from 650 to 800 V. Consequently, irradiance does not scale linearly with supply voltage, as some spectral components are modified less than others. This yields to a

voltage-dependent short-circuit current density underestimation up to −2.75%. The proposed methodology requires an expensive spectrometer, but offers an easy in-situ implementation, providing spectral measurements for each flash hitting the DUT. Results in this work suggest to extend the measurement range further into the infrared region, as infrared spectra at higher supply voltages could differ even more from the lower supply voltages ones.

Funding

This work was financed by the Research Fund for the Italian Electrical System under the Three-Year Research Plan 2025–2027 (DM MASE n. 388, 06.11.2024), in compliance with the Decree of April 12th, 2024.

Conflicts of interest

All authors have nothing to disclose.

Data availability statement

Data available upon reasonable request.

Author contribution statement

Conceptualization and Methodology, M.V. I., F. T., and A. M; Data Curation, Writing, and Original Draft Preparation, M. V. I.; Review & Editing, Supervision, G. T.

References

1 *NASA GISS Surface Temperature Analysis (v4)*. https://data.giss.nasa.gov/gistemp/graphs_v4/
2 REN21-Renewable Energy Policy Network for the 21st Century, *Renewables 2024 Global Status Report – Renewables in Energy Supply* (2024). https://www.ren21.net/wp-content/uploads/2019/05/GSR2024_Demand_Full_Report.pdf.
3 King RR et al., Solar cell generations over 40% efficiency, *Prog. Photovolt: Res. Appl.* **20**, 801–815 (2021). https://doi.org/10.1002/pip.1255.
4 Fraunhofer ISE press release, *Fraunhofer ISE Develops the World's Most Efficient Solar Cell with 47.6 Percent Efficiency* (2022). https://www.ise.fraunhofer.de/en/press-media/press-releases/2022/fraunhofer-ise-develops-the-worlds-most-efficient-solar-cell-with-47-comma-6-percent-efficiency.html.
5 Sharma T et al., A short review on long persistent luminescence materials and their application prospects in emerging photovoltaic devices, *Heliyon* **10**, e40341 (2024). https://doi.org/10.1016/j.heliyon.2024.e40341.
6 IEC Standard 60904-1. *Measurement of Photovoltaic Current-Voltage Characteristics* (2020). https://webstore.iec.ch/en/publication/32004.
7 IEC Standard 60904-3. *Measurement Principles for Terrestrial Photovoltaic (PV) Solar Devices with Reference Spectral Irradiance Data* (2019). https://webstore.iec.ch/en/publication/61084.
8 International Standard ISO *9845-1 Solar Energy – Reference Solar Spectral Irradiance at the Ground at Different Receiving Conditions* (1992). https://www.iso.org/standard/79989.html.
9 IEC standard 60904-7. *Computation of the Spectral Mismatch Correction for Measurements of Photovoltaic Devices* (2019). https://webstore.iec.ch/en/publication/26502.
10 Gerstenberger F et al., Spectrally accurate calibration of pulsed solar simulators for multijunction CPV measurement, *AIP Conf. Proc.* **2550**, 020004 (2022). https://doi.org/10.1063/5.0099803.
11 Pravettoni M et al., Characterization of a pulsed solar simulator for concentrator photovoltaic cell calibration, *Measure. Sci. Technol.* **21**, 115901 (2010). https://doi.org/10.1088/0957-0233/21/11/115901.
12 Dong X et al., Time-resolved spectra of solar simulators employing metal halide and xenon arc lamps, *Sol. Energy* **115**, 613–620 (2015). https://doi.org/10.1016/j.solener.2015.03.017.
13 Veinberg-Vidal E et al., Characterization of dual-junction III-V on Si tandem solar cells with 23.7% efficiency under low concentration, *Prog. Photovolt: Res. Appl.* **27**, 652–661 (2019). https://doi.org/10.1002/pip.3128.

European Optical Society

The European Optical Society is a non-profit society, and an umbrella organization for all national optical societies around Europe. Our members extend from Europe to all over the world.

Our mission is to foster collaboration among individuals and organizations interested in optics, optoelectronics, and related fields. We strive to advance research, facilitate its practical applications, and promote the industrial potential of optics. To achieve this, we organize international conferences, workshops, training sessions, exhibitions, and other events. Additionally, we maintain strong connections with like-minded societies and organizations worldwide.

By joining EOS, you become part of a vibrant and expanding community dedicated to optics and photonics. With over 4,000 members—including societal, corporate, individual, associate, and student categories—we welcome both individuals and institutions.

EOS membership includes various benefits. You can explore the benefits and advantages here:
https://www.europeanoptics.org/pages/members/join-us/benefits.html

EOSAM

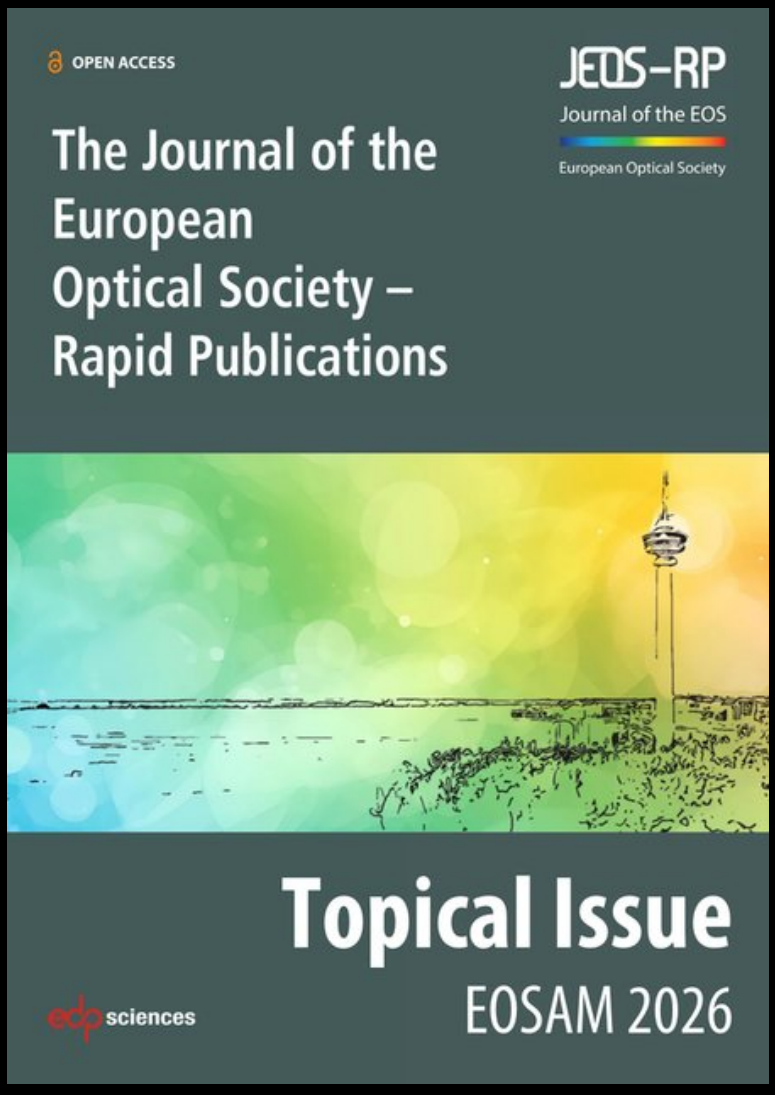
The European Optical Society Annual Meeting, EOSAM, is a major international scientific conference and exhibition covering all aspects of optics and photonics.

EOSAM covers all aspects of optics and photonics within several topical meetings and sessions. It is attended annually by around 500 top researchers, key leaders, students, and industry experts from over 30 different countries all over the world.

EOSAM is a great platform for presenting the most recent research results, connecting and catching up, as well as bridging the gap between research, education, and industry. Each year EOSAM moves into different optics and photonics hubs better serving the local communities and at each time creating a unique EOSAM experience for the attendees. EOSAM has been organized since 2006 in France, Scotland, Germany, the Netherlands, Italy, Portugal, and Finland.

EOSAM 24 - 28 August 2026 Tampere, Finland

European Optical Society Annual Meeting



EOSAM 2026 TOPICAL ISSUE: PUBLISH AN EXTENDED PAPER

EOSAM 2026 took place in Tampere, Finland, from 24 to 28 August 2026.
Participants can still submit extended articles based on their conference presentations to the EOSAM 2026 Topical Issue of the Journal of the European Optical Society-Rapid Publications (JEOS-RP).
The submission deadline is **December 31, 2026**.

JEOS-RP published by EDP Sciences in cooperation with:

DGaO Deutsche Gesellschaft für angewandte Optik
SIOF Società Italiana di Ottica e Fotonica
Prom Optica
SPOF
LOS Latvian Optical Society
Photonics Finland
SEDOPTICA Sociedad Española de Óptica
PhotonicSweden
SFO Société Française d'Optique

Why Publish in JEOS-RP?

- Learned society journal
- Rigorous peer review
- Strict ethical policies following the COPE code of conduct (read more)
- Fast publication times – articles are published online within 3 days of acceptance
- Very fair publication fees with no page charges and further discounts for EOS members
- Open-access Creative Commons licence (CC BY)
- Author services: performant search engine, article metrics, bookmarking, online submission, and peer review portal...
- Reader services: free e-mail alerts and RSS Feeds, CrossRef®, and DOI
- Large visibility of the published papers through indexation in main databases

jeos.edpsciences.org

Editor-in-Chief: Silvia Soria
Co-Editor-in-Chief: Martina Delgado-Pinar

The official journal of
EOS European Optical Society