TU Delft

Delft University of Technology

Coherent Fourier Scatterometry for sensitive detection of subwavelength structures and particles

Kolenov, D.

DOI
10.4233/uuid:c189b14b-c454-4f13-8658-5c6210bb652e

Publication date
2022

Document Version
Final published version

Citation (APA)
Kolenov, D. (2022). *Coherent Fourier Scatterometry for sensitive detection of subwavelength structures and particles*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:c189b14b-c454-4f13-8658-5c6210bb652e

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Coherent Fourier Scatterometry for sensitive detection of subwavelength structures and particles

Coherent Fourier Scatterometry for sensitive detection of subwavelength structures and particles

Dissertation

for the purpose of obtaining the degree of doctor at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Tuesday of 25 October 2022 at 10:00 o'clock

by

Dmytro KOLENOV

Master in Radio Physics and Electronics,
V. N. Karazin Kharkiv National University, Kharkiv, Ukraine,
born in Kharkiv, Ukraine.

This dissertation has been approved by the promotors

promotor: Prof. dr. H.P. Urbach
promotor: Dr. ir. S.F. Pereira

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. H.P. Urbach	Technische Universiteit Delft, promotor
Dr. ir. S.F. Pereira	Technische Universiteit Delft, promotor

Independent members:

Prof. dr. A. J. den Boef	Vrije Universiteit Amsterdam
Prof. dr. W. M. J. M. Coene	Technische Universiteit Delft
Prof. dr. M. Hofmann	Ruhr-Universität Bochum
Prof. dr. S. Stallinga	Technische Universiteit Delft
Dr. P. van der Walle	TNO (Toegepast Natuurwetenschappelijk Onderzoek)

This work has been financially supported by High Tech Systems and Materials Research Program, Applied and Technical Sciences division (14660)



Keywords: Scatterometry, particle detection, computational physics

Printed by: Proefschriftspecialist

Copyright © 2022 by D. Kolenov

ISBN 978-90-832797-1-8

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

Summary ix

Samenvatting xi

1 Introduction 1
1.1 Scattering . . . 1
1.2 Quality control for the semiconductor industry . . . 2
1.3 Surface inspection in the semiconductor industry . . . 3
1.4 Goals of this thesis . . . 6
1.5 Outline . . . 7
References . . . 8

2 Introduction to Electromagnetic theories and Numerical Modeling 11
2.1 Coherent Fourier Scatterometry . . . 12
2.2 Vectorial diffraction integral . . . 13
2.3 Depth of focus and high NA . . . 19
2.4 A dipole radiation in free space and its far-field . . . 20
2.5 Field in the detector for a dipole at an interface excited by the focused spot . . . 23
2.6 Numerical modeling . . . 27
2.6.1 FDTD . . . 27
2.6.2 FEM . . . 29
2.6.3 Convergence test . . . 29
2.6.4 Definition of the simulation domain . . . 31
References . . . 32

3 Heterodyne Detection System for Nanoparticle Detection 34
3.1 Introduction . . . 35
3.2 Methods and materials . . . 35
3.3 Results . . . 38
3.4 Conclusions . . . 40
References . . . 40

4 Polarization effects in evanescent wave amplification for particle detection 42
4.1 Introduction . . . 43
4.2 Enhancement mechanism . . . 44
4.3 Thin film material analysis . . . 44
4.4 Numerical Results . . . 46
4.5 Experimental Results . . . 56
4.6 Discussion and Conclusion . . . 59

References . . . 60

5 Highly-sensitive laser focus positioning method with sub-micrometer accuracy. 62
5.1 Introduction . . . 63
5.2 Methods . . . 63
5.2.1 Measurement Approach . . . 63
5.2.2 Position-dependent phase . . . 66
5.2.3 Uncertainty Estimation . . . 69
5.3 Results . . . 72
5.4 Discussion and Conclusions . . . 74
References . . . 75

6 Efficient signal processing with the aid of Machine Learning. 78
6.1 Introduction . . . 79
6.2 Methods . . . 80
6.3 Sub-problems . . . 81
6.3.1 Pre-processing . . . 81
6.3.2 Selection of suitable Amplitude and Width . . . 82
6.3.3 Multiple line particle detection identification . . . 82
6.4 Algorithm . . . 82
6.4.1 Pre-processing . . . 82
6.4.2 Selection of suitable Amplitude and Width . . . 84
6.4.3 Fine adjustment for the boundaries of particle's signal . . . 86
6.4.4 Clustering of data from one single particle . . . 87
6.5 Results . . . 91
6.5.1 Pre-processing and search. . . . 91
6.5.2 Comparing the accuracy of clustering routines on a dataset with drift. . . . 93
6.5.3 Benefit of the centroids re-assignment . . . 95
6.6 Discussion and Conclusions . . . 98
References . . . 99

7 Convolutional Neural Network applied for nanoparticle classification 101
7.1 Introduction . . . 102
7.2 Method . . . 103
7.3 Results . . . 106
7.3.1 Closed set classification . . . 106
7.3.2 Comparison with thresholding classification method . . . 107
7.3.3 Towards Multi-class Open Set Classification . . . 109
7.4 Discussion and Conclusions . . . 111
References . . . 113

8 The limits of detection 115
8.1 Introduction . . . 116
8.2 Experimental results on the minimum size limit of particle detection . . . 116

8.3 AFM for background surface and no particle CFS scans 118
8.4 Effects of the roughness of the surface 120
8.5 Accessing the noise level of the detector 124
8.6 Noise originating from vibrations 126
8.7 Limit of detection 129
8.8 Analysis of pellicle sample 131
8.9 Conclusions 132
References 134

9 Conclusion and Outlook **135**
9.1 Conclusion 135
9.2 Outlook 1. Detection below the background level. 136
9.3 Outlook 2. Nanoparticle shell-layer quantification 137

A Theoretical background **140**
A.1 Maxwell equations, boundary conditions, time-harmonic fields and Green's function. 140
A.2 Angular spectrum representation, far-field and Fourier transform by a lens 143
References 147

B Schematic of the legacy split detector **148**

C Fabrication of experimental samples **149**
C.1 Spin-coated PSL particles 149
C.2 Protocol for PSL sample analysis 151
C.3 Negative resist particles 153

Acknowledgements **155**

Curriculum Vitæ **158**

List of Publications **159**

Summary

Light incident on a medium that contains localized irregularities deviates from the straight path in more than one direction, both in reflection and in transmission. This phenomenon is called scattering. A straightforward example is light that is reflected or transmitted into several angles from a rough surface. Optical scatterometry is a method that can be used for the characterization of unknown properties of a medium by measuring some parameters or properties of the scattered light, such as its intensity distribution, polarization and coherence. The scattered field is usually detected or observed at a distance of many wavelengths from the medium, in the so-called far-field region. Far-field detection is convenient in terms of data acquisition, but the drawback is that, in the far field, high spatial frequency information contained in the near field is lost. By measuring the scattered far field - one does not obtain a direct image, but instead by means of solving an inverse electromagnetic model and taking a priori knowledge into account, one can infer some parameters of interest of the medium/scatterer.

In this thesis, we explore a particular scatterometry technique, in which a coherent light beam is focused by a lens on a surface to be inspected and the scattered light is collected and detected in the far field. Since the focused spot illuminates a small region of the sample, lateral scanning of the focused beam is required in order to cover the entire area of interest. This technique is called Coherent Fourier Scatterometry (CFS). Because CFS enables the detection of subwavelength particles of low optical contrast and at low light levels, the detection, localization and classification of isolated particles on surfaces is an important application of CFS. In this dissertation, we explore several aspects of nanoparticle detection using CFS and introduce new applications of the technique.

In Chapter 1, we introduce the application of scattering for the detection and localization of tiny objects on top of a surface. Furthermore, we show how the technique can be applied for quality control of surfaces, in particular for the semiconductor industry, with the substrate being a silicon wafer. We furthermore summarize the limitations of CFS as compared to existing surface inspection tools.

The concept of CFS and the main optical setup are outlined and explained in Chapter 2. This chapter also lays the foundations for the analytical and numerical modeling that has been used in this dissertation.

In contrast to imaging techniques that are limited by diffraction, CFS is essentially limited by noise. In Chapter 3, we show the implementation of a heterodyne detection system in order to suppress the experimental noise. The demonstrated improvement of the signal-to-noise ratio (SNR) allows us to detect particles of smaller sizes than in previous experiments. The ability to minimize the noise in the measurement process enables us to push the technique for the detection of particles in the domain of deep subwavelength sizes, which is critical in many applications.

In Chapter 4, we investigate the unique properties of highly focused vector beams,

possessing conventional (linear and circular) and cylindrical (radial and azimuthal) polarization states, combined with substrates coated with a specially designed thin film. The design takes into account the phenomenon of evanescent wave amplification, resulting in an improvement in the sensitivity of detection.

In order to maximize the scattering due to the nanoparticle, the focused beam that is used for the illumination should have its focal plane at the surface that contains the nanoparticle. In Chapter 5, we study the effects of defocusing the illuminating beam, and as result, we propose a novel method of determining the focus position on a surface that relies on measuring the scattering of an isolated particle. We show that our method has superior focus position performance as compared to existing techniques based on astigmatism.

One of the main applications of scatterometry is to estimate the amount of contamination on a surface. The quality/cleanliness of, for example, a silicon wafer is judged by determining the density, location and size distribution of the detected particles. This data is extracted by monitoring the detector signal as a function of the position of the illumination spot on the surface by means of a lateral raster scan along the surface to be inspected. The smaller the particle size we are aiming to measure, the smaller the scanning step should be, providing more measurements at more positions. This leads to a growing amount of data to be processed and stored, which is a challenging scenario.

In Chapters 6 and 7, we look at ways to improve data acquisition and post-process workflow. These ways involve the searching, clustering, calibration, and classification of particle signals, as well as the rejection of anomalies. Our results illustrate that the proposed algorithms can detect and classify surface contaminants correctly and effectively for the dataset, both with and without drifts in the raster scan positions.

Armed with the improvements of the system from the previous chapters, we investigate in Chapter 8, the detection of particles in the deep subwavelength regime and determine the ultimate limits of our instrument. We analyze the electronic noise, vibration sources and the effect of surface roughness, resulting in the ability to detect a latex sphere of 29 nm diameter on a silicon wafer at a wavelength of 405 nm.

Samenvatting

Licht dat invalt op een medium met lokale onregelmatigheden wijkt af van het rechtdoorgaande pad in meer dan één richting, zowel bij weerspiegeling als bij doorvalling. Dit fenomeen heet verstrooiing. Een eenvoudig voorbeeld is licht dat onder meerdere hoeken wordt weerspiegeld of doorgelaten door een onregelmatig oppervlak. Optische verstrooiingsmeting is een methode die kan worden gebruikt voor het bepalen van onbekende eigenschappen van een medium door een aantal parameters of eigenschappen van het licht te meten, zoals zijn intensiteitsverdeling, polarisatie en coherentie. Het verstrooide veld wordt gewoonlijk gedetecteerd of waargenomen op een afstand van meerdere golflengtes van het medium, in het zogenaamde verre veld. Detectie in het verre veld is geschikt voor gegevensverwerving, maar het nadeel is dat hoogfrequente informatie aanwezig in het nabije veld verloren raakt in het verre veld. Door het verstrooide verre veld te meten verkrijgt men niet een direct beeld, maar in plaats daarvan, door het oplossen van een invers elektromagnetisch model en a priori kennis in acht te nemen, kan men een aantal belangrijke parameters van het medium/de verstrooier afleiden. In dit proefschrift verkennen we een speciale verstrooiingsmeettechniek waarbij een coherente lichtstraal door een lens wordt gefocusseerd op een te inspecteren oppervlak en het verstrooide licht wordt verzameld en gedetecteerd in het verre veld. Omdat de gefocusseerde lichtvlek slechts een klein gebied van het monster belicht, is zijdelingse aftasting van de gefocusseerde straal vereist zodat het volledige gebied van belang wordt gedekt. Deze techniek wordt Coherente Fourier Verstrooiingsmeting (CFV) genoemd. Omdat CFV de detectie van deeltjes kleiner dan de golflengte en van laag optisch contrast mogelijk maakt bij weinig licht, is de detectie, lokalisatie en classificatie van afgezonderde deeltjes op oppervlakken een belangrijke toepassing van CFV. In dit proefschrift verkennen we meerdere aspecten van de detectie van nanodeeltjes door middel van CFV en introduceren we nieuwe toepassingen van de techniek. In hoofdstuk 1 introduceren we de toepassing van verstrooiing voor de detectie en lokalisatie van kleine deeltjes bovenop een oppervlak. Verder tonen we hoe de techniek kan worden toegepast op de kwaliteitscontrole van oppervlakken, in het bijzonder voor de halfgeleiderindustrie waarbij het substraat een silicium wafer is. Verder vatten wij de beperkingen van CFV, vergeleken met bestaande inspectiegereedschappen voor oppervlakken, samen. Het CFV-concept en de belangrijkste optische opstelling worden geschetst en uitgelegd in hoofdstuk 2. Dit hoofdstuk legt ook de basis voor de analytische en numerieke modellering die in dit proefschrift gebruikt is. In tegenstelling tot beeldvormingstechnieken die door diffractie beperkt worden, wordt CFV hoofdzakelijk beperkt door ruis. In hoofdstuk 3 tonen we de verwezenlijking van een heterodyne detectiesysteem voor het onderdrukken van experimentele ruis. De aangetoonde verbetering van de signaal-ruisverhouding (SRV) stelt ons in staat deeltjes te detecteren die kleiner zijn dan die in eerdere experimenten. Het vermogen de ruis te minimaliseren stelt ons in staat de techniek van deeltjesdetectie naar het domein van diep-subgolflengte afmetingen uit te breiden, wat van cruciaal belang is

voor vele toepassingen. In hoofdstuk 4 onderzoeken we de unieke eigenschappen van sterk gefocusseerde vectorstralen, die gebruikelijke (lineair en circulair) en cilindrische (radiaal en azimutaal) polarisatietoestanden bezitten, gecombineerd met substraten gecoat met een speciaal ontworpen dunne film. Het ontwerp houdt rekening met het verschijnsel van verdwijnende golfversterking, met een verbeterde detectiegevoeligheid als resultaat. Om de verstrooiing door het nanodeeltje te maximaliseren, moet de gefocusseerde straal die voor de belichting gebruikt wordt zijn brandvlak hebben in het oppervlak dat de nanodeeltjes bevat. In hoofdstuk 5 bestuderen we de effecten van het onscherp maken van de belichtingsstraal. Als resultaat stellen we een nieuwe methode voor, gebaseerd op het meten van de verstrooiing door een afgezonderd deeltje, voor het bepalen van de brandpuntpositie op een oppervlak. We laten zien dat onze methode beter presteert als het gaat om brandpuntposities vergeleken met bestaande technieken op basis van astigmatisme. Eén van de voornaamste toepassingen van verstrooiingsmeting is het inschatten van de hoeveelheid vervuiling op een oppervlak. De kwaliteit/reinheid van bijvoorbeeld een silicium wafer wordt beoordeeld door het bepalen van de dichtheid, plaats en grootteverdeling van de gedetecteerde deeltjes. Deze gegevens worden gewonnen door het volgen van het detectorsignaal als functie van de positie van de belichtingsvlek op het oppervlak tijdens een zijdelingse rastervormige aftasting van het te inspecteren oppervlak. Hoe kleiner de deeltjesgrootte is die willen we meten, hoe kleiner de aftastingsstap moet zijn, waardoor meer metingen op meer posities worden verschaft. Dit leidt tot een groeiende hoeveelheid gegevens die verwerkt en opgeslagen moet worden, wat een uitdagend scenario is. In hoofdstuk 6 en 7 kijken we naar manieren om het proces van gegevensverwerving en nabewerking te verbeteren. Deze manieren hebben betrekking op het zoeken, groeperen, kalibreren en classificeren van signalen van deeltjes, evenals het verwerpen van anomalieën. Onze resultaten illustreren dat de voorgestelde algoritmes oppervlakteverontreinigingen juist en doeltreffend kunnen detecteren én classificeren voor de meetreeks, zowel met als zonder afdrijving van de aftastingsposities in het raster. Bewapend met de verbeteringen van het systeem uit de vorige hoofdstukken, onderzoeken we in hoofdstuk 8 de detectie van deeltjes in het diep-subgolflengteregime en bepalen we de uiterste limieten van ons instrument. We analyseren de elektronische ruis, trillingsbronnen en het effect van oppervlakteruwheid, met als resultaat het vermogen een latex bol met een diameter van 29 nm op een silicium wafer te detecteren met een golflengte van 405 nm.

1

Introduction

1.1. Scattering

The phenomenon of scattering is ubiquitous. The human eye sees it as a "blue" sky in a summer morning or a diffuse glow during the night, the color of a laser or fog in the air. Alternatively, scattering recorded with a state-of-the-art instrument manifests itself in collisions between atoms, electrons, and photons, such as processes in nuclear reactors or inside an accelerator, and high-energy electrons precipitation in the atmosphere. Scientists across the world are interested in studying the scattering effects that take part in the interaction between the light and matter in order to determine the physical properties of materials. The dimensions of an arbitrary structure and its quality is adequately studied with optical metrology, which is the science of measurements with light. Staring from the height and width determination of a distant object by triangulation over many decades, optical metrology has developed to the vibrant area of technological advancements [1]. Numerous optical techniques and instruments, such as microscopes, wavefront sensors, optical comparators, and interferometers are available now where subnanometer precision is achieved. By studying the properties of the electromagnetic field that is generated from the interaction between the probe and the unknown target, it is nowadays possible to retrieve intricate parameters of the object such as its shape and roughness. Commonly, the measurement in the far-field regime is adopted since it is noninvasive. In the far-field, for successful information retrieval of features smaller than the Rayleigh limit, the inverse problem of scatterometry (optical metrology with scattered light) requires a priori information. In many cases, we assume that the target under the study is guaranteed to exist, and it is partially known. For example, one can deposit particles of certified material on top of a surface, measure the intensity of the scattered field, and by combining this information with electromagnetic models, one can deduce parameters such as size and position of the scatterer. However, when the target becomes extremely small, as for example, a fraction of the wavelength, the question arises: given the measuring instrument, would we still detect this target? The answer is yes, if the sensitivity of the instrument is high enough. Many areas of optics and physics rely on the detection and localization of tiny objects on top of a surface. The main examples

include contamination and nanofabricated features and defects in the semiconductor industry [2, 3], the studies of viruses and bacteria for biological and medical sciences [4, 5], air and water pollution with toxic particles in environmental science [6, 7].

In this thesis, we will concentrate on the application of scatterometry for isolated nanoparticle detection aimed at quality control in the semiconductor industry.

1.2. Quality control for the semiconductor industry

One of the big challenges in the semiconductor industry is the quality control of the entire lithograph process; it includes the unwanted contamination on the silicon wafers in the nanometer-size scale (Figure 1.1). This contamination can occur at different stages of the lithography process on the wafer, and it is vital to check the bare or patterned wafer. Contamination of the reticle (mask) is also an important problem that hinders the manufacture of integrated circuits. The reticle quality and reticle defects continue to be a top industry risk [8].

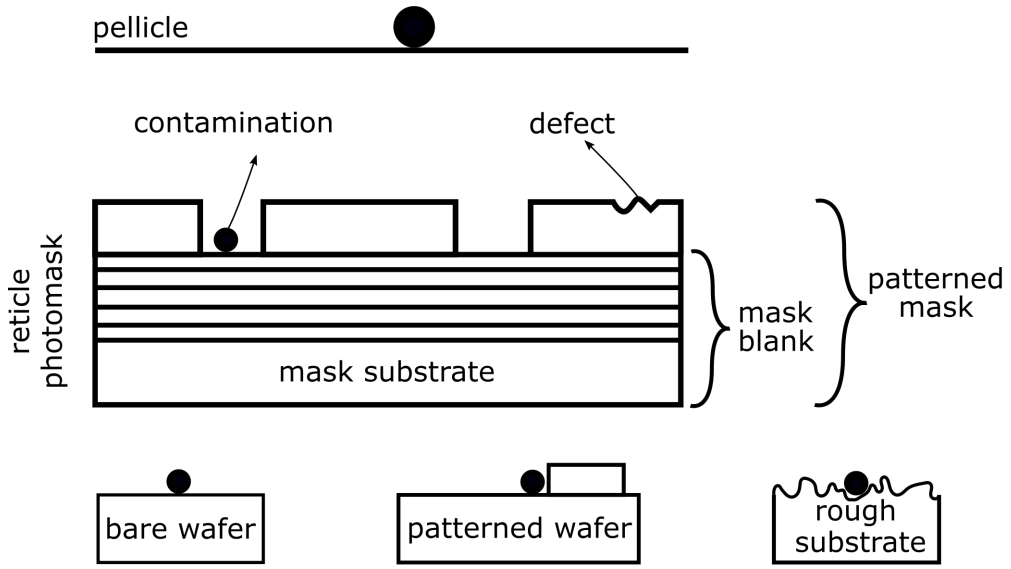


Figure 1.1: Different inspection scenarios of a reticle (mask) for the particle scanner. The fabrication of the reticle photomask includes the intermediate steps wherein the reticle substrate, blank or the final patterned mask is analyzed for the presence of contamination and defects. The pellicle film is intended to protect the mask from additional contamination. The cleanliness of the wafer is analyzed both before and after the mask projection.

The cause of the contamination ranges from the abrasion of the mechanical parts and dust introduced during the assembly to particles induced by the plasma light source and to remainders of the photoresist. To ensure the quality and high yield in semiconductor manufacturing, contamination due to isolated particles in the size range from 20 nm to 1 μm in diameter should be detected and, specifically for high

concentrations, localized and removed.

Before the in-line inspection, contamination due to, for example, isolated nanoparticles was studied using certified-size polystyrene latex spheres (PSLs) that were deposited on the surface of interest in a controlled manner. When analyzing other particles consisting of different materials, the scattering cross-sections can be converted to the Latex Sphere Equivalent (LSE) [9]. From the Rayleigh scattering cross-section [10]:

$$\sigma_s \propto \frac{d^6}{\lambda^4}\left(\frac{m^2-1}{m^2+2}\right)^2. \tag{1.1}$$

where, σ_s represents the amount of scattering, d is the diameter of a sphere, m represents optical contrast of the particle and is equal to $m = \sqrt{n^2 + k^2}$, where n and k are the real and imaginary parts of the refractive index, respectively, and λ is the wavelength of the light. Accordingly, from the perspective of the detector, there is no difference between e.g. 100 nm PSL ($n = 1.57$) and Au 80.8 nm ($n = 1.46, k = 1.95$) diameter sphere at $\lambda = 405$ nm and one can assume that 80.8 nm gold particle in free space has a LSE of 100 nm. Depending on the instrument's collection optics and the wavelength regime, particle scattering on surfaces can be a stronger or weaker function of particle diameter.

Electronic chips are made with the aid of an optical photolithography system. The goal is to transfer the patterns of the future chips from a photomask (reticle) to a silicon wafer coated with photoresist. The estimate of the minimum particle size on the mask which will cause the defects on the wafer can be calculated from [11]:

$$d_{def} = \frac{1}{2} \cdot \frac{CD \cdot M}{E}. \tag{1.2}$$

where the d_{def} is the diameter/size of the defect, CD is the critical dimension, which is defined as the smallest feature that can be patterned, M is the de-magnification factor of the imaging system and $E = \Delta CD_{wafer}/\Delta CD_{mask}$ is the error factor that describes the influence of the imperfections of the mask on the final printed pattern. Error factor is defined as the derivative of the ΔCD_{wafer} at the wafer to that at the reticle ΔCD_{mask} [12]. It depends on feature size, pattern type, exposure parameters, illumination conditions etc. Typical values are usually calculated numerically and fall within a range from 1.5 to 3.5. It is common that the estimate for d_{def} is in the order of the feature size of the pattern CD. For instance, with $M = 4$ and $E = 1.75$ it is critical to eliminate 25 nm or larger particles for the $CD = 22$ nm node. Further, we summarize the estimates for the maximum allowed particle size within the other parts of an Extreme ultraviolet lithography (EUV) system [13–15]:

- EUV reticle load 50 nm.
- The front side of the pellicle (Figure 1.1) $\approx 1\ \mu m$.
- The front side 20 nm and the back side $\approx 1\ \mu m$ of the reticle blank.
- The size of the contamination on bare/patterned wafer is < 20 nm.

1.3. Surface inspection in the semiconductor industry

The semiconductor industry uses metrological tools for product qualification. The main techniques to study nanometer-size features are scanning electron (SEM) and dark and

1

bright field microscopes (Figure 1.2). For electrically conductive materials, surface analysis in the reflection mode is straightforward with SEMs. If the scattering objects are buried inside the structures, transmission electron microscope (TEM) or scanning TEM (STEM) using a beam or a focused spot of electrons can be used [16]. With these techniques, subnanometer resolution can be achieved. However, the implementation of these techniques in the production line can be complicated, and they are slow. In addition, if high beam current and acceleration voltage for the electrons are used, SEM can also produce cracks on the surface or permanent thermal damage.

In order to solve this problem, noninvasive in-line surface inspection, based on optical scatterometry is a powerful option. It relies on illuminating the surface under study with a probing light beam and processing light that is diverted from the surface using a far-field detection system. The contamination can be detected if the scattering produces a sufficient amount of light (photons) on the detector. It is also crucial to scan the entire area because the location of the contamination is essential in product monitoring and root cause analysis.

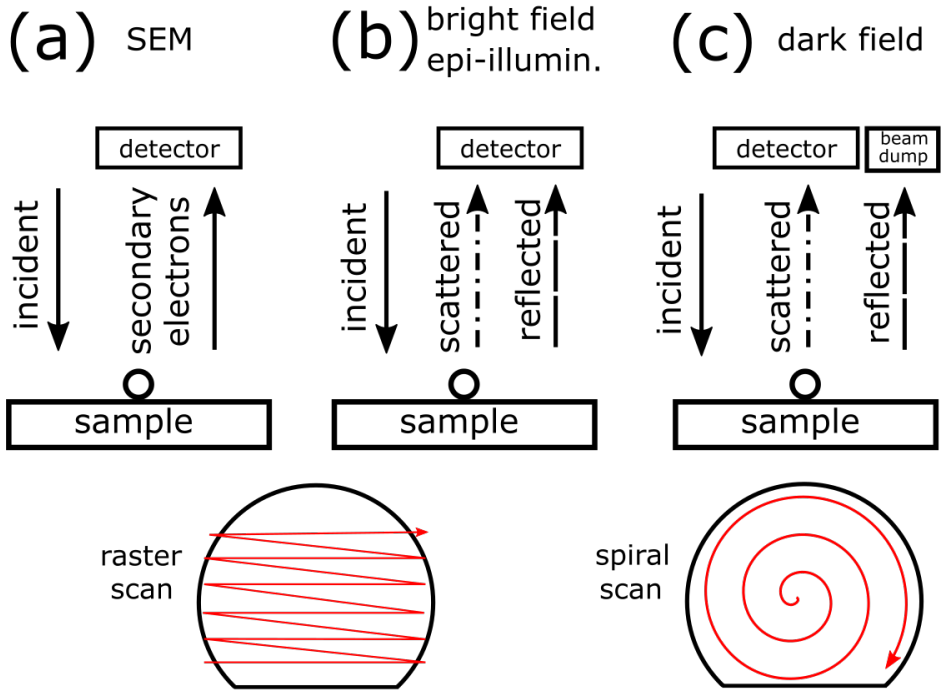


Figure 1.2: (a) Basic principle of the scanning electron microscope (SEM), (b) bright-field epi-illumination, and (c) dark-field microscope. In SEM primary electron beam is raster-scanned over the sample surface where the secondary electrons are generated. An image is created by combining the detector signal (information about topography) and the rastering voltages. In epi-illumination, a light at the detector is a combination of reflected and scattered from the sample while in darkfield microscopy only the scattered light is captured. Commonly, dark-field particle scanners rely on the spiral scanning where the wafer is rotated underneath the beam.

Dark-field techniques, where only the scattered portion of the light is captured, are powerful tools for high-throughput contamination analysis. The state-of-art systems work with bare wafers, smooth and rough films, and deliver defect detection sensitivity aimed at the 7 nm logic and advanced memory device nodes [17, 18]. There are also types of scanners designed for either blank or patterned masks inspection, that aim to achieve the high-resolution, low noise imaging of defects present inside a multi-layer structure (e.g. Mo-Si). These are called actinic, which means they operate at the same wavelength as used for writing the pattern to the wafer [19]. For all techniques in the dark-field family, the direct reflected light is eliminated from the measured field. Consequently, the incident power has to be high in order to produce enough scattering and a sufficient signal-to-noise ratio (SNR). Hence, similar to SEM, in dark-field measurements, there is a a danger of altering or damaging the sample under study due to thermal effects [20].

The so-called bright field technique, where both the reflected and the scattered light from the surface are measured, solves the issue of the sample damaging since it uses very low incident power. However, similarly to dark-field, the small inherent scattering causes too low a SNR. In this context, it is hard to detect tiny particle sizes, with diameters smaller than 200 nm in bright field mode using wavelengths in the visible spectrum. To solve this issue and to allow for the detection of such particle sizes, researchers have proposed various methods including label-free interferometric reflectance imaging sensor (IRIS) or interferometric scattering microscopy (ISCAT) [21] and non-conventional sensing with optical forces in optical pseudo-electrodynamics microscopy (OPEM) [3]. Unfortunately, these approaches suffer from shot noise in the camera, mechanical instability, and low throughput.

Alternatively, one can obtain high sensitivity and low power of the illumination by measuring the light that is scattered from the particles to the far-field in a smart way such that the SNR can be improved as compared to dark-field techniques operating with the same power. Bright field differential detection is the core of the technique, called coherent Fourier scatterometry (CFS) [22] that has been recently applied to particle detection, to achieve a low cost, robust technique, suitable for the detection of low index nanoparticles not only on wafers but also on low contrast substrates such as plastic or glass [23–26]. In this application of CFS, the surface is raster-scanned with the aid of high-NA tightly focused spot. The key-enabling principle is the differential detection of the scattered and reflected fields from the nanoparticle and sample surface, where depending on the position of the particle w.r.t. the focused spot, the scattering is asymmetric, while the bare low roughness surface renders nearly zero signal.

The performance of the surface inspection tools based on the manufacturer's specification as well as on published results, are summarized in Table 1.1. Compared to the existing particle scanners, the method based on CFS was slow and less sensitive. However, at that time ([23], 2014), the detector noise and scanning issues substantially limited the highest sensitivity that could be obtained.

Table 1.1: The initial sensitivity and other parameters of the technique applied to particle detection (by the year 2014) are compared to the state-of-the-art systems for the particle size detection, including the recent ones (by the year 2020). The Latex Sphere Equivalent (LSE) is one of the critical parameters when any scatterometric particle is calibrated against contamination standards of certified-size PSLs. The speed performance of the techniques is estimated based on the settings required to detect 100 nm PSL and to analyze the full area of 200 mm (8-inch) wafer.

Scanner		Sensitivity (LSE)	Source	Inspection target	Time as normalized to 100 [nm] sensitivity to cover 200 [mm] wafer	Reference
Dark-field (conventional)	scientific	35 nm	193 nm	mask blank reticle substrate	2 [hr]	2017 [27, 28]
	industrial	<16 nm	193 nm	bare wafers films (smooth and rough)	3 [m]	2013-2016 [29–31]
Dark-field (actinic)	ABI	33 nm	13.5 nm	mask blank	26 [m]	2013-2020 [19, 32]
	APMI	32 nm	13.5 nm	patterned mask	1.16 [h]	2015-2019 [33–35]
SEM		<5 nm	Accelerating voltage 500 V - 30 kV	patterned wafer	4.36 [hr]	2013 [36, 37]
CFS (single beam)		>100 nm	405 nm	plastic sheet	*9 [m]	2014 [23]

Hence, discovering the limits of the CFS technique in the context of nanoparticle detection is an essential and, according to the author, exciting topic. If the sensitivity and/or speed of the CFS technique is improved, the scatterometer can be integrated into current semiconductor manufacturing processes to facilitate inspection of nanoscale features/unpatterned wafers in-line with lithography processes. The scatterometer can also be used in highly-accurate inspection of polymer webs or pellicle structures. In addition, we envision that this system could be employed in bridging the gap between SEM and other metrology tools such as inline/3D AFM [38–40]. Moreover, it could be explored in new applications such as extending the capabilities of optical microscopy by enabling the analysis of subwavelength bio nanoparticles on surfaces [41, 42].

1.4. Goals of this thesis

For the detection of very small particles of diameter < 100 nm at a fast rate using CFS, it is crucial to optimize the entire system. The goal of this thesis is to boost the sensitivity of CFS, by using an optimal illuminating focused spot in conjunction with optimized detection schemes with both depending on the inspected sample. We have used numerical tools to understand the parameters that could influence the scattering process such as polarization or the surface design. We revisited the way in which the data should be collected and defocusing effects. We also introduced the pre-, post-data process tools and methods to maximize sensitivity for nanoparticle detection on different types of surfaces such as plastic, glass or silicon wafers. Furthermore, we experimentally and numerically established the detection limit of our current system.

*the speed is estimated for $100\ m/s$ spiral scanning and was not optimized in the real system (refers to Table 1.1)

1.5. Outline

In this dissertation, we deal with the following subjects:

- **Chapter 2**: Introduction to electromagnetic theories and numerical modeling. We start by showing the schematics of CFS. Next, we highlight the importance of understanding the polarization and phase effects when focusing light with a high numerical aperture lens. After that, we introduce an analytical model of a dipole at the interface to understand scattering effects that take place in the system. Furthermore, we discuss the principle of finite-difference time-domain (FDTD) and finite element method (FEM) numerical solvers for Maxwell's equations. Finally, we explain the importance of meshing and the geometry for a computational domain.
- **Chapter 3**: Heterodyne detection system for nanoparticle detection. To mitigate the experimental noise, we implement the heterodyne detection system. The underlying principle of heterodyne is explained, and experimentally we have determined the suitable frequencies and waveforms for both modulation and reference signals. The result of this implementation is the first improvement detecting polystyrene nanoparticles.
- **Chapter 4**: Polarization effects in evanescent wave amplification for particle detection. We describe the problem of the particles being index-matched with the substrate, such as in the case of polystyrene particle on top of plastic or glass, analyze the role of the polarization of the input beam (linear, radial and circular), and the coupling mechanism to cavity modes when the surface is coated with a thin film. The studied enhancing mechanism of evanescent wave amplification is accompanied by simulations where both the near-field and far-field gain are considered. We conclude the study by showing the best materials/thickness and polarization of the input light that should be used to optimize the evanescent wave amplification effect.
- **Chapter 5**: Highly-sensitive laser focus positioning method with sub-micrometer accuracy. We analyze the near-field amplitude and phase of the probe interaction with the particle on a surface. We give a detailed description of the novel focus determination approach based on the far-field detection of the scattered light from a particle deposited on the sample surface. The new method is highly sensitive; a two-fold improvement in the slope of an error curve over other methods, such as astigmatism-based methods, is demonstrated.
- **Chapter 6**: Efficient signal processing with the aid of machine learning. It has been shown that it is possible to minimize the noise in the system, set the optimal focusing position, and further, it is beneficial to use linearly polarized light. The process of the detection in the domain less than 100 nm ($< \lambda/4$) as compared to particles of diameter larger than 100 nm ($< \lambda/4$) is associated with the processing of much bigger differential voltage datasets. In this chapter, we develop the framework of pre-processing and semi-automatic feature extraction. Starting from the raw data with the user input for the search criteria, and by modifying the techniques of unsupervised clustering, we can get the particle size distribution histogram with high selectivity between the particle classes.

- **Chapter 7**: Convolutional neural network applied for nanoparticle classification. We rely on the possibilities of convolutional neural networks (CNN) to automatically extract features from the scanned maps to distinguish between the 40, 50, 60, 80 nm particles and the background. We optimize the architecture of the network, to achieve the right balance between the training time and the high-enough accuracy. We show that based on the relatively small set of ≈ 1300 images, we can train the network with two convolutional layers and batch normalization to achieve 95% accurate results. Furthermore, the CNN method illustrates the superior performance in terms of specificity and sensitivity compared to the thresholding and search approach of Chapter 6. We rely on the three methods of the penultimate layer thresholding, such as baseline, mean activation vectors (MAV), and OpenMax to capture the unknown samples. The approach of MAV turns out to be an optimal method where we achieve the best result of capturing 80% of the fooling images at the cost of only dropping the 10% of the original dataset. The main contribution of this chapter is to emphasize that CNN can be exploited as a high-performance analysis tool that is most suitable to treat the big datasets.

- **Chapter 8**: The limits of detection. In the final chapter, we conduct validation experiments on the basis of specially fabricated lithography samples using electron beam technique. We carry out the analysis of the noise due to electronics, surface roughness, and vibration and merge it with a quantitative assessment of low-index particles.

- **Chapter 9**: Conclusion and Outlook. Here, we summarize the thesis and discuss possible future work.

References

[1] Peter deGroot. Optical metrology. *The Optics Encyclopedia: Basic Foundations and Practical Applications*, pages 2085–2117, 2007.

[2] K. Lindfors, T. Kalkbrenner, P. Stoller, and V. Sandoghdar. Detection and spectroscopy of gold nanoparticles using supercontinuum white light confocal microscopy. *Phys. Rev. Lett.*, 93:1–4, Jul 2004.

[3] Jinlong Zhu, Yanan Liu, Xin Yu, Renjie Zhou, Jian Ming Jin, and Lynford L. Goddard. Sensing sub-10 nm wide perturbations in background nanopatterns using optical pseudoelectrodynamics microscopy (opem). *Nano Letters*, 19(8):5347–5355, 8 2019.

[4] Derin Sevenler, Oğuzhan Avci, and M. Selim Ünlü. Quantitative interferometric reflectance imaging for the detection and measurement of biological nanoparticles. *Biomed. Opt. Express*, 8(6):2976–2989, 6 2017.

[5] Jeong-Eun Park, Keunsuk Kim, Yoonjae Jung, Jae-Ho Kim, and Jwa-Min Nam. Metal nanoparticles for virus detection. *ChemNanoMat*, 2(10):927–936, 2016.

[6] Ferdinand Brandl, Nicolas Bertrand, Eliana Martins Lima, and Robert Langer. Nanoparticles with photoinduced precipitation for the extraction of pollutants from water and soil. *Nature Communications*, 6(1):1–10, Jul 2015.

[7] Sayed Mohammad et al. Taghavi. Effects of nanoparticles on the environment and outdoor workplaces. *Electronic physician*, 5(4):706–712, Nov 2013.

[8] William Broadbent Jr. et al. 1X HP EUV reticle inspection with a 193nm inspection system. In Peter D. Buck and Emily E. Gallagher, editors, *Photomask Technology 2017*, volume 10451, pages 149–157. International Society for Optics and Photonics, SPIE, 2018.

[9] R Schmolke, D Gräf, M Suhren, R Kirchner, H Piontek, and P Wagner. Impact of chemical and epitaxial treatment on surface defects on silicon wafers. *MRS Online Proceedings Library Archive*, 442, 1996.

[10] AJ Cox, Alan J DeWeerd, and Jennifer Linden. An experiment to measure mie and rayleigh total scattering cross sections. *American Journal of Physics*, 70(6):620–625, 2002.

[11] Luigi Scaccabarozzi, Niels A. Lammers, Roel Moors, and Vadim Banine. Particle cleaning of euv reticles. *Journal of Adhesion Science and Technology*, 23(12):1603–1622, 2009.

[12] Hye-Young Kang, Sung-Hyuck Kim, Chang-Ho Lee, and Hye-Keun Oh. Mask error enhancement factor variation with pattern density. In J. Tracy Weed and Patrick M. Martin, editors, *25th Annual BACUS Symposium on Photomask Technology*, volume 5992, pages 903 – 910. International Society for Optics and Photonics, SPIE, 2005.

[13] J. K. Stortelder et al. Particle qualification procedure for the TNO EUV reticle load port module of the HamaTech MaskTrackPro cleaning tool. In Bruno M. La Fontaine and Patrick P. Naulleau, editors, *Extreme Ultraviolet (EUV) Lithography II*, volume 7969, pages 597 – 606. International Society for Optics and Photonics, SPIE, 2011.

[14] Carmen Zoldesi et al. Progress on EUV pellicle development. In Obert R. Wood II and Eric M. Panning, editors, *Extreme Ultraviolet (EUV) Lithography V*, volume 9048, pages 430 – 439. International Society for Optics and Photonics, SPIE, 2014.

[15] Rogier Verberk, Norbert Koster, Edwin te Sligte, and Wilbert Staring. Latest developments on EUV reticle and pellicle research and technology at TNO. In Uwe F.W. Behringer and Jo Finders, editors, *33rd European Mask and Lithography Conference*, volume 10446, pages 1 – 8. International Society for Optics and Photonics, SPIE, 2017.

[16] Maria Benelmekki. An introduction to nanoparticles and nanotechnology. In *Designing Hybrid Nanoparticles*, 2053-2571, pages 1–14. Morgan & Claypool Publishers, 2015.

[17] Benjamin Bunday, Eric Solecky, Alok Vaid, A. F. Bello, and Xintuo Dai. Metrology capabilities and needs for 7nm and 5nm logic nodes. In Martha I. Sanchez, editor, *Metrology, Inspection, and Process Control for Microlithography XXXI*, volume 10145, pages 102–142. International Society for Optics and Photonics, SPIE, 2017.

[18] I. Tolle and M. Daino. Advanced defect inspection techniques for nfet and pfet defectivity at 7nm gate poly removal process. In *2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pages 1–4, 2018.

[19] Hiroki Miyai, Tsunehito Kohyama, Tomohiro Suzuki, Kiwamu Takehisa, and Haruhiko Kusunose. Actinic patterned mask defect inspection for EUV lithography. In Jed H. Rankin and Moshe E. Preil, editors, *Photomask Technology 2019*, volume 11148, pages 162 – 170. International Society for Optics and Photonics, SPIE, 2019.

[20] Jacques Dubochet. High resolution dark-field electron microscopy. *Journal of microscopy*, 98(3):334–344, 1973.

[21] J. T. Trueb, O. Avci, D. Sevenler, J. H. Connor, and M. S. Ünlü. Robust visualization and discrimination of nanoparticles by interferometric imaging. *IEEE Journal of Selected Topics in Quantum Electronics*, 23(2), 2017.

[22] D. Kolenov, R. C. Horsten, and S. F. Pereira. Heterodyne detection system for nanoparticle detection using coherent Fourier scatterometry. In Peter Lehmann, Wolfgang Osten, and Armando Albertazzi Gonçalves Jr., editors, *Optical Measurement Systems for Industrial Inspection XI*, volume 11056, pages 336 – 342. International Society for Optics and Photonics, SPIE, 2019.

[23] S. Roy, M. Bouwens, L. Wei, S.F. Pereira, H.P. Urbach, and P. Van Der Walle. High speed low power optical detection of sub-wavelength scatterer. *Review of Scientific Instruments*, 86(12), 2015.

[24] Omar El Gawhary and Stefan J.H. Petra. Method and apparatus for determining structure parameters of microstructures. *US Patent*, Mar 2015.

[25] S. Roy, S. F. Pereira, H. P. Urbach, Xukang Wei, and O. El Gawhary. Exploiting evanescent-wave amplification for subwavelength low-contrast particle detection. *Phys. Rev. A*, 96:1–5, Jul 2017.

[26] S. Roy, A. C. Assafrao, S. F. Pereira, and H. P. Urbach. Coherent fourier scatterometry for detection of nanometer-sized particles on a planar substrate surface. *Opt. Express*, 22(11):13250–13262, Jun 2014.

[27] P. van der Walle et al. Deep sub-wavelength metrology for advanced defect classification. In Peter Lehmann, Wolfgang Osten, and Armando Albertazzi Gonçalves Jr., editors, *Optical Measurement Systems for Industrial Inspection X*, volume 10329, pages 1137 – 1146. International Society for Optics and Photonics, SPIE, 2017.

[28] TNO. Rapid nano particle scanner, 2017. https://tinyurl.com/y8aua9ch.

[29] William Broadbent, Gregg Inderhees, Tetsuya Yamamoto, Isaac Lee, and Phillip Lim. EUV reticle inspection with a 193nm reticle inspector. In Kokoro Kato, editor, *Photomask and Next-Generation Lithography Mask Technology XX*, volume 8701, pages 245 – 255. International Society for Optics and Photonics, SPIE, 2013.

[30] KLA-Tencor. Kla-tencor introduces inspection and review portfolio for leading ic technologies, 2014. https://tinyurl.com/yazpl4ku.

[31] EVG50. Novati technologies selects applied materials to create advanced 200mm line, 2016. https://tinyurl.com/yapl7vld.

[32] GP Naaijkens. *Reticle side wall clamping*. PhD thesis, Technische Universiteit Eindhoven, 2013.

[33] KLA-Tencor. Solutions for euv mask inscpection, 2012. https://tinyurl.com/y9obvs6j.

[34] Ted Liang et al. EUV mask infrastructure and actinic pattern mask inspection. In Nelson M. Felix and Anna Lio, editors, *Extreme Ultraviolet (EUV) Lithography XI*, volume 11323, pages 204 – 216. International Society for Optics and Photonics, SPIE, 2020.

[35] Anna Tchikoulaeva, Hiroki Miyai, Tsunehito Kohyama, Kiwamu Takehisa, and Haruhiko Kusunose. Enabling EUVL high-volume manufacturing with actinic patterned mask inspection. In Nelson M. Felix and Anna Lio, editors, *Extreme Ultraviolet (EUV) Lithography XI*, volume 11323, pages 316 – 320. International Society for Optics and Photonics, SPIE, 2020.

[36] FEI. Nova nanosem 50 series, 2013. https://tinyurl.com/yc79jnxq.

[37] Davies William de Lima Monteiro, O Bonnaud, and N Morimoto. Microelectronics technology and devices–sbmicro 2009. pages 44–47. The Electrochemical Society, 2009.

[38] Hirokazu Koyabu, Ken Murayama, Yukio Kembo, and S Hosaka. In-line atomic force microscope for semiconductor process evaluation. *Hitachi Rev*, 51:130–135, 2002.

[39] Y. Dufrêne. *Life at the nanoscale: Atomic force microscopy of live cells*, chapter 2. Life at the Nanoscale: Atomic Force Microscopy of Live Cells. 2011.

[40] Jan Thiesler, Rainer Tutsch, Karsten Fromm, and Gaoliang Dai. True 3d-afm sensor for nanometrology. *Measurement Science and Technology*, 2020.

[41] Ugur Aygun, Hakan Urey, and Ayca Yalcin Ozkumur. Label-free detection of nanoparticles using depth scanning correlation interferometric microscopy. *Scientific reports*, 9(1):1–8, 2019.

[42] Oguzhan Avci, Maria I. Campana, Celalettin Yurdakul, and M. Selim Ünlü. Pupil function engineering for enhanced nanoparticle visibility in wide-field interferometric microscopy. *Optica*, 4(2):247–254, 02 2017.

2

Introduction to Electromagnetic theories and Numerical Modeling

We present a review of the electromagnetic theory to treat the problem of particle detection with CFS. In Section 2.1, we give the schematics of the CFS experimental setup. In Sections 2.2-2.3, we describe the theory necessary to account for the high numerical aperture of our system and its consequence to the field distribution in the focal region. Next, in Sections 2.4 and 2.5 we develop the relations necessary to analytically account for the interaction between the incident light and the isolated sub-wavelength particle (dipole) at the interface between two media. Finally, in Section 2.6, we describe numerical modeling with FDTD and FEM, and we test the convergence of our template models and explain the computational geometry.

2.1. Coherent Fourier Scatterometry

Coherent Fourier Scatterometry is a relatively young technique based on the scattering of coherent light into the far field of a laser beam that is focused on the sample. It was initially applied to the problem of grating shape reconstruction with the collection of overlapping diffracted orders in the far field [1, 2]. Also, it has been shown that CFS has outstanding capabilities for the detection of nanoparticles on top of silicon, glass, and plastic substrates [3]. Afterwards, by replacing the original CCD camera with a differential detection scheme using a bi-cell, CFS has been modified to better suit the problem of particle detection. As a benefit, effects of spurious reflections from a sample surface were eliminated and a radical increase in the SNR and speed were achieved [4].

The experimental setup is shown in Figure 2.1.

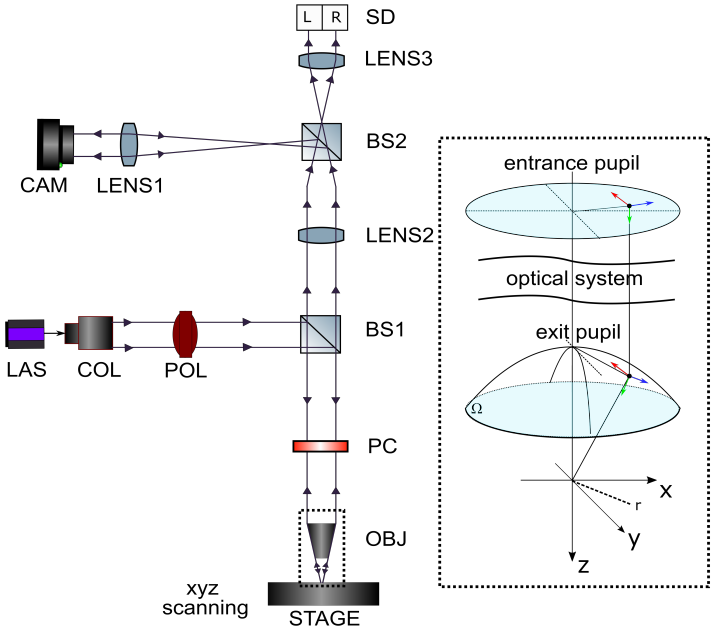


Figure 2.1: A sketch of the experimental setup: STAGE: piezo-electric xyz translator; OBJ: objective lens; PC: on-demand polarizator convertor (zero-order vortex half-wave retarder); BS1, BS2: beam splitters; POL: linear polarizer, COL: light collimator, LENS1, LENS2, LENS3 converging lenses; CAM: CCD camera used for localizing the sample; SD: split detector (bi-cell silicon photodiode), LAS: blue diode laser. A schematic of the objective with entrance and exit pupil as optical system that focuses the collimated beam.

The light source is a blue diode laser (405 nm by Power Technology, model: IQ1A25) that is polarized before it approaches the focusing lens of numerical aperture $NA = 0.9$ (OBJ in the figure). The objective is a non-commercial objective designed for mastering CDs in optical data storage. The focused spot is < 1 μm. The high numeri-

cal aperture allows for the collection of a large number of diffracted angles within the illumination cone of the light (up to 64 degrees w.r.t. optical axis in air). These include the scattered and reflected fields from the nanoparticle and substrate. A telescopic arrangement formed by LENS2 and LENS3 is used to fit the reflected beam into the area of the split detector (SD). Recorded signals from the photodetector are the basis for the signal maps. In our system, SD is bi-cell silicon photodiode. The sample is mounted on a 3D piezo-electric stage whose position can be controlled with sub-nm precision (P-620.ZCD stacked on P-629.2CD by Physik Instrumente). STAGE is programmed to move in raster-scan (introduced earlier in Figure 1.2). As it will be clear later, the state of polarization is crucial to determine the various contributions of the electric field components of the focused spot on the scattering of the isolated particle. In order to generate azimuthal or radial polarization, we add a liquid crystal polarizator converter (PC in the figure, zero-order vortex half wave retarder from Thorlabs) before the objective (OBJ), otherwise the polarization is linear. This converter will only be used in the Chapter 4. Finally, the camera (CAM) is used for localizing the sample.

2.2. Vectorial diffraction integral

For the strong focusing regime (high NA), the widely used scalar theory becomes invalid, because it cannot predict the polarization-induced effects such as the creation of strong longitudinal components or the broadening of the electric field distribution in the focal plane. A vectorial focusing theory is required instead. The focal region description of the field was firstly described by Ignatowsky [5] and further re-derived by Richards and Wolf [6]. In this thesis, we will rely on the calculation steps given in reference [7].

We consider a collimated beam as a monochromatic time-harmonic plane wave, with frequency $\omega > 0$ and electric field $\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) = Re[\boldsymbol{E}(\boldsymbol{r})e^{-i\omega t}]$, by implication, contains time-dependence in the form of $e^{-i\omega t}$. The wave is incident on the lens that is rotationally symmetric and fulfills the Abbe's sine condition $R_{max} = R_f \sin\alpha$ where the maximum radius of the exit pupil R_{max} equals the focal length of the imaging system R_f times the sine of the maximum incidence angle α [8]. Threrefore, the numerical aperture is defined as $NA = n\sin\alpha$, where n is the refractive index of the focusing media. The lens assumed to be aberration-free, causes no absorption or reflection, and the focusing medium is isotropic, homogeneous and nonmagnetic ($\mu = \mu_0$).

Given the vector field distribution in the exit pupil $\boldsymbol{a}(k_x, k_y)$, by integrating over the exit pupil (dk_x, dk_y) (Ω is the solid angle subtended by the aperture of the lens as seen from the focal point), the field in the focal region is given by: $\boldsymbol{E}(\boldsymbol{r})$

$$\boldsymbol{E}(\boldsymbol{r}) = -\frac{i}{2\pi}\int\int_{\Omega}\frac{\boldsymbol{a}(k_x,k_y)}{k_z}\exp\left[i\boldsymbol{k}\cdot\boldsymbol{r}\right]dk_x dk_y. \tag{2.1}$$

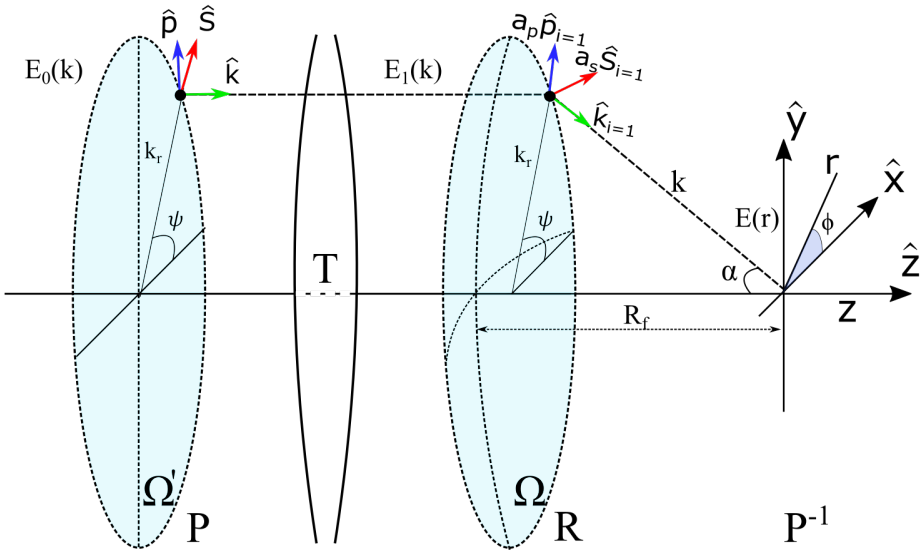


Figure 2.2: Propagation of the field distribution in the entrance pupil $\boldsymbol{a}_0(k_r,\psi)$ via the exit pupil $\boldsymbol{a}_1(k_r,\psi)$ to the focal region $\boldsymbol{a}_1(r)$. The entrance, with the unit vectors $(\hat{\boldsymbol{s}},\hat{\boldsymbol{p}},\hat{\boldsymbol{k}})$, to exit, with the unit vectors $(\hat{\boldsymbol{s}}_1,\hat{\boldsymbol{p}}_1,\hat{\boldsymbol{k}}_1)$, lens transformation with matrix $\boldsymbol{T}$. The observation plane (x,y) with unit vectors $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{z}})$ is in the Gaussian focus with focal length R_f.

The set of cylindrical coordinates in the exit pupil is introduced $\boldsymbol{k}=(k_r,\psi,k_z)$, and in the focal region $\boldsymbol{r}=(r,\phi,z)$ such that:

$$\boldsymbol{k}=k_r\cos\psi\,\hat{\boldsymbol{x}}+k_r\sin\psi\,\hat{\boldsymbol{y}}+k_z\,\hat{\boldsymbol{z}}, \tag{2.2}$$

$$\boldsymbol{r}=r\cos\phi\,\hat{\boldsymbol{x}}+r\sin\phi\,\hat{\boldsymbol{y}}+z\,\hat{\boldsymbol{z}}. \tag{2.3}$$

In order to discretize Eq. 2.1 in cylindrical coordinates a necessary transformation for the elements of integration is given by

$$dk_xdk_y\equiv\begin{pmatrix}\frac{\partial k_x}{\partial k_r} & \frac{\partial k_x}{\partial\psi}\\ \frac{\partial k_y}{\partial k_r} & \frac{\partial k_y}{\partial\psi}\end{pmatrix}dk_rd\psi=k_rdk_rd\psi. \tag{2.4}$$

By expanding the dot product in the exponential term and re-writing the differential, the equation 2.1 becomes:

$$\boldsymbol{E}(\boldsymbol{r})=-\frac{i}{2\pi}\int\int_{\Omega}\frac{\boldsymbol{a}(k_r,\psi)}{k_z}\exp\left[irk_r\cos(\psi-\phi)+ik_zz\right]k_rdk_rd\psi. \tag{2.5}$$

In the aplanatic imaging system, the transition from the entrance pupil, a disc defined by Ω', to the exit pupil, a spherical shell defined by Ω, can be considered as a rotation of the field vector around the angular axis $\hat{\boldsymbol{\phi}}$ in a cylindrical basis $\hat{\boldsymbol{r}},\hat{\boldsymbol{\phi}},\hat{\boldsymbol{z}}$. This lens operation is described by the transfer matrix $\boldsymbol{T}$ (Figure 2.2).

With propagation vector on the cylindrical coordinate frame given by $k_i = (k_{ri}, k_{\phi i}, \pm k_{zi})$, a Fourier transform can always be used $\boldsymbol{E}_i(\boldsymbol{r}) = \boldsymbol{A}\exp\left[\pm i k_i^{\pm}\right]$ to decompose the total field into plane waves and to propagate it to the point of observation. The subscript i indicates the medium in which the vector is defined, while the superscript $\pm$ indicates the direction of propagation. Because the transition from medium i to next medium $i+1$ at position $z = d_i$ should be independent of the coordinates (r, ϕ), plane wave decomposition implies that $k_{ri} = k_{ri+1} = k_r$ and $\psi_i = \psi_{i+1} = \psi$. The propagation vector of monochromatic waves has a fixed length $k_i^2 = k_{zi}^2 + k_r^2$. The sign of the square root that should be taken to obtain the propagation vector is so that the wave decreases exponentially in the direction of propagation (determined by energy conservation).

As mentioned above, the transfer matrix is necessary to take the lens effect into account, i.e. the rotation effect of the lens on the field vector and square root multiplication factor $R_f\sqrt{kz/k}$ (Abbe's sine condition). If the field distribution in the entrance pupil is defined in polar coordinates such that $\boldsymbol{E}_0(k_r, \psi)$, the field distribution in exit pupil is defined by $\boldsymbol{E}_1(k_r, \psi) = \boldsymbol{T} \cdot \boldsymbol{E}_0(k_r, \psi)$. To describe the rotation of field vector three operations are necessary (1) operation $\boldsymbol{P}$ for projection of the Cartesian to cylindrical coordinates, (2) operation $\boldsymbol{R}$ for rotation of the field vectors in cylindrical coordinates, and (3) $\boldsymbol{P}^{-1}$ for inverse of the first operation, i.e. projection from cylindrical to Cartesian coordinates. The matrix operators are given by

$$\boldsymbol{P} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.6}$$

$$\boldsymbol{R} = 1/k_1 \begin{pmatrix} k_{z1} & 0 & k_r \\ 0 & k_1 & 0 \\ -k_r & 0 & k_{z1} \end{pmatrix}, \tag{2.7}$$

$$\boldsymbol{T} = \boldsymbol{P^{-1}RP^1} = \frac{1}{k_1}\begin{pmatrix} k_{z1}\cos^2\psi + k_1\sin^2\psi & (k_{z1}-k_1)\sin\psi\cos\psi & k_r\cos\psi \\ (k_{z1}-k_1)\sin\psi\cos\psi & k_{z1}\sin^2\psi + k_1\cos^2\psi & k_r\sin\psi \\ -k_r\cos\psi & -k_r\sin\psi & k_{z1} \end{pmatrix}. \tag{2.8}$$

where $k_1 = |\boldsymbol{k}_1|$ defines the amplitude of the propagation vector in the first medium. In the experimental scheme, we define a certain polarization state for the light beam (is described in the polar coordinate system), which is transformed by the objective lens and converges to the focal spot (cylindrical coordinate system). The set of unit vectors in the entrance pupil as $(\hat{\boldsymbol{k}}, \hat{\boldsymbol{s}}, \hat{\boldsymbol{p}})$ (red cross blue is green Figure 2.2) given by

$$\hat{\boldsymbol{k}} = \frac{\boldsymbol{k}}{k}, \qquad \hat{\boldsymbol{s}} = \frac{\hat{\boldsymbol{k}} \times \hat{\boldsymbol{z}}}{|\hat{\boldsymbol{k}} \times \hat{\boldsymbol{z}}|}, \qquad \hat{\boldsymbol{p}} = \hat{\boldsymbol{k}} \times \hat{\boldsymbol{s}}. \tag{2.9}$$

The unit vector $\hat{\boldsymbol{k}}$ is parallel, $\hat{\boldsymbol{s}}$ is perpendicular to the direction of wave propagation, and $\hat{\boldsymbol{p}}$ is perpendicular to the plane created by $\hat{\boldsymbol{k}}$ and $\hat{\boldsymbol{s}}$. By performing a scalar multiplication of the field vector (E_x, E_y, E_z) with the three unit vectors below (in the cylindrical coordinate system of exit) one can get the three components of the electric field.

$$\hat{\boldsymbol{k}}_i^{\pm} = \frac{1}{k_i}\begin{pmatrix} k_r\cos\psi \\ k_r\sin\psi \\ \pm k_{iz}\end{pmatrix}, \quad \hat{\boldsymbol{s}}_i^{\pm} = \hat{\boldsymbol{s}}_i = \begin{pmatrix} \sin\psi \\ -\cos\psi \\ 0\end{pmatrix}, \quad \hat{\boldsymbol{p}}_i^{\pm} = \frac{1}{k_i}\begin{pmatrix} \pm k_{iz}\cos\psi \\ \pm k_{iz}\sin\psi \\ -k_r\end{pmatrix}. \tag{2.10}$$

Using the transfer matrix $\boldsymbol{T}$ Eq. 2.8, and, for convenience integrating over the entrance pupil Ω' rather than exit pupil Ω, the Eq. 2.5 now can be written in the final form

$$\boldsymbol{E}_{spot}^{free}(\boldsymbol{r}) = -\frac{iR_f}{2\pi}\int\int_{\Omega'}\sqrt{\frac{k_z}{k}}\frac{\boldsymbol{T}\,\boldsymbol{E}_0(k_r,\psi)}{k_z}\exp\left[irk_r\cos(\psi-\phi)+ik_z z\right]k_r dk_r d\psi. \tag{2.11}$$

The integration/summation is defined over the cone Ω' given by $\psi \in (0, 2\pi]$ and $k_r \in [0, NA\,k_0)$, where the wave number in vacuum given by $k_0 = \omega\sqrt{\epsilon_0\mu_0}$. The integral over the ψ is performed analytically using the integral representation of the Bessel function of the first kind and the integral over the k_r is estimated numerically [7, 9].

In order to compute focal fields for various polarization states at the input, we consider six different possibilities:

Table 2.1: The $E_x(k_r,\psi)$ and $E_y(k_r,\psi)$ coefficients to set the polarization of entrance pupil.

Input polarization	$E_x(k_r,\psi)$	$E_y(k_r,\psi)$
linear x	1	0
linear y	0	1
right circular	$1/\sqrt{2}$	$i/\sqrt{2}$
left circular	$i/\sqrt{2}$	$1/\sqrt{2}$
radial	$\cos\psi$	$\sin\psi$
azimuthal	$-\sin\psi$	$\cos\psi$

We continue by providing intuition for the rotation of polarization induced by the high NA lens.

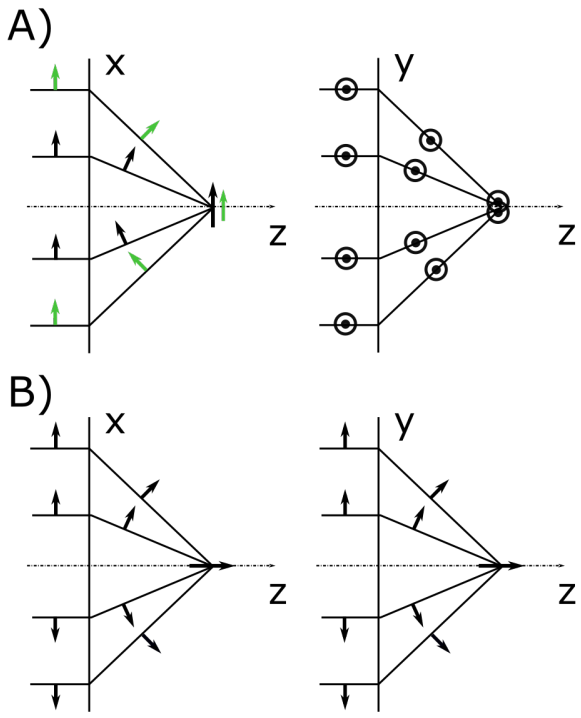


Figure 2.3: Focusing properties with high-NA lenses for different beam input polarization states. The focusing of linearly (x) A) and radially B) polarized light in xz and yz cross-section of the beam.

For linearly polarized light along the x-axis, the rays propagating in the xz- and the yz-plane contribute differently to the focal field (left and right image of Figure 2.3 A). In the xz-plane, we see that the rays that propagate do not add up ideally at the focus. Field component in direction orthogonal to the yz plane does not cancel, and as a consequence, depolarization of the field in focus occurs (i.e., the x linearly polarization at the input is decomposed into x, y and z polarization components in the focal plane). The result is that the intensity field distribution in focus is not cylindrically symmetric anymore but elongated in the x-axis. This has implications for the resolution of the imaging system, as it is well known in microscopy. In contrast, the contributions at focus from the different spatial portions of the radially polarized beam (Figure 2.3 B) constructively interfere to create an intense z-component of the electric field at the focal point. Since the radially polarized input beam is rotation invariant in amplitude and polarization, the field in focus is cylindrically symmetric. Using the calculations according to Eq. 2.11 , we show here the absolute value and phase of the electric field components in the focal plane (xy) for the case $NA = 0.9$ and wavelength $\lambda = 405$ nm.

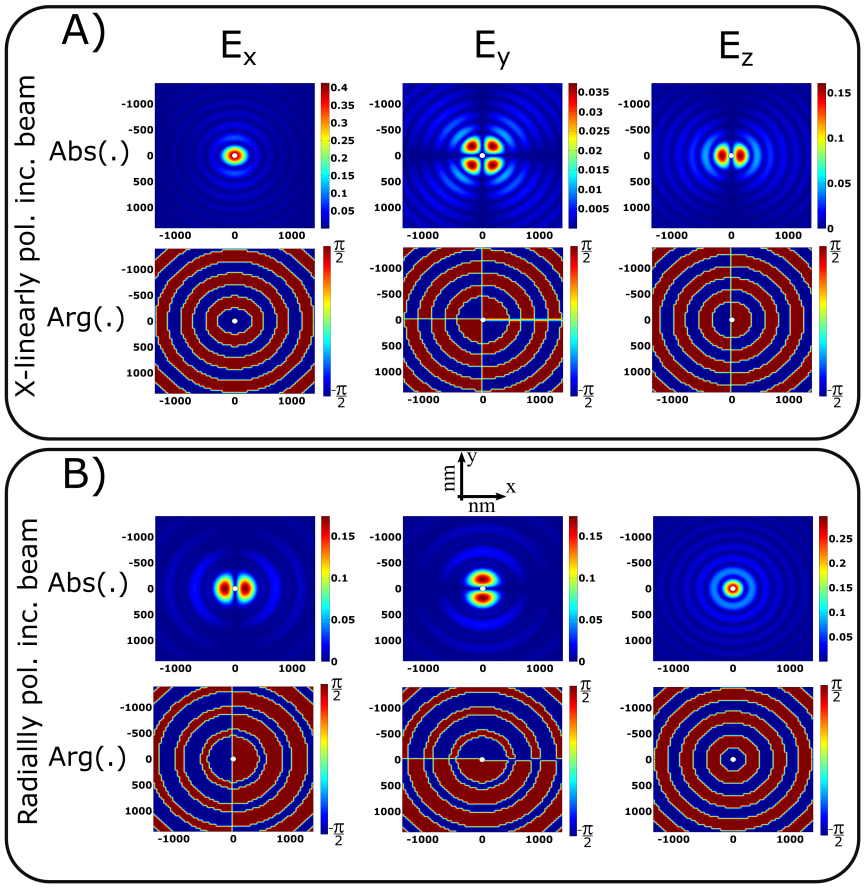


Figure 2.4: The field components of the linearly A) and radially B) polarized incident beam on the focal plane $[nm]$. For each panel, the upper row is absolute value and the bottom row is the phase of the E_x, E_y and E_z component for the case $NA = 0.9$ and wavelength $\lambda = 405$ nm. In the center of each field distribution, the sphere/circle particle of 100 nm is indicated. Note that the colormap is not in the same scale for different field components. For the case of linearly polarized light along x-axis, the component E_x is the strongest, and for radially polarized light, E_z is the strongest.

Figure 2.4 A) for x-linearly polarized beam shows that the $|E_x|$ is elliptically shaped with major axis parallel to x. The $|E_y|$ component has four lobes with maxima along the ± 45 degrees directions. The $|E_z|$ component contains two lobes and contributes to a further broadening of the total electric field in the x-direction. The dominant component is $|E_x|$ with the magnitude ratio of $1 : 0.09 : 0.39$ for $max(|E_x|) : max(|E_y|) : max(|E_z|)$. For the radial polarization (Figure 2.4 B), the $|E_x|$ and $|E_y|$ components of the electric field are the same except for a rotation over 90 degrees, and their sum results in a cylindrically symmetric, doughnut-shaped field distribution. The $|E_z|$ is

rotationally symmetric and dominant where the ratio of the maximum amplitudes of the three components is $0.59 : 0.59 : 1$. The small diameter of $|E_z|$ in focus makes it attractive for applications that require high positional accuracy. The white circle in the figures represents a spherical particle with diameter of 100 nm in the middle of the spot. Scanning is required for the focal field components $|Ey|$, $|Ez|$ of the x-polarized beam and $|Ex|$ and $|Ey|$ focal field components of the radially polarized beam to produce the substantial scattering from the particle.

2.3. Depth of focus and high NA

The high NA of the objective not only causes polarisation effects but also has consequences for the effective depth of focus. From the theory on aberration analysis in the optical system [10], the Rayleigh focal depth is defined by a wavefront deviation of a quarter wavelength at the rim of the exit pupil of the imaging system, between the focused wavefront and the defocused one. The equation for the Rayleigh range is given by

$$\Delta z_R = \frac{\lambda}{4[1-\cos\alpha]}, \tag{2.12}$$

where incidence angle is $\alpha = \arcsin(NA)$. The approximation of $\cos\alpha = \sqrt{1-\sin^2\alpha}$ up to the first order by developing the square root as $\sqrt{1-a} = 1 - a/2 + \ldots$ yields the result $(1-\cos\alpha)^{-1} \approx 2(\sin^2\alpha)^{-1}$. Using this approximated expression gives the common definition of the Rayleigh focal depth

$$\Delta z_R^{approx} = \frac{\lambda}{2\sin^2\alpha}, \tag{2.13}$$

The relative error between the approximated and exact formula can be defined as

$$\delta Z_r = \frac{\Delta z_R^{approx} - \Delta z_R}{\Delta z_R}. \tag{2.14}$$

and is shown in the plot below

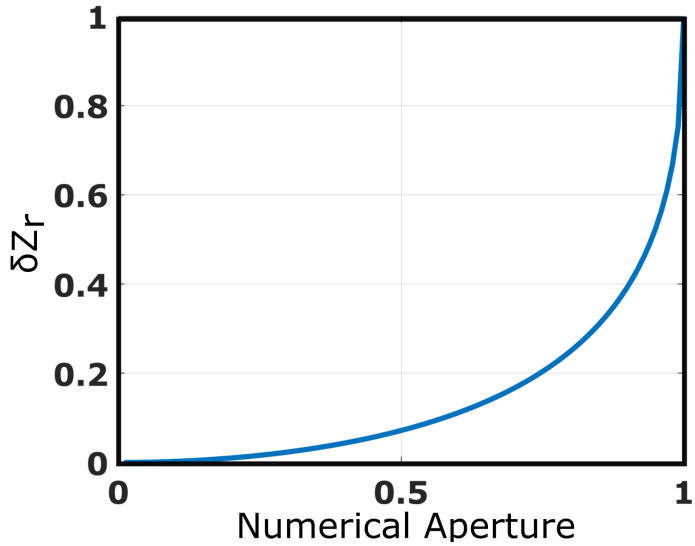


Figure 2.5: The normalized error δZ_r between the two formulas Eq. 2.12 and Eq. 2.13 for the Rayleigh range versus the value of NA

If α is relatively small, for instance smaller than 15 degrees, the error is of the order of a percentage, see Figure 2.5. However, at $\alpha = 30$ degrees ($NA = 0.5$), the error is already 7%. For $NA = 0.9$, the relative error amounts to $\approx 40\%$. The 'exact' formula yields a smaller value for the Rayleigh focal depth than the approximated value. At the limiting maximum aperture of $\alpha = \pi/2$, the error is exactly 100%. It is to some extent arbitrary which definition of the focal length is chosen, however in Chapter 5, when comparing the sensitivity of the proposed focusing method we will rely on the vectorial definition.

2.4. A dipole radiation in free space and its far-field

In this section, as a preparatory step for modelling the interaction of a light beam being focused on an isolated nanoparticle at an interface given in the next section, we show the mathematical model of dipole in a free space. We first refer to Appendix A, where we discuss that similarity exists between the formulas for Fourier transform, angular spectrum representation (plane wave decomposition) and far field computation (propagation to the infinite distance). Finally, we show that for the scatterometer in this thesis, the objective lens performs a Fourier transform up to a multiplicative factor. It is clear thus that in our setup the far-field pattern is revealed at the back focal plane of the lens, and there is no need to experimentally realize a large propagation distance to achieve it. The fourier part in coherent Fourier scatterometry stems from this transformation property of the lens.

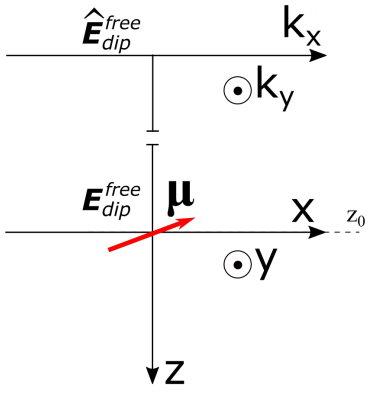


Figure 2.6: The dipole moment $\boldsymbol{\mu}$ radiates at the origin.

Consider a right-handed Cartesian system (x, y, z) with downward propagation direction z. An electric dipole in air with dipole vector $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$ at the origin radiates electric field $\boldsymbol{E}^{free}_{dip}$ (Figure 2.6). Starting from Maxwell's law (Eq. A.19) in free space

$$\boldsymbol{\nabla} \times \boldsymbol{H} = -i\omega\epsilon_0 \boldsymbol{E} + \boldsymbol{J}, \tag{2.15}$$

The current source density $\boldsymbol{J}$ for the dipole moment $\mu(t) = Re[\boldsymbol{\mu} e^{-i\omega t}]$ is given by

$$\boldsymbol{J} = -i\omega\delta(\boldsymbol{r} - \boldsymbol{r}_0)\boldsymbol{\mu}. \tag{2.16}$$

Further, because origin coincides with the position of the dipole moment and $r_0 = (0, 0, 0)$, let's write delta function as $\delta(\boldsymbol{r})$, where δ is the Dirac delta function. By applying the curl operator $\boldsymbol{\nabla}\times$ using Maxwell's equation Appendix Eq. A.15 and A.16, and the identity $\nabla \times \nabla\times = -\nabla^2 + \nabla\nabla\cdot$, the equation 2.15 becomes

$$\omega^2\epsilon_0\mu_0\boldsymbol{H} + \Delta\boldsymbol{H} = i\omega\boldsymbol{\nabla} \times \big[\boldsymbol{\mu}\delta(\boldsymbol{r})\big]. \tag{2.17}$$

The x-component of the RHS of Eq. 2.17 is given by

$$i\omega\, \boldsymbol{i}_x \cdot \boldsymbol{\nabla} \times \boldsymbol{\mu}\delta(\boldsymbol{r}) = i\omega\, \boldsymbol{i}_x \cdot \begin{vmatrix} \boldsymbol{i}_x & \boldsymbol{i}_y & \boldsymbol{i}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \mu_x\delta(\boldsymbol{r}) & \mu_y\delta(\boldsymbol{r}) & \mu_z\delta(\boldsymbol{r}). \end{vmatrix} = i\omega\left(\mu_z \frac{\partial\delta(\boldsymbol{r})}{\partial y} - \mu_y \frac{\partial\delta(\boldsymbol{r})}{\partial z}\right). \tag{2.18}$$

With $k = \omega/c_0 = \omega\sqrt{\epsilon_0\mu_0}$, one can notice that Eq. 2.17 is the inhomogeneous Helmholtz equation similar to Appendix Eq. A.26. Furthermore, the Green's function G, given by $G = -e^{ikr}/4\pi r$, of the scalar Helmholtz equation satisfies:

$$[\omega^2\epsilon_0\mu_0 + \Delta]G = \delta(\boldsymbol{r}). \tag{2.19}$$

By comparing the two equations above, and the radiation condition for $r \to \infty$, we see that the x-component of Eq. 2.17 is satisfied if the x-component of the magnetic field

strength is given by:

$$H_x(\boldsymbol{r}) = i\omega\left(\mu_z \frac{\partial G(\boldsymbol{r})}{\partial y} - \mu_y \frac{\partial G(\boldsymbol{r})}{\partial z}\right), \tag{2.20}$$

The other components should be handled in the same manner, so that Eq. 2.17 is satisfied with

$$\boldsymbol{H}(\boldsymbol{r}) = i\omega\boldsymbol{\nabla} \times \boldsymbol{\mu} G(\boldsymbol{r}). \tag{2.21}$$

Re-arranging Eq. 2.15, using expression Eq. 2.16 for dipole current density gives:

$$\boldsymbol{E} = -\frac{\boldsymbol{\nabla} \times \boldsymbol{H} + \boldsymbol{J}}{i\omega\epsilon_0} = \frac{i}{\omega\epsilon_0}\boldsymbol{\nabla} \times \boldsymbol{H} - \frac{\boldsymbol{\mu}}{\epsilon_0}\delta(\boldsymbol{r}). \tag{2.22}$$

Substituting Eq. 2.21 into Eq.2.22 gives

$$\begin{aligned}\boldsymbol{E}_{dip}^{free}(\boldsymbol{r}) &= -\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \left(\frac{\boldsymbol{\mu}}{\epsilon_0} G(\boldsymbol{r})\right) - \frac{\boldsymbol{\mu}}{\epsilon_0}\delta(\boldsymbol{r}) = \Delta\left(\frac{\boldsymbol{\mu}}{\epsilon_0} G(\boldsymbol{r})\right) - \boldsymbol{\nabla}\boldsymbol{\nabla} \cdot \left(\frac{\boldsymbol{\mu}}{\epsilon_0} G(\boldsymbol{r})\right) - \frac{\boldsymbol{\mu}}{\epsilon_0}\delta(\boldsymbol{r}) \\ &= -\frac{k^2}{\epsilon_0}\boldsymbol{\mu} G(\boldsymbol{r}) - \boldsymbol{\nabla}\boldsymbol{\nabla} \cdot \left(\frac{\boldsymbol{\mu}}{\epsilon_0} G(\boldsymbol{r})\right).\end{aligned} \tag{2.23}$$

The above equation represents the field of radiation of an electric dipole in air located at $z = z_0 = 0$. To compute the far-field radiation, we will require the 2D Fourier transform with respect to the (x, y)

$$\mathcal{F}[\boldsymbol{E}](k_x, k_y, z_0) = \frac{1}{(2\pi)^2} \iint e^{-i[k_x x + k_y y]} \boldsymbol{E}(x, y, z_0) dx dy, \tag{2.24}$$

where the wave vector is defined as $\boldsymbol{k} = k_x\boldsymbol{x} + k_y\boldsymbol{y} + k_z\boldsymbol{z}$. Importantly, let's also introduce the inverse 2D Fourier transform because it is essential for plane wave expansion.

$$\boldsymbol{E}(x, y, z_0) = \iint e^{i[k_x x + k_y y]} \mathcal{F}[\boldsymbol{E}](k_x, k_y, z_0) dk_x dk_y. \tag{2.25}$$

N.B. When performed on a uniform spatial grid, the forward and inverse Fourier transforms are performed with fast Fourier transform (FFT). Hence, the forward 2D Fourier transform with respect to the (x,y) variables of Eq. 2.19 is performed

$$(\omega^2\epsilon_0\mu_0 - k_x^2 - k_y^2)\mathcal{F}(G)\left(\frac{k_x}{2\pi}, \frac{k_y}{2\pi}, z\right) + \frac{d^2}{dz^2}\mathcal{F}(G)\left(\frac{k_x}{2\pi}, \frac{k_y}{2\pi}, z\right) = \delta(z). \tag{2.26}$$

Solving separately for the field in the upper $z < 0$ and lower $z > 0$ halfspace, with wave vector's z-components $k_z^- = -\sqrt{(k^2n^2 - k_x^2 - k_y^2)}$ and $k_z^+ = +\sqrt{(k^2n^2 - k_x^2 - k_y^2)}$ correspondingly, implies

$$\mathcal{F}(G)\left(\frac{k_x}{2\pi}, \frac{k_y}{2\pi}, z\right) = A(k_x, k_y)e^{-ik_z z}, \tag{2.27}$$

$$\mathcal{F}(G)\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right)=B(k_x,k_y)e^{ik_z z}. \tag{2.28}$$

where A and B are functions that are still to be determined

The Green's function in Eqs. 2.27 and 2.28 must be continuous and the derivative with respect to z must take a unit jump for $z=0$. Imposing these conditions implies

$$A(k_x,k_y)=B(k_x,k_y)=1/2ik_z, \tag{2.29}$$

and therefore,

$$\mathcal{F}(G)\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right)=\frac{e^{\pm ik_z z}}{2ik_z}. \tag{2.30}$$

$+$ for $z>0$ and $-$ for $z<0$. Note that, the z derivative of Eq. 2.30

$$\frac{d}{dz}\mathcal{F}(G)\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right)=\text{sgn}(z)ik_z\frac{e^{\pm ik_z z}}{2ik_z}=ik_z\text{sgn}(z)\mathcal{F}(G)\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right), \tag{2.31}$$

where sign function is defined by

$$\text{sgn}(z)=\begin{cases}1, & if \quad z>0,\\ -1, & \text{otherwise.}\end{cases} \tag{2.32}$$

With substitution of $k_x\hat{\boldsymbol{x}}+k_y\hat{\boldsymbol{y}}+\frac{d}{dz}\hat{\boldsymbol{z}}$ into the formula 2.23 the for electric field strength and performing the Fourier transform with respect to x and y gives

$$\begin{aligned}\mathcal{F}(\mathbf{E}_{dip}^{free})\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right)=\Bigg[&-\omega^2\epsilon_0\mu_0\boldsymbol{\mu}\\ &+[k_x\mu_x+k_y\mu_y+k_z\mu_z\text{sgn}(z)]\begin{pmatrix}k_x\\ k_y\\ k_z\text{sgn}(z)\end{pmatrix}\Bigg]\frac{e^{\pm ik_z z}}{2i\epsilon_0 k_z}.\end{aligned} \tag{2.33}$$

The compact version of Eq. 2.33 above is given by,

$$\hat{\boldsymbol{E}}_{dip}^{free}\left(\frac{k_x}{2\pi},\frac{k_y}{2\pi},z\right)=-\left(k^2\boldsymbol{\mu}-(\boldsymbol{k}.\boldsymbol{\mu})\boldsymbol{k}\right)\frac{e^{\pm ik_z z}}{2i\epsilon_0 k_z}. \tag{2.34}$$

2.5. Field in the detector for a dipole at an interface excited by the focused spot

The goal of this section is to account for the interaction of the focused spot with the dipole moment positioned close to the surface. We assume now that the substrate is flat. Then the reflected field can be derived by considering the reflection by the substrate of every individual plane wave in the angular spectrum decomposition of

the field radiated by the dipole. All outgoing electrical field components (Figure 2.7) are combined in the reference plane $(x_0, y_0; z_0 = 0)$, in our case, the focal plane of the lens that coincides with the centre of the particle (dipole moment), and Fourier transforming it to achieve the detector plane.

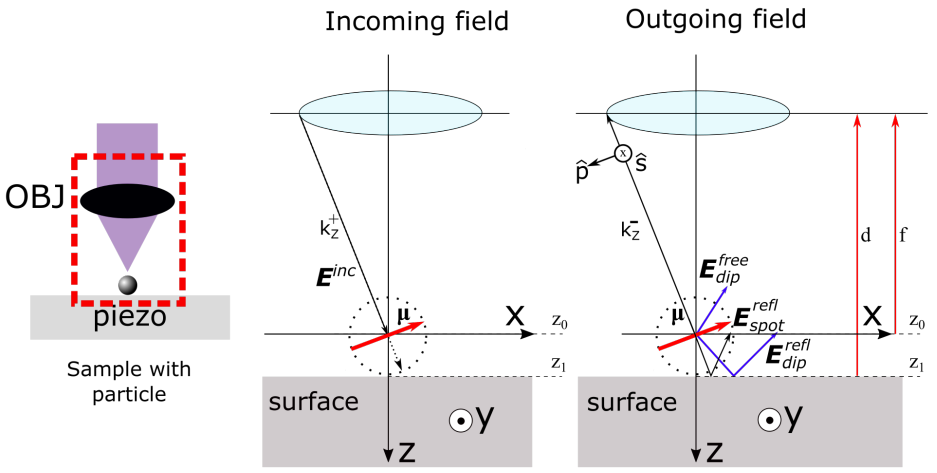


Figure 2.7: The schematic of collimated light beam is focused by a high numerical aperture objective (OBJ) onto an interface containing an isolated nanoparticle. The dipole moment $\boldsymbol{\mu}$ is located in the geometrical focus of the objective lens and it is excited by the $\boldsymbol{E}^{inc}$ focal field. The outgoing scattered field consists of focal field reflected by the surface $\boldsymbol{E}^{refl}_{spot}$, free-space radiation of a dipole $\boldsymbol{E}^{free}_{dip}$, and field of a dipole reflected by the surface $\boldsymbol{E}^{refl}_{dip}$.

Let's define the background incident field as $\boldsymbol{E}^{inc}(\boldsymbol{r})$ that consists of two components: focal field in free space $\boldsymbol{E}^{free}_{spot}(\boldsymbol{r})$, which can be given in the origin plane above the surface $z = z_0$ (derived in cylindrical coordinates Eq. 2.11) and focal field reflected from the substrate $\boldsymbol{E}^{refl}_{spot}(\boldsymbol{r})$ in the absence of the dipole moment.

$$\boldsymbol{E}^{inc}(\boldsymbol{r}) = \boldsymbol{E}^{free}_{spot}(\boldsymbol{r}) + \boldsymbol{E}^{refl}_{spot}(\boldsymbol{r}). \tag{2.35}$$

In order to compute the second term of Eq. 2.35 the defocused field at the surface must be multiplied with Fresnel reflection coefficients and propagated back to the reference plane. In fact, the transformation matrix of Eq. 2.11 can be written to support the s and p - components according to

$$\boldsymbol{E}^{free}_{spot}(\boldsymbol{r}) = -\frac{iR_f}{2\pi}\int\int_{\Omega'}\frac{1}{\sqrt{kk_z}}e^{-ik_z z}\left[\boldsymbol{T}^{s-} + \boldsymbol{T}^{p-}\right]\cdot\boldsymbol{E}_0(k_r,\psi)\; e^{irk_r\cos(\psi-\phi)}k_r dk_r d\psi. \tag{2.36}$$

transformation matrix $\boldsymbol{T}$ is replaced with sum of two matrices separately for s and

p-components [7] (orthogonal basis perpendicular to the wave vector), defined as

$$\boldsymbol{T}^{s-} = r_s(k_x, k_y)\begin{pmatrix} \sin^2 k_\phi & -\sin k_\psi \cos k_\psi & 0 \\ -\sin k_\psi \cos k_\psi & \cos^2 k_\psi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.37}$$

$$\boldsymbol{T}^{p-} = \frac{r_p(k_x, k_y)}{k}\begin{pmatrix} -k_z \cos^2 k_\psi & -k_z \cos k_\psi \sin k_\psi & 0 \\ -k_z \cos k_\psi \sin k_\psi & -k_z \sin^2 k_\psi & 0 \\ -k_r \cos k_\psi & -k_r \sin k_\psi & 0 \end{pmatrix}. \tag{2.38}$$

where for non-magnetic materials $\mu_1 = \mu_2 = 1$, the Fresnel reflection coefficients are

$$r_s(k_x, k_y) = \frac{k_{z_1} - k_{z_2}}{k_{z_1} + k_{z_2}}, \qquad r_p(k_x, k_y) = \frac{\epsilon_2 k_{z_1} - \epsilon_1 k_{z_2}}{\epsilon_2 k_{z_1} + \epsilon_1 k_{z_2}}. \tag{2.39}$$

The k_x, k_y dependency is due to longitudinal vector is given by $k_{z_i} = \sqrt{k_i^2 - (k_x^2 + k_y^2)}$.

Because the position of the surface under the spherical particle is further in the positive z-direction, the propagation of $\boldsymbol{E}_{spot}^{free}(\boldsymbol{r})$ is required to represent the spot field at z_1. This is achieved with plane wave decomposition method

$$\boldsymbol{E}(x, y, z) = \iint \hat{\boldsymbol{E}}\left(k_x, k_y, z_0\right) \exp\left[i(k_x x + k_y y \pm k_z z)\right] dk_x dk_y \tag{2.40}$$

where a plus sign is for propagation downwards and minus sign is for propagation upwards. With $x = x_0$, $y = y_0$ and $z = z_0$ the defocused to the surface free-space focal filed is defined as

$$\boldsymbol{E}_{spot}^{free}(x, y, z_1) = \iint \hat{\boldsymbol{E}}_{spot}^{free}(k_x, k_y, z) \exp\left[i(k_x x + k_y y + k_z(z_1 - z))\right] dk_x dk_y \tag{2.41}$$

modified by the Fresnel reflection coefficients

$$\hat{\boldsymbol{E}}_{spot}^{refl}(k_x, k_y, z_1) = r_{s/p}(k_x, k_y)\, \hat{\boldsymbol{E}}_{spot}^{free}(k_x, k_y, z_1) \tag{2.42}$$

it is further propagated back to the focal plane

$$\boldsymbol{E}_{spot}^{refl}(\boldsymbol{r}) = \iint \hat{\boldsymbol{E}}_{spot}^{refl}(k_x, k_y, z_1) \exp\left[i(k_x x + k_y y + k_z(z - z_1))\right] dk_x dk_y \tag{2.43}$$

Further, let's introduce a dipole at $\boldsymbol{r}_0$ with a dipole moment $\boldsymbol{\mu}$ that also has two components: the free space radiation of dipole $\boldsymbol{E}_{dip}^{free}(\boldsymbol{r})$ and the radiation of the dipole reflected by the surface $\boldsymbol{E}_{dip}^{refl}(\boldsymbol{r})$ given by

$$\boldsymbol{E}_{dip}(\boldsymbol{r}) = \boldsymbol{E}_{dip}^{free}(\boldsymbol{r}) + \boldsymbol{E}_{dip}^{refl}(\boldsymbol{r}). \tag{2.44}$$

Notice that there is no excitation field interacting with the dipole so far. The alternative to Eq.2.23 way to represent the field of a dipole in a homogeneous medium, given in Reference [11], is

$$\boldsymbol{E}_{dip}^{free}(\boldsymbol{r}) = \omega^2\mu_0\underline{\underline{\boldsymbol{G}_0}}(\boldsymbol{r},\boldsymbol{r_0})\boldsymbol{\mu}, \tag{2.45}$$

here $\underline{\underline{\boldsymbol{G}_0}}(\boldsymbol{r},\boldsymbol{r_0})$ is a Green's tensor given by

$$\underline{\underline{\boldsymbol{G}_0}}(\boldsymbol{r},\boldsymbol{r_0}) = \frac{i}{8\pi^2}\iint \underline{\underline{\boldsymbol{M}}}\exp\left[ik_x(x-x_0)+ik_y(y-y_0)+ik_z|z-z_0|\right]dk_xdk_y, \tag{2.46}$$

with expression for second rank tensor and the dipole position above the surface

$$\underline{\underline{\boldsymbol{M}}} = \frac{1}{k^2k_z}\begin{bmatrix} k^2-k_x^2 & -k_xk_y & k_xk_z \\ -k_xk_y & k^2-k_y^2 & k_yk_z \\ k_xk_z & k_yk_z & k^2-k_z^2 \end{bmatrix}. \tag{2.47}$$

The propagation and reflection of the dipole field must be defined in the same way that the reflected focal field was. Firstly, the propagation of the free-space dipole field to the surface plane.

$$\underline{\underline{\boldsymbol{G_0}}}(x,y,z_1) = \iint \underline{\underline{\hat{\boldsymbol{G}}_0}}(k_x,k_y,z_0)\exp\left[i(k_xx_0+k_yy_0+k_z(z_1-z_0))\right]dk_xdk_y, \tag{2.48}$$

$$\hat{\boldsymbol{E}}_{dip}^{free}(k_x,k_y,z_1) = \omega^2\mu_0\underline{\underline{\hat{\boldsymbol{G}}_0}}(k_x,k_y,z_1)e^{i\boldsymbol{k}^-\cdot\boldsymbol{r_0}}\boldsymbol{\mu}, \tag{2.49}$$

$$\hat{\boldsymbol{E}}_{dip}^{refl}(k_x,k_y,z_1) = r_{s/p}(k_x,k_y)\,\hat{\boldsymbol{E}}_{dip}^{free}(k_x,k_y,z_1). \tag{2.50}$$

Secondly, propagation of the reflected Green's tensor to the focal plane.

$$\begin{aligned}\boldsymbol{E}_{dip}^{refl}(\boldsymbol{r}) &= \iint \hat{\boldsymbol{E}}_{dip}^{refl}(k_x,k_y,z_1)\exp\left[i(k_xz_0+k_yz_0+k_z(z_0-z_1))\right]dk_xdk_y \\ &= \omega^2\mu_0\iint r_{s/p}(k_x,k_y)\,\underline{\underline{\hat{\boldsymbol{G}}_0}}(k_x,k_y,z_1)e^{i\boldsymbol{k}^-\cdot\boldsymbol{r_0}}e^{i(k_xz_0+k_yz_0+k_z(z_0-z_1))}\,dk_xdk_y\,\boldsymbol{\mu} \\ &= \omega^2\mu_0\underline{\underline{\boldsymbol{G}}}^{refl}(\boldsymbol{r},\boldsymbol{r_0})\,\boldsymbol{\mu}\end{aligned} \tag{2.51}$$

So far the dipole moment $\boldsymbol{\mu}$ was assumed to be known, however the appropriate formalism is needed to account for both the total excitation field $\boldsymbol{E}^{exc}(r) = \boldsymbol{E}^{inc}(\boldsymbol{r}) + \boldsymbol{E}_{dip}^{refl}(\boldsymbol{r})$ and the interaction of the dipole moment with itself. A reasonable approximation of the dipole moment $\boldsymbol{\mu}$ is to assume that it is proportional to the local excitation electric field $\boldsymbol{E}^{exc}(r)$ at the centre of the nano-particle with the proportionality factor given by the complex polarizability α

$$\boldsymbol{\mu} = \alpha\boldsymbol{E}^{inc}(\boldsymbol{r_0}) + \alpha\boldsymbol{E}_{dip}^{refl}(x,y,0). \tag{2.52}$$

where, according to Reference [12], α has the following relation with the material ϵ, radius r, and surrounding of the particle ϵ_m.

$$\alpha = 4\pi\epsilon_m r^3\frac{\epsilon-\epsilon_m}{\epsilon+2\epsilon_m}. \tag{2.53}$$

We can re-write the second term of Eq. 2.52 by using Eq. 2.51

$$\boldsymbol{\mu} = \alpha \boldsymbol{E}^{inc}(\boldsymbol{r_0}) + \alpha\omega^2\mu_0 \underline{\underline{\boldsymbol{G}}}^{refl}(x, y, 0)\,\boldsymbol{\mu}, \tag{2.54}$$

Re-arranging terms and solving for $\boldsymbol{\mu}$

$$\boldsymbol{\mu} = \left[\underline{\underline{\boldsymbol{I}}} - \alpha\omega^2\mu_0 \underline{\underline{\boldsymbol{G}}}^{refl}(x, y, 0)\right]^{-1} \alpha \boldsymbol{E}^{inc}(\boldsymbol{r_0}). \tag{2.55}$$

Finally the total field at the detector is written as

$$\begin{aligned} \hat{\boldsymbol{E}}^{tot}_{det}(k_x, k_y) &= \hat{\boldsymbol{E}}^{refl}_{spot}(k_x, k_y, 0) + \hat{\boldsymbol{E}}^{free}_{dip}(k_x, k_y, 0) + \hat{\boldsymbol{E}}^{refl}_{dip}(k_x, k_y, 0) \\ = \hat{\boldsymbol{E}}^{refl}_{spot}(k_x, k_y, 0) &+ \omega^2\mu_0 \underline{\underline{\hat{\boldsymbol{G}}_{\boldsymbol{0}}}}(k_x, k_y, 0)\underline{\underline{\boldsymbol{\mu}}} + \omega^2\mu_0 \underline{\underline{\hat{\boldsymbol{G}}_{\boldsymbol{0}}}}^{refl}(k_x, k_y, z_1) e^{-ik_z z_1} \underline{\underline{\boldsymbol{\mu}}} \end{aligned} \tag{2.56}$$

One can notice that Green's tensor Eq. 2.46, contains singularity at $\boldsymbol{r} = \boldsymbol{r_0}$, that is when $|r| = 0$. It thus not straightforward to compute the field exactly at the position of the dipole moment. It has been shown that singularity could be excluded [11] or the formula could be estimated numerically [13]. Further, Eq. 2.11 is defined in the cylindrical coordinates and cannot be directly applied in the formula for the coupled dipole moment Eq. 2.55. The possible solution is to interpolate complex field data from polar to Cartesian coordinates. It is convenient to compute the first term of Eq. 2.56 by using fast fourier transform (FFT) algorithm defined on the Cartesian grid.

2.6. Numerical modeling

The analytical model of a dipole, as shown in the previous section, gives us insight into the problem we are dealing with. But, to make the simulation more realistic, one should consider the effects of finite size and the shape of the particle. Also effects due to thin layers that could be present on the surface should be considered. The latter is possible using a straightforward extension of the analytical model to a multilayer substrate provided that all interfaces are straight and flat. However, when some interfaces are not flat, a numerical simulation is required.

2.6.1. FDTD

One of the two numerical methods that has been used in this thesis to solve the Maxwell's equations is the "Finite-difference time-domain" (FDTD) which abbreviation coined by Allen Taflove [14]. It can be considered as the method to create "movies" of the electromagnetic field. The computational volume is divided by the square/cube grid where the Maxwell's time-dependent differential equations are solved by updating the electric $\mathcal{E}$ and magnetic fields $\mathcal{H}$. To get started, it is necessary to have the Maxwell's equations A.1b and A.1a Appendix A written in the finite-difference approximation from which, we get the the update equations for the fields. Below, in Eq. 2.57 left column is the derivation of the update for $\mathcal{H}$ and the right column is the derivation of the update

for $\mathcal{E}$:

$$\begin{aligned}
\nabla \times \mathcal{E} &= -\frac{\partial \mathcal{B}}{\partial t} & \nabla \times \mathcal{H} &= \epsilon \frac{\partial \mathcal{E}}{\partial t} \\
\frac{\partial \mathcal{H}}{\partial t} &= -\frac{1}{\mu}\nabla \times \mathcal{E} & \frac{\partial \mathcal{E}}{\partial t} &= \frac{1}{\epsilon}\nabla \times \mathcal{H} \\
\frac{\mathcal{H}\left(t+\frac{\Delta t}{2}\right)-\mathcal{H}\left(t-\frac{\Delta t}{2}\right)}{\Delta t} &= -\frac{1}{\mu}\nabla \times \mathcal{E}(t) & \frac{\mathcal{E}(t+\Delta t)-\mathcal{E}(t)}{\Delta t} &= \frac{1}{\epsilon}\nabla \times \mathcal{H}\left(t+\frac{\Delta t}{2}\right) \\
\mathcal{H}\left(t+\frac{\Delta t}{2}\right) &= \mathcal{H}\left(t-\frac{\Delta t}{2}\right) - \frac{\Delta t}{\mu}\nabla \times \mathcal{E}(t) & \mathcal{E}\left(t+\Delta t\right) &= \mathcal{E}(t) + \frac{\Delta t}{\epsilon}\nabla \times \mathcal{H}\left(t+\frac{\Delta t}{2}\right)
\end{aligned} \tag{2.57}$$

Notice that the right-hand side (RHS) of the last equations from Eq. 2.57 contains the values computed at the previous time step. The engine of the FDTD is described in the following diagram, starting from the top left corner:

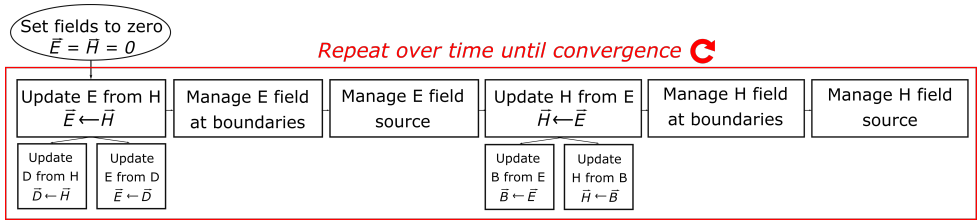


Figure 2.8: The block diagram of the generic finite difference time-domain (FDTD) algorithm. The explicit vector notation is applied for visibility instead of the bold symbol.

In Figure 2.8, the algorithm starts with assigning zero values to $\mathcal{E}$ and $\mathcal{H}$. Within the loop, the following steps occur: update $\mathcal{E} \leftarrow \mathcal{H}$ and $\mathcal{H} \leftarrow \mathcal{E}$, introduction of the source via overwriting the field values at specific locations on the grid, waves that approach the simulation boundary go to infinity and not being reflected from the edges at the volume. It is further possible to incorporate the materials properties into the simulation by using the constitutive relations for $\mathcal{D}$ and $\mathcal{B}$ (Appendix A Eq. A.2 and Eq. A.4). The equations that account for the material properties, for a diagonally anisotropic material such as occurs in the perfectly matched layer (PML), can be written as

$$\begin{aligned}
\frac{\partial \mathcal{E}_z}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial z} &= -\frac{\mu_{xx}}{c_0}\frac{\partial \tilde{\mathcal{H}}_z}{\partial t} & \frac{\partial \tilde{\mathcal{H}}_z}{\partial y} - \frac{\partial \tilde{\mathcal{H}}_y}{\partial z} &= \frac{\epsilon_{xx}}{c_0}\frac{\partial \mathcal{E}_x}{\partial t} \\
\frac{\partial \mathcal{E}_x}{\partial z} - \frac{\partial \mathcal{E}_z}{\partial x} &= -\frac{\mu_{yy}}{c_0}\frac{\partial \tilde{\mathcal{H}}_y}{\partial t} & \frac{\partial \tilde{\mathcal{H}}_x}{\partial z} - \frac{\partial \tilde{\mathcal{H}}_z}{\partial x} &= \frac{\epsilon_{yy}}{c_0}\frac{\partial \mathcal{E}_y}{\partial t} \\
\frac{\partial \mathcal{E}_y}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial y} &= -\frac{\mu_{zz}}{c_0}\frac{\partial \tilde{\mathcal{H}}_z}{\partial t} & \frac{\partial \tilde{\mathcal{H}}_y}{\partial x} - \frac{\partial \tilde{\mathcal{H}}_x}{\partial y} &= \frac{\epsilon_{zz}}{c_0}\frac{\partial \mathcal{E}_z}{\partial t}
\end{aligned} \tag{2.58}$$

The tilde represents the normalization of the field values, such that there is the same order of magnitude for the electric and magnetic field. The group of equations 2.58 can also be extended to have e.g. current density source $\mathcal{J}$ such that time-dependent differential equations are solved in the presence of the source that is launched at the

$t = 0$ [10]. Also, the crucial part of algorithm is to re-write the differential equations 2.58 on the cells of the so called Yee grid [15]. The criteria for the algorithm to stop (reach the steady-state) is if the difference between maximum and minimum field values on the "lattice" is small for time steps differing by one period of the light wave. The algorithm becomes unstable if a sufficiently small time step is not taken. At the output of the algorithm, the time domain impulse response of the system can be Fourier transformed to get the frequency response.

2.6.2. FEM

The second numerical method used in this thesis is the finite element method (FEM). It is particularly suited to solve boundary value problems and scattering problems for Maxwell's equations. The simplified scheme of FEM is as follows. From the family of variational methods, the FEM discretizes the electromagnetic (EM) problem domain Ω into non-overlapping small regions called finite elements Ω^e. One way of discretizing the space is by meshing. In 2D, the mesh consists of triangles or quadrilaterals, and in 3D of tetrahedral, pyramids, hexahedral or prism elements. Suppose, a differential equation has to be solved $L[f(x)] = g$, where L is an operator, g is the source term, and $f(x)$ is the unknown function to be computed in Ω. In the Galerkin method [16] the unknown solution $f(x)$ is approximated by an expansion, set of basis functions v_n with coefficients a_n, according to $f(x) = \sum_n a_n v_n(x)$. The residual $r = L[f(x)] - g$ is formed, and it should be reduced to a minimum. The weighting functions w_m, with $m = 1, 2, \cdots, n$ are defined for weighting the residual r in the form of the inner product $\langle w_m, r \rangle$. Note that performing the inner product with the test weighting functions w_m that are the same as basis functions v_n is what makes the Galerkin method different from the generic method of weighted residuals. The final step is to solve the set of equations $\langle w_m, r \rangle = \int_\Omega w_m r d\Omega = 0$ for coefficients a_n.

Unlike the FDTD method, which requires a mesh of equidistant grid points, the FEM method does not. In comparison to FDTD, FEM can handle more complex geometries. One of the primary benefits of the finite-element method is that subsequent reductions in the mesh size in the computational volume ensure absolute convergence to the end result. The practical issue with this is that as mesh size increases, so does memory and computation time.

2.6.3. Convergence test

It is necessary to set specific optimal discretization of the volume when simulating electromagnetic problems in order to reduce the error between the numerical output and the "true" solution. A fine mesh, especially in 3D (Figure 2.9 A), increases the simulation time and requires more memory. Most of the time the "true" solution is unknown. Therefore, certain field values have to be monitored for different values of the numerical parameters (mesh size, time step etc.) In this thesis, we will mostly rely on the Lumerical (FDTD) [17], Cyclops 3D (FEM) [18], and JCMWave [19] (FEM) software to compute the solution of our scattering problem. The approach to check the convergence of the algorithms was to record the total electric near field at different locations of the volume. The selected test problem is that of a plane wave normally incident on the interface between air and glass with PSL sphere of diameter $d = 50$ nm on the glass.

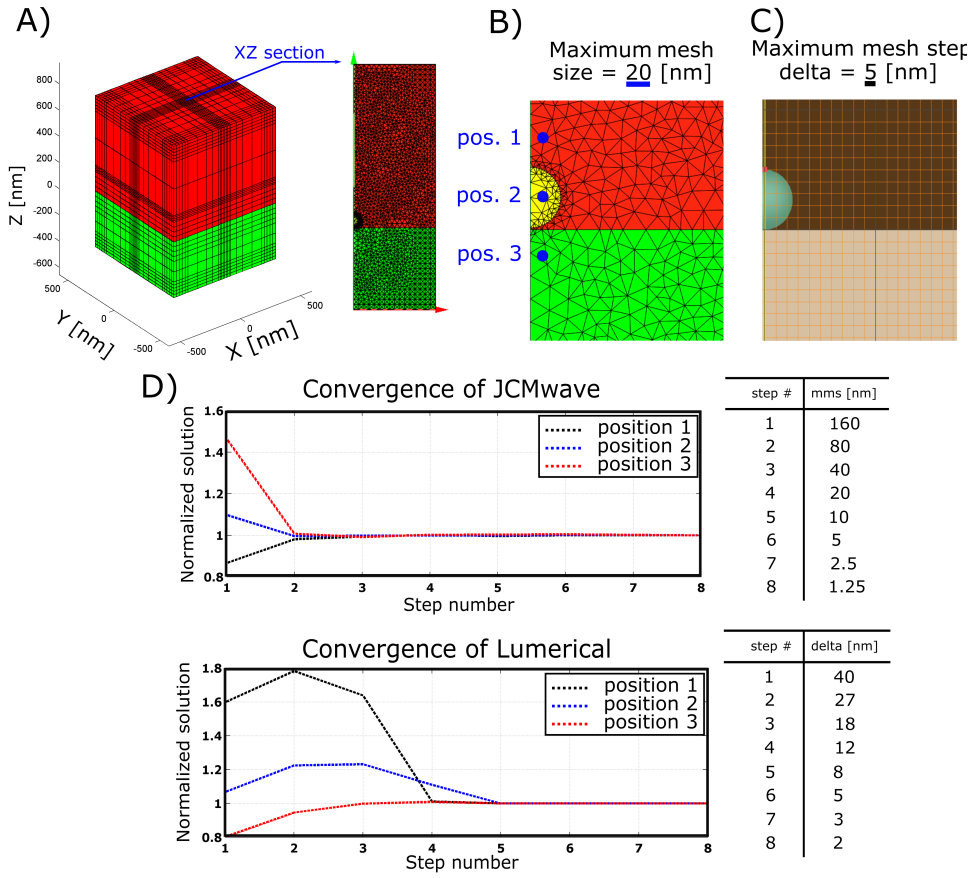


Figure 2.9: A) The meshing of a 3D domain, a sphere of $d = 50$ nm at an interface (environment above and, e.g. silicon slab below). B) The convergence test is performed for three positions. In FEM solvers the parameter of maximum mesh size mms as in figure B) is changed while in the FDTD case shown in C) the maximum mesh step $delta$ is varied. The numerical methods converged for $mms = 40$ nm $mms_{scat} = 4.9$ nm for FEM and $delta = 8$ nm for FDTD.

Unlike the square mesh of FDTD, the mesh of FEM is adaptive, meaning it is non-uniform in different regions. Two types of meshing sizes are defined: 1) mesh size in the regions outside of the scatterer and 2) the mesh size inside the scatterer. For the mesh inside the scatterer the concept of refinement order is introduced in the FEM solver. It is to provide finer mesh at the geometry of the scattering object such that the maximum mesh size for scatterer mms_{scat} at the sphere surface is $\pi d/2^n_{ref} = 4.9$ nm, with the refinement order $n_{ref} = 5$. By keeping constant both finite element degree $p = 2$ and hexahedral mesh with $mms_{scat} = 4.9$ nm, we increase the discretization of the domain outside the particle by decreasing the value of maximum mesh size mms (Figure 2.9 B). In the case of the FDTD, we vary the size of the maximum mesh step $delta$, which is uniform in the whole computational volume (Figure 2.9 C). The three

points of interest for recording the numerically computed electric field are above the sphere, inside the sphere, and inside the substrate medium (Figure 2.9 B). We notice that in JCM solver we achieve the convergence (Figure 2.9 D top) at $mms = 40$ nm and with refinement for a sphere of $mms_{scat} = 4.9$ nm. For the Lumerical FDTD solver, the convergence is established at $delta = 8$ nm (Figure 2.9 D bottom). The corresponding converged settings for the Cyclops FEM solver is $mms = 25$ nm and $mms_{scat} = 6.25$ nm.

2.6.4. Definition of the simulation domain

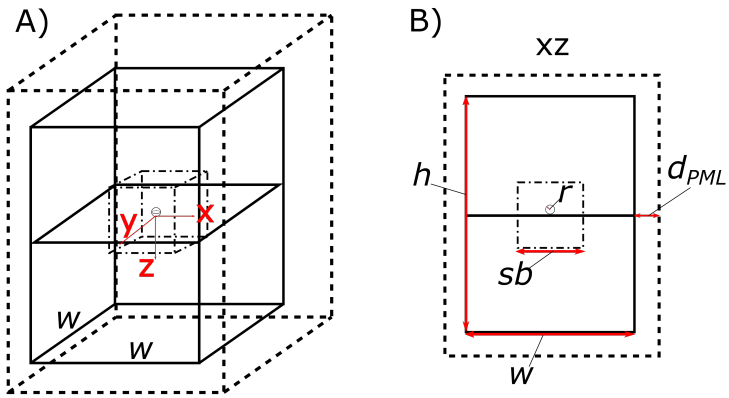


Figure 2.10: A) The sketch of the 3D geometry for the computational electromagnetic problem. B) The xz cross-section showing the size parameters of width w, height h, perfectly matched layer thickness d_{PML}, radius of scatterer r, and size of the sampling box sb.

With other computational parameters being fixed, it is expected that if the domain size increases, the memory requirements will grow in proportion to the number of elements being meshed. It is desired thus to define a computational domain that is as small as possible while still capturing all the properties of the EM interaction. Simulations done in 3D are crucial if one wants to take into account the interaction with the focused spot, and to compute the directionality of the scattering as well as the scattering from an object without rotational symmetry. When adjusting the geometry, there are two things to keep in mind. Firstly, visualizing near-field interactions is only possible over the region inside the computational domain. Secondly, the smallest computational domain for far-field computation, which relies on the near-field, is such that the scatterer and, for example, multilayer structure are included in the interior. There are no other requirements because the PML will ensure that the scattered field's outgoing radiation condition is met, even if the domain is very small.

The typical 3D simulation volume used in this thesis is shown in Figure 2.10 A). The interior domain, which is bounded by the solid line, has the shape of a rectangular prism, which is divided horizontally by the surface of the substrate, with the air above and the substrate below the dividing surface. In the case shown, the particle is on the z-axis of the coordinate system with origin in the surface of the substrate and with

x-axis parallel to the scan direction. The incident field is launched from above the computational volume, and the scattered field from the inhomogenity will propagate either upward - negative z or downward - positive z. The decay of the outgoing waves is treated by the perfectly matched layer with thickness d_{PML}, dashed line, which is used to avoid the reflection from the boundaries of the interior domain. In this thesis, typically we place the sphere of radius r right on top of the substrate. If a user-defined near- to far-field propagation is required, the six electric and magnetic components should be calculated on the faces of the cube that fully surround the local inhomogenity and is inside the physical volume. The edge size of this cube is given by sb and is shown as the dash-dotted line in Figure 2.10 B). The near- to far-field propagation over the computational domain in 2D or 3D in general can not be computed from a simple FFT because in general the computation domain is too small. Instead the Stratton-Chu based formula [18], which is implemented in Cyclops 3D (FEM) [18], or the computation of radiation diagram based on Lorentz reciprocity theorem [20] should be used. The latter one is based on estimating the plane-wave amplitude coefficients for the outgoing waves, based on computing a surface integral on a closed "box" surrounding the local inhomogeneities. These more sophisticated descriptions are beyond the scope of this thesis. Finally, in Table 2.2 we summarize customary parameters for the size of the computational domain.

Table 2.2: Typical values chosen for the geometrical parameters in the numerical parameters shown in Fig. 2.10.

Geometry parameter	Value
w	2λ
h	3λ
d_{pml}	$\lambda/3$
r	$\lambda/8$ to $\lambda/4$
sb	$\lambda/2$ + 2r

References

[1] Omar El Gawhary and Stefan J.H. Petra. Method and apparatus for determining structure parameters of microstructures. *US Patent*, Mar 2015. https://goo.gl/zJkGPq.

[2] Nitish Kumar, Peter Petrik, Gopika K P Ramanandan, Omar El Gawhary, Sarathi Roy, Silvania F Pereira, Wim M J Coene, and H. Paul Urbach. Reconstruction of sub-wavelength features and nano-positioning of gratings using coherent fourier scatterometry. *Opt. Express*, 22(20):24678–24688, 10 2014.

[3] S. Roy, A. C. Assafrao, S. F. Pereira, and H. P. Urbach. Coherent fourier scatterometry for detection of nanometer-sized particles on a planar substrate surface. *Opt. Express*, 22(11):13250–13262, Jun 2014.

[4] S. Roy, Maryse Bouwens, Lei Wei, S. F. Pereira, H. P. Urbach, and P. van der Walle. High speed low power optical detection of sub-wavelength scatterer. *Review of Scientific Instruments*, 86(12):123111, 2015.

[5] VS Ignatowsky. Diffraction by a lens of arbitrary aperture. *Trans. Opt. Inst*, 1(4):1–36, 1919.

[6] B Richards and Emil Wolf. Electromagnetic diffraction in optical systems, ii. structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 253(1274):358–379, 1959.

[7] AS Van de Nes, L Billy, SF Pereira, and JJM Braat. Calculation of the vectorial field distribution in a stratified focal region of a high numerical aperture imaging system. *Optics Express*, 12(7):1281–1293, 2004.

[8] Ernst Abbe. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–468, 1873.

[9] Joseph JM Braat, Peter Dirksen, Augustus JEM Janssen, and Arthur S van de Nes. Extended nijboer–zernike representation of the vector field in the focal region of an aberrated high-aperture optical system. *JOSA A*, 20(12):2281–2292, 2003.

[10] G Bouwhuis, Joseph Braat, A Huijser, J Pasman, G van Rosmalen, and Kees Schouhamer Immink. *Principles of Optical Disc Systems*. 01 1987.

[11] Lukas Novotny and Bert Hecht. *Principles of nano-optics*. Cambridge university press, 2012.

[12] Nilesh Kumar Pathak, P Senthil Kumar, and RP Sharma. Plasmonic perovskite solar cells utilizing noble metal–metal oxide hybrid nanoparticles. In *Noble Metal-Metal Oxide Hybrid Nanoparticles*, pages 487–498. Elsevier, 2019.

[13] Michael Paulus, Phillipe Gay-Balmaz, and Olivier JF Martin. Accurate and efficient computation of the green□s tensor for stratified media. *Physical Review E*, 62(4):5797, 2000.

[14] Allen Taflove and Susan C Hagness. *Computational electrodynamics: the finite-difference time-domain method*. Artech house, 2005.

[15] Kane Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307, 1966.

[16] Bruce A Finlayson. *The method of weighted residuals and variational principles*. SIAM, 2013.

[17] FDTD Solutions. Lumerical solutions. 2003. http://www.lumerical.com.

[18] Xiuhong Wei, Arthur J. Wachters, and H. Paul Urbach. Finite-element model for three-dimensional optical scattering problems. *J. Opt. Soc. Am. A*, 24(3):866–881, 03 2007.

[19] Sven Burger et al. 3D simulations of electromagnetic fields in nanostructures using the time-harmonic finite-element method. In Harald Bosse, Bernd Bodermann, and Richard M. Silver, editors, *Modeling Aspects in Optical Metrology*, volume 6617, pages 255 – 265. International Society for Optics and Photonics, SPIE, 2007.

[20] Jianji Yang, Jean-Paul Hugonin, and Philippe Lalanne. Near-to-far field transformations for radiative and guided waves. *ACS photonics*, 3(3):395–402, 2016.

3

Heterodyne Detection System for Nanoparticle Detection

We present one of the most important limiting factors of CFS, namely the noise in detection. In Sections 3.1-3.3 of this chapter, we point out the importance of nanoparticle detection, propose the heterodyne detection system for CFS, and experimentally investigate its capability for noise suppression. Finally, in Section 3.4 we present the conclusions.

Parts of this chapter have been published in SPIE Proceedings **11056**, 110561A (2019) [1]

3.1. Introduction

As semiconductor devices shrink, the requirement for contamination detection has become more stringent, i.e., to the level of deep sub-wavelength particles ($d << 100$ nm). Contamination can originate, for example, from storage, cleaning, and the handling of wafers [2, 3]. Additionally, undesirable particles can land on top of (or buried in) coated surfaces due to multilayer deposition processes. Accordingly, instruments and techniques for inspection of surfaces have been gaining importance in recent years [4–6]. Ideally, inspection systems should work fast, be sensitive, and should not thermally damage the samples with excessive illuminating power. The scattering-based techniques, such as CFS, operating under bright-field illumination can operate at a low-power, and it has also been demonstrated that high contrast can be achieved even when the particle's index of refraction matches that of the substrate. The latter makes them also suitable for other applications, such as in biology [7–9]. Given that the ultimate limit of the technique has not been established yet, there is room for improvement of its sensitivity. As noise remains one of the critical limitations, in this chapter, we explore the possibility of suppressing its contribution in order to achieve a new milestone in nanoparticle inspection.

We demonstrate improvement on the SNR of about 45 dB with our new detection scheme as compared to the previous detection scheme of CFS [10]. Specifically, we demonstrate the detection of 80 nm PSL particles ($n = 1.57$) on top of a silicon wafer $\underline{n} = 5.43 + i0.34$ [11] at wavelength $\lambda = 405$ nm with an SNR > 58 dB. Moreover, using even lower illumination power in gathering the signal from nanoparticles with diameters $\geq$100 nm becomes possible. In order to achieve these figures, we have merged the original CFS system with the well-known principle in optics and photonics called heterodyne detection.

3.2. Methods and materials

In this section, we present the three building blocks of the experimental setup that has been used to implement heterodyne detection. It consists of the laser source modulation, the up-converted experimental signal as measured by the differential detector, and a lock-in amplifier.

Both modulation and external reference signals are provided by the same device, a HP 8904A multi-function synthesizer. The blue laser diode (wavelength of 405 nm, maximum power of 60 mW) is used in direct modulation mode by modulating its driving current. When a binary "one" is given, light is on while binary "zero" means no light (Figure 3.1 A)). We modulate our laser with a square pulse waveform of either $V_{pp} = 1$ V or $V_{pp} = 1.4$ V so that the peak power at the objective (before being focused on the substrate) is either $P_{low} = 0.58$ mW or $P_{high} = 1.64$ mW, respectively. The same output powers can also be produced without direct modulation, which is used to perform the comparative measurements with the no lock-in configuration.

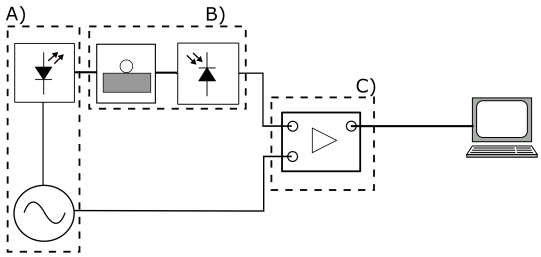


Figure 3.1: The schematic diagram of heterodyne CFS to illustrate the layout for sensing/detection of tiny (d <100 nm, PSL) nanoparticles on top of silicon wafers. A) Represents the waveform synthesizer that is used to modulate the laser amplitude. B) The modulated probing light travels further through the optical system, where the interaction between the light with the scatterer takes place, and is further recorded by the split photodiode (or bi-cell) at the far-field. C) A higher harmonic of the modulation frequency is used for the reference channel of the lock-in amplifier.

Although direct modulation has the limitation of the maximum applied frequencies to be below 3 GHz, it is sufficient to move far from the major noise sources by applying the modulation in the range of a few tens of kHz. In the preferred embodiment, the modulation is in the frequency band from 8 to 12 kHz, suitable for detecting contamination for multiple reasons. Firstly, at such frequencies, the modulated signal is sufficiently far from $1/f$ noise of the electronic components (typically < 200 Hz), acoustic and electrical interference (50 - 60 Hz), and sits at the region of predominantly white noise. Secondly, we choose a frequency which is not a harmonic (multiple) of the known noise sources that are present in the system. Hence, in our experiment $f_m = f_r = 11.111$ kHz. The characteristic frequency of the particle's signal is two orders smaller than the modulation of the laser. Lastly, one should neither use modulation frequencies that are higher than the lock-in amplifier can accept nor sample the signal at a speed that is much lower than the modulation frequency.

Following the schematics of CFS presented in Chapter 2 one can see that the scattered signal in the far-field is directed to a split detector where its two halves are integrated and subtracted from each other, resulting in a differential signal that changes as the sample is scanned. This configuration eliminates the background from spurious reflections from the sample. This is vital for the preliminary suppression of the noise (Figure 3.1 B)). In our new implementation of CFS, the optical system as well as the detector remained the same; only laser amplitude modulation is added to the diode laser in combination with a lock-in amplifier (Figure 3.1 C)) with the goal to improve the SNR of the signal originated from the scattering of nanoparticles. Subsequently, for comparative studies with the previous CFS configuration, the SNR with and without lock-in in the detection of nanoparticles was studied under the same ambient and experimental conditions.

The working principle of the lock-in amplifier (EG&G Princeton Applied Research 5101) is shown in Figure 3.2. In this configuration, also called $1f$ mode, the carrier

signal V_s is connected to the input port of the lock-in amplifier, where the initial DC offset is blocked by the capacitor. The AC signal is further amplified to ensure a large-enough level of the signal. The second input port takes the reference signal V_r. Importantly, the phase of the reference signal can be adjusted, which allows us to control the polarity of the output waveform. The two signals V_s and V_r are multiplied at the mixer to produce the sum and the difference of the corresponding frequency pairs. Finally, a low-pass filter is applied to reject high frequency components.

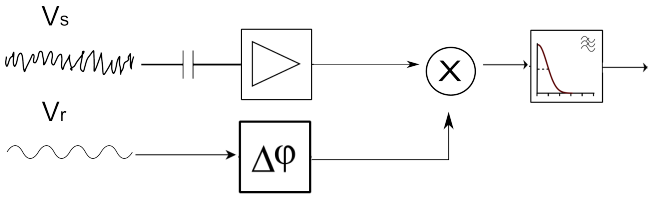


Figure 3.2: Working principle behind the lock-in amplifier. Signal from the photodetector that contains the modulated carrier signal V_s goes to the input channel of the lock-in amplifier. The reference waveform V_r goes to the reference channel of the lock-in. The phase difference between V_r and V_s is controlled by $\Delta\varphi$.

Any phase shift between the reference signal V_r and the measured signal V_s will produce a smaller output than if they were entirely in phase. Hence, we always adjust the phase in such a way that the lock-in amplifier gives the maximum Peak-to-Peak output. One should choose the cut-off frequency of the low pass filter such that on the one hand it reduces the noise to an "acceptable level" and on the other hand, the shape of the signal is not distorted. We also analyze the system in the $2f$ mode, which is obtained by doubling the frequency of the reference signal in order to see if higher SNR gain can be achieved as compared to the $1f$ mode [12].

Preparation of the samples

Three samples have been considered in this chapter; these substrates are silicon wafers covered with PSL nanoparticles. The PSL nanosphere suspension was firstly stabilized by ultra-sonic treatment and shaking and secondly dispersed by spin coating on the silicon wafer [10]. In Table 3.1, we show the diameters and densities of the particles.

Table 3.1: Description of the samples: the particle's material is polystyrene (PSL) with silicon wafer as substrate.

Sample #	Particle [nm]	Density of [particles/mm^2]
1	200	4100
2	100	26000
3	60 and 80	28000

The densities are such that the average distance between them is larger than the focused laser spot on the wafer. In this case, the signal from one particle will not overlap with the signal of neighboring particles, i.e., they can be considered as isolated particles. For each sample, we scan arbitrary areas until we detect a few particles. Essentially, any area of the sample is suitable because the particles are evenly distributed. When comparing the SNRs with and without lock-in, the sample remains in a fixed position with respect to the probing light and hence the same area is scanned in both cases.

3.3. Results

In order to estimate the benefit of using the heterodyning technique for particle detection, we carried out several experiments. We studied bare silicon wafers covered with artificial contamination as highlighted in the previous section.

It is well known that, when applying heterodyne technique, any periodic waveform such as square, sine or triangle can be used for modulating or referencing purposes [12]. In our case, the reference waveform can be chosen freely. However, the choice can affect the results, as shown below. For this work, we have considered a square and a sine reference waveform as well as a lock-in configuration in $1f$ and $2f$ modes.

In order to compare the SNRs with lock-in and no lock-in configuration, we define the SNR gain G (in dB units) as:

$$G = \mathrm{SNR}_{\text{(lock-in)}} - \mathrm{SNR}_{\text{(no lock-in)}}, \quad \text{with } \mathrm{SNR}_{(.)} = 10\log_{10}\left(\frac{\mathrm{S}_{(.)}}{\mathrm{N}_{(.)}}\right). \tag{3.1}$$

In Table 3.2, we show a quantitative comparison between a square and a sine wave reference in the $1f$ and $2f$ modes for the cases of detection of nanoparticles with 100 and 200 nm diameters. Note that the gain G can be higher than 10 dB for the detection of 100 nm spheres with a $\sin(2f_r)$ reference, and, in all cases, the gain G is more than 3 dB. However, only the sinusoidal $1f$ mode shows slight improvement in G for the detection of 200 nm nanoparticles, while the other modes give a negative gain. For this particle size, the original $\mathrm{SNR}_{\text{(no lock-in)}}$ is already high due to large scattering cross-section of these particles, and so, there is no benefit in using lock-in techniques. For this data, the power of the laser at the objective was $P_{low} = 0.58$ mW.

Table 3.2: SNR gain G [dB] for the cases of detection of 100 and 200 nm particles, for various lock-in configurations.

	$\sin(f_r)$	square(f_r)	$\sin(2f_r)$	square($2f_r$)
200 nm PSL	1.98	-14.03	-13.41	-18.21
100 nm PSL	4.39	3.34	10.81	10.07

From the results shown in Table 3.2, one can conclude that the sinusoidal reference waveform is superior to the square one.

We also compare the f and $2f$ modes for the sample that contains a mixture of

60 and 80 nm particle sizes. The results presented in Table 3 show that gain in the SNR can be observed for all the tested modes. If $1f$ mode is used we can estimate the G > 9 dB and G > 25 dB for the $2f$ mode. However, for the case of $1f$ mode, the introduced gain is not enough to detect a particle with high reliability as final SNR for either square or sine reference is < 8 dB. For the detection of 80 nm particles, the $2f$ mode has better performance than the $1f$ mode because both sine and square reference waveform in the $2f$ mode can recover the particle's signals with excellent SNR > 19 dB. For this data, the power of the laser at the objective was $P_{low} = 0.58$ mW.

To illustrate the improvement in SNR for the $2f$ mode, we superimpose the signal from the lock-in configuration (red solid line), with the no lock-in configuration (blue dashed line), as shown in Figure 3.3. For this data, the power of the laser at the objective is P_{high}. From this figure, one can see that a distinctive gain is achieved when detecting 80 nm particles (G of 45 dB). However, we point out that, with this detection configuration and laser power, the 60 nm particles have not been detected.

Table 3.3: SNR gain G [dB] for the detection of 80 nm particles for various configurations of the lock-in system.

80 nm PSL	sin(f_r)	square(f_r)	sin($2f_r$)	square($2f_r$)
SNR [dB]	9.89	5.45	25.11	20.23

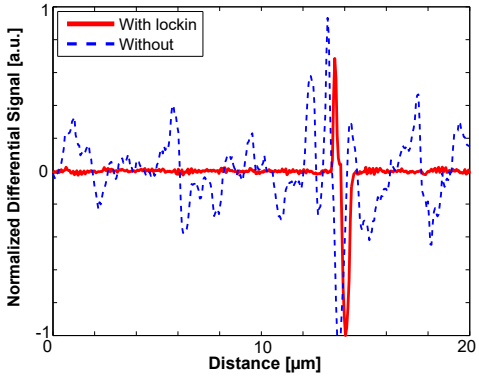


Figure 3.3: The normalized differential signal that originates from 80 nm particles with lock-in configured in $2f$ mode and sinusoidal reference (solid line, red) compared to no lock-in configuration (dashed line, blue). The gain G in SNR is of 45.17 dB.

In view of the achieved improvement in SNR, we emphasize that when dealing with larger particles such as 100 and 200 nm spheres and knowing that scattering varies linearly with the light intensity [2], the power of the laser can be much lower, since the SNR ratio is large enough for these particles in the present configuration of our system. That could be of benefit in applications that require gentle treatment

of inspected surfaces (e.g. polymers, biological samples). However, the detection of particles smaller than 80 nm still requires sufficient power to overcome the noise at the detector (> 1.5 mW), due to the smaller scattering cross-section [13, 14].

3.4. Conclusions

In this chapter, we applied lock-in amplifier techniques to CFS to increase the SNR of the detection of PSL nanoparticles. Compared to the previously available conventional operational regime of the CFS [10], our new implementation has led to an improvement in the SNR of about 45 dB for the detection of nanoparticles with diameters of 80 nm ($\approx \lambda/5$) deposited on a silicon wafer. Given the improvement in SNR, one can also explore the possibility of using lower illumination power, which could be beneficial for biological applications or inspection of plastic substrates, where too much power would mean damage of either the substrate or the specimen, or both. An improvement of the SNR will also help when dealing with samples where the contrast between the nanoparticles and substrate is very low, such as index matching PSL nanoparticles on glass. Although the proposed heterodyne CFS technique already shows excellent performance for detection of polystyrene nanoparticles on silicon wafer, there is still room for improving the sensitivity towards even smaller particles with the upgrade for split-detector schematic (see Appendix B).

References

[1] D. Kolenov, R. C. Horsten, and S. F. Pereira. Heterodyne detection system for nanoparticle detection using coherent Fourier scatterometry. In Peter Lehmann, Wolfgang Osten, and Armando Albertazzi Gonçalves Jr., editors, *Optical Measurement Systems for Industrial Inspection XI*, volume 11056, pages 336 – 342. International Society for Optics and Photonics, SPIE, 2019.

[2] Takeshi Hattori, Akira Okamoto, and Hitoshi Kuniyasu. Challenges of finer particle detection on unpatterned silicon wafers. *AIP Conference Proceedings*, 683(1):271–277, 2003.

[3] Alan Allan et al. International technology roadmap for semiconductors 2.0. *Executive Report*, 2015. http://tinyurl.com/yxrwy5oo.

[4] Steve Yifeng Cui, Jay Huang, Chunhai Wang, Christian Wolters, Bret Whiteside, Anatoly Romanovsky, Chuanyong Huang, and Donald Peetibone. Surface defect inspection with large particle monitoring and laser power control. *US Patent*, (WO2018027010), Feb 2018. https://goo.gl/x8fKPT.

[5] Nan Zhang, Zhiyuan Gu, Shuai Liu, Yujie Wang, Shuai Wang, Zonghui Duan, Wenzhao Sun, Yun-Feng Xiao, Shumin Xiao, and Qinghai Song. Far-field single nanoparticle detection and sizing. *Optica*, 4(9):1151–1156, Sep 2017.

[6] U. Okoroanyanwu et al. Towards the optical inspection sensitivity optimization of euv masks and euvl-exposed wafers. volume 8352, pages 8352 – 8352 – 14, 2012.

[7] S. Roy, S. F. Pereira, H. P. Urbach, Xukang Wei, and O. El Gawhary. Exploiting evanescent-wave amplification for subwavelength low-contrast particle detection. *Phys. Rev. A*, 96:013814, Jul 2017.

[8] Shuming Nie and Richard N. Zare. Optical detection of single molecules. *Annual Review of Biophysics and Biomolecular Structure*, 26(1):567–596, 1997.

[9] Oguzhan Avci, Celalettin Yurdakul, and M. Selim Ünlü. Nanoparticle classification in wide-field interferometric microscopy by supervised learning from model. *Appl. Opt.*, 56(15):4238–4242, May 2017.

[10] S. Roy, Maryse Bouwens, Lei Wei, S. F. Pereira, H. P. Urbach, and P. van der Walle. High speed low power optical detection of sub-wavelength scatterer. *Review of Scientific Instruments*, 86(12):123111, 2015.

[11] D. E. Aspnes and A. A. Studna. Dielectric functions and optical parameters of si, ge, gap, gaas, gasb, inp, inas, and insb from 1.5 to 6.0 ev. *Phys. Rev. B*, 27:985–1009, Jan 1983.

[12] George Greenwald. The lock-in amplifier: Noise reduction and phase sensitive detection, 2007. https://goo.gl/XcxjwK.

[13] Elton Williams Howard R. Huff, Randal K. Goodall. Measurement of Silicon Particles by Laser Surface Scanning and Angle-Resolved Light Scattering. *Electrochem. Soc.*, 144:243–250, 1997.

[14] N. Dahan and J.-J. Greffet. Enhanced scattering and absorption due to the presence of a particle close to an interface. *Opt. Express*, 20(S4):A530–A544, Jul 2012.

4

Polarization effects in evanescent wave amplification for particle detection

This chapter focuses on the index-matched particle-substrate problem. If the substrate has single or multiple thin layers on its surface, the sensitivity can be improved by engineering the near-field enhancement mechanism. We considered a thin dielectric layer added to the top of the original substrate making a 2-layer structure. Also, given the fact that we use focused light on the sample, polarization effects play an essential role. With numerical simulations that have been confirmed by experiments, we were able to arrive at a comprehensive understanding of the problem as well as suggest the most suitable materials for the cover layer and input polarization state.

Recent literature [1] has shown improvement to bio-detection by adding a thin layer of dielectric on the substrate. Although the enhancement mechanism is different from the one proposed here, we believe that our work could have a contribution in this area. Thus, in this chapter, we start with a discussion on applying CFS in the context of the detection of particles in biological specimens. Then, in Section 4.2 we outline the evanescent wave amplification (EWA) enhancing mechanism. Further, in Section 4.3 and Section 4.4, we investigate the design of the dielectric layer, far-field gain due to cylindrically or conventionally polarized beams, and the robustness of the EWA. In Section 4.5, we present the experimental results for the detection of $\lambda/8$ particles on the 2-layer structure. In Section 4.6, we finalize the chapter with discussions and conclusions.

Parts of this chapter have been published in OSA Continuum **3**, 742-758 (2020) [2]

4.1. Introduction

The need for the detection and size identification of (bio-) nanoparticles has been one of the main driving forces for a number of the near-field and far-field optical techniques in the last few decades. In turn, single molecule detection is one of the significant challenges of optics applied to modern biology [3]. The behavior of individual particles and molecules can have significant implications for both properties of individual cells and biochemical processes [4]. Notably, bio-nanoparticles such as exosomes or viruses are known to be important bio-markers for a range of medical experiments [5]. They are challenging to visualize by optical means because they are index-matched with the substrate [6]. Additionally, their size is small: the diameters of viruses range from 18 to 300 nm and of exosomes from 30 to 100 nm [7]. When imaging nanoparticles with diameters $< \lambda/8$, the diffraction limit of optical systems becomes a prohibiting factor [1]. Existing labeling techniques found in fluorescence microscopy can overcome the diffraction limit. For example, wide-field epi-illumination confocal microscopy combined with fluorescence are popular methods for single particle detection. A review of these methods can be found in Ref. [7]. However, such techniques have drawbacks due to the complexity of sample preparation and labeling [5]. Another popular approach is near field optical microscopy (NSOM). But this also has difficulties because of tip perturbations, risk to add contamination of the sample, and limited throughput [7].

In the last few years, new detection schemes have been introduced that eliminate the need of fluorophores and work in the far field. For instance, the technique called common path wide-field interferometric microscopy [1] uses a layered substrate. It works in reflection mode, and the function of the extra layer is to eliminate the background due to destructive interference. The strength of this technique is that the detection is in the far-field and is very sensitive to subwavelength nanoparticles with small optical contrast [8].

The use of layers on the substrate as extra "tool" to allow subwavelength detection of single particles is also at the heart of the technique that is discussed in this chapter. However, the principle is quite different than common path interference microscopy. Essentially, our method is based on exploring the interaction between the incoming light with the nanoparticle and the substrate with the latter having a thin layer of dielectric material deposited on it. This interaction results in evanescent wave amplification in the near field and enhancement of the scattering to the far field [9, 10]. In this chapter, we go one step forward with this method by presenting a study of appropriate materials (dielectric) and appropriate thicknesses for the thin layer and an analysis of the influence of the polarization state of the input light. We show that by optimizing the system, one can drastically improve the sensitivity of the detection of nanoparticles. Further, we demonstrate the detection of 40 nm PSL nanoparticles in diameter with a wavelength of 405 nm.

4.2. Enhancement mechanism

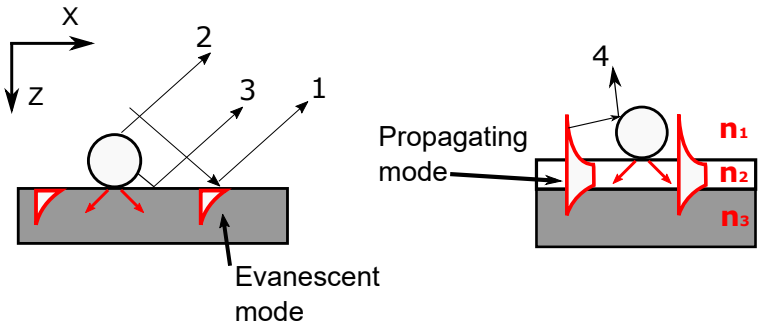


Figure 4.1: Left: the scattered and reflected components that contribute to the far field signal and evanescent modes decay quickly in the substrate and in air. Right: evanescent waves at the interface due to the presence of a thin layer on the substrate are amplified and re-scattered by the particle.

The light at the split detector is a superposition of mainly three contributions (see Figure 4.1 left): the mirror-like reflected field from the substrate (1), the direct scattered field by the particle (2), the field scattered by the particle that is further reflected from the substrate (3). The scattering from the particle on top of substrate can generate the weak evanescent mode that decays very fast and cannot reach the detector. When switching to, e.g. the 3-layer system (air, enhancement layer, substrate with refractive indices n_1, n_2 and n_3), a fourth channel (4) can be invoked; firstly, the particle couples light to the modes that can propagate in the layer (Figure 4.1 right), hence it is critical that $n_2 > n_1$ and $n_2 > n_3$, and secondly, the generated decaying tail is converted to the propagating modes by re-scattering from particle [9, 10]. It is crucial to create such an enhancement layer that the amount of the decay towards the first medium is maximum.

We believe that the evanescent wave amplification that we employ is highly relevant for biological applications because it manifests itself even with low power illumination such as in the range of smaller than 1 mW, as we will show experimentally. Low illumination power is important in systems where the specimen could be damaged through heat [11]. Our approach is more efficient than only using the massive power of illumination for one more reason: in the proposed mechanism, we can amplify a characteristic signal from the particle rather than from its surroundings. To overcome the noise level, we introduce and tune the thickness and material of the enhancement layer until it has a sizable contribution to the evanescent field, and this light is being re-scattered specifically by a sphere.

4.3. Thin film material analysis

One should choose a suitable material to achieve enhanced scattering. Firstly, for the probing wavelength of interest, the layer should have both a higher refractive index than the substrate surface $n_2 > n_3$ and low absorption to allow the guiding.

Secondly, in applications in which the transmission properties of the resulting stack should be high, the enhancement layer should be very thin (of the order of a few tens of nanometers) in order to avoid much light loss because of absorption. Finally, the thickness should be tuned so that the most substantial evanescent part of the spectra builds up near the interface between the host medium n_1 (e.g. air) and top layer n_2.

Enhancing the far-field response by coating the glass or plastic substrate with a single layer of dielectric is convenient because, in the visible regime, it is relatively straightforward to find materials that do have a higher refractive index than glass or plastic. See, for example, the set of common dielectric and semiconductor materials shown in Table 4.1.

Table 4.1: Examples of dielectric and semicon. materials that can act as single layer EWA. The optical properties are defined for the wavelength 405 nm. The substrate should have a lower refractive index than a cover layer to allow the guiding. Deposition processes are Evaporation: conventional thermal evaporation in high vacuum, EPVD: Electron-beam physical vapor deposition, CVD: chemical vapor deposition techniques, Arc-PVD: cathodic arc plasma deposition. Ext. coeff. refers to extinction coefficient.

Material	Index	Ext. coeff.	Process
TiO_2	2.66	1e-3	Evaporation + Arc-PVD
Ta_2O_5	2.15	1e-4	EPVD
Si_3N_4	2.09	≈ zero	CVD
c-Si	4.4 or 5.437	2.3 or 0.34	CVD
GaP	4.15	0.25	CVD
Ge	4.11	2.18	Evaporation
InSb	3.39	2	CVD
alpha-Si	3.8	2.5 or 0.5 process dependent	EPVD

Our goal is to maximize the scattering that reaches the detector. In the case of very small particles, it is crucial that the surface under the particle is smooth and does not introduce too much scattering. Dielectric materials are favored because they allow large area and homogeneous deposition with high levels of smoothness.

In addition, one should add that the enhancement mechanism considered here is different from conventional ways of getting greater optical response from the surface under inspection such as high reflecting surfaces by introducing single-layer metal or by producing multilayer reflectors. By using one single dielectric layer, our approach offers the elegance regarding both simplicity and surface uniformity. The selection of the dielectric materials which have sufficiently high real part of the refractive index and small absorption (low imaginary part), yields the effect that is similar to metal covering layers while it also contributes to the enhancement of the evanescent waves. The concept of the enhancement of evanescent energy will hold for any combination

of the considered substrate and cover layer (as given in Table 4.1), while the medium between the layered substrate and lens is either air, water or oil (the latter two for immersion applications), so that, together, it creates an asymmetric waveguide $n_2 > n_1$ and $n_2 > n_3$. A minimum thickness of the dielectric layer is also required for the existence of a guided wave.

The introduction of the covering layer of large (compared to the wavelength) thickness is a no-go solution. The original properties of the substrate, such as transmission, will be hindered if the covering layer is too thick. Conversely, thin films of approximately $\lambda/20$ which allow for guiding of at least one mode and consequently boost of the evanescent field, keep the design simple and compact [12].

Another important issue in the choice of materials is the deposition process. In Table 4.1, we include, for each material, the traditional deposition processes. The fabrication accuracy and the temperature requirement vary. So, for instance, CVD processes run at much higher temperatures than PVD processes (usually between 300°C and 900°C) which can be inappropriate to cover glass or plastic substrates directly because they cannot tolerate such temperatures. Also, PVD has a distinct advantage because it has a more precise film thickness control, in the order of sub-nm precision. This is important in our application, considering the required thickness of the thin film (around 10-20 nm).

4.4. Numerical Results

For the simulations, we performed both the 1D multilayer method [13] to analyze the Fresnel reflection coefficients when the substrate plus layer system is excited by evanescent waves, and rigorous 3D vectorial simulations using a homemade finite element method [14] to analyze the scattering enhancement due to a the presence of a particle on the top of the substrate plus layer system. In both 1D and 3D problems, the substrate roughness is neglected. The 1D simulations give insight about the background field that would be created if either an s- or p-polarized plane wave is incident at the interface in the absence of a scattering object. This step is necessary to determine the thickness that maximizes the evanescent components at the interface where the particle will lie. It is also important to look at both polarization components, s and p, because the incident field will include them both, for example, a tightly focused linearly polarized beam. The complex refractive index $\underline{n} = n + i * k$ for a thin film of varying thickness that accounts for small changes in absorption is simulated through the software package OpenFilters [15]. The simulations are done for the wavelength of 405 nm, and the numerical aperture of the system (for the 3D simulations) was equal to $NA = 0.9$.

Corresponding to the 3-layer system introduced in the Figure 4.1 B), we show the cavity in Figure 4.2 A). We consider a middle layer whose index of refraction is n_2 and thickness is d_2 surrounded by an input medium of index n_1 and an output medium with n_3. We have incident wave $\boldsymbol{k}_1 = \boldsymbol{k}_{z1} + \boldsymbol{k}_x$ where $\boldsymbol{k}_{z1}$ and $\boldsymbol{k}_x$ are the normal and tangential wave vectors (incident from the left in Figure 4.2 A).

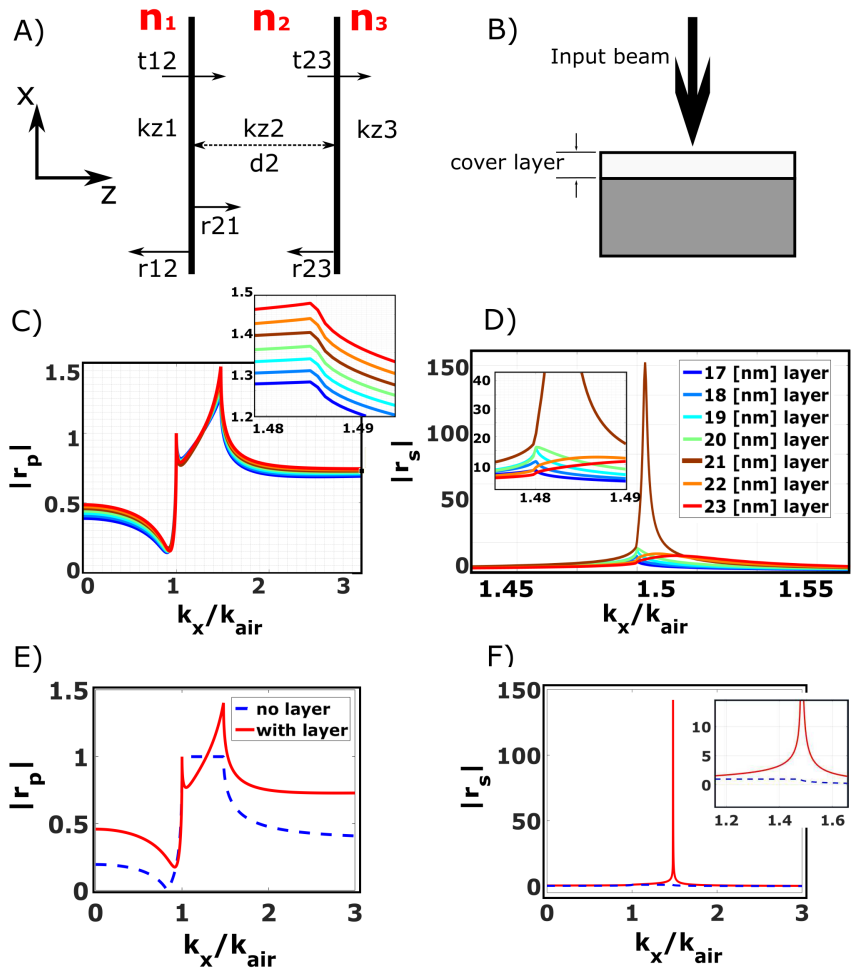


Figure 4.2: A) Incident reflected and transmitted waves for three media layered structure. B) Scheme of the simulation system. Fresnel reflection coefficients for the three layer medium (air as input medium, TiO_2 as thin layer and glass as substrate). C) and D): absolute value of the Fresnel reflection coefficients as a function of the thickness of the TiO_2 layer for p and s polarizations, respectively, when the thickness of the covering layer changes from 17 to 23 nm. For the case of 21 nm thick TiO_2 layer, we show in E) and F) the Fresnel reflection coefficients with (red) and without (blue) the thin layer.

The incident wave is partially reflected and partially transmitted through the layer. The transmitted part reaches subsequently the second interface and will either be reflected or transmitted again. According to the literature [13] and [16] by adding all the coherent amplitudes for reflection or for transmission, we arrive at the equations known as Fabry-Perot (FP) formulas for a single cavity which are exact, valid for the input of evanescent waves, work with any thickness of d_2. Here we only use the

reflected part:

$$r = r_{12}^{\nu} + \frac{t_{12}t_{23}r_{23}exp(ik_{z2}d_2)}{1 - r_{23}r_{21}exp(i2k_{z2}d_2)}, \tag{4.1}$$

where superscript $\nu = s, p$ defines either s or p polarization, $k_{z2} = \sqrt{k_0^2 n_2^2 - k_x^2}$ is the z component of the wavevector in the medium 2 and the r_{jm}^{ν} (t_{jm}^{ν}) with j and $m \in$ 1, 2 or 3 is the reflection (transmission) coefficient corresponding to the wave reaching the interface from medium j to medium m. We have

$$r_{jm}^{(s)} = \frac{\mu_m k_{zj} - \mu_j k_{zm}}{\mu_m k_{zj} + \mu_j k_{zm}}, \quad r_{jm}^{(p)} = \frac{\epsilon_m k_{zj} - \epsilon_j k_{zm}}{\epsilon_m k_{zj} + \epsilon_j k_{zm}}, \tag{4.2}$$

$$t_{jm}^{(s)} = \frac{2\mu_m k_{zj}}{\mu_m k_{zj} + \mu_j k_{zm}}, \quad t_{jm}^{(p)} = \frac{2\epsilon_m k_{zj}}{\epsilon_m k_{zj} + \epsilon_j k_{zm}}, \tag{4.3}$$

where ϵ is the material's relative permittivity, and μ is its relative permeability. One notices that a particular combination of $k_x > k_0 n$ will render the imaginary value for k_z that represents an evanescent wave. Since we are interested in maximizing the evanescent wave in the reflection mode next to the particle, i.e. in the medium of air with $n_1 = 1$, we will study the reflection as the function of in-plane wavevector not only for the propagating modes $k_x < k_0 n_1$ but also higher spatial frequencies $k_x > k_0 n_1$. Further, the wavevector in the first medium is written as $k_1 = k_0 n_1 = k_{air}$. Here we show simulations for TiO_2 as a thin layer on glass; the same procedure was used for the Ta_2O_5. For Ta_2O_5, the Sellmeir equation [17] is used to approximate the real part of refractive index and Urbach absroption tail Eq. (4.4) to model the imaginary part of the weakly absorbing thin film:

$$\begin{aligned} n^2(\lambda) &= 1 + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} + \frac{B_3\lambda^2}{\lambda^2 - C_3} \\ k(\lambda) &= D\exp\left[E\left(12400\left(\frac{1}{\lambda} - \frac{1}{F}\right)\right)\right] \end{aligned} \tag{4.4}$$

with $B_1 = 3.3$, $C_1 = 0.005[\mu m^2]$, $B_2 = 0.2$, $C_2 = 0.01[\mu m^2]$ and $B_3 = 0.1$, $C_3 = 0.02[\mu m^2]$ and for absorption $D = 180\text{e}{-9}$, $E = 8[\mu m]$, $F = 4000[\mu m]$.

The Cauchy-Urbach dispersion model [18, 19] Eq. (4.5) is used for the TiO_2 film:

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \tag{4.5}$$

with $A = 2.1959$, $B = 0.025614[\mu m^2]$, $C = 0.0059846[\mu m^4]$ and $D = 0.0025745$, $E = 8.508[\mu m]$, $F = 4000[\mu m]$ The choice for this configuration is obtained by selecting the system with the strongest contribution of the evanescent waves.
For a three layer system (air-thin layer-glass, see scheme in subfigure 4.2 B), we plot in subfigures C) and D) of Figure 4.2 the Fresnel reflection coefficients of s and p input polarization for layer thicknesses ranging from 17 to 23 nanometers for the case of TiO_2 as thin layer. In subfigures E) and F) we show the comparison between the

absolute value of the Fresnel reflection coefficients for the bare (no thin layer) and covered substrate with 21 nm of TiO_2.

Specifically, the region of spatial frequencies $k_x/k_{air} > 1$ is interesting because it represents the non-propagating waves, where for values of r_p and $r_s > 1$, it means amplification. It is evident that the presence of a thin layer such as TiO_2 results in a growth of the evanescent field that correspond to a guided mode. Moreover, the reflected amplitude $k_x/k_{air} < 1$ is also higher with the layer than without, but not as much as the evanescent part. This effect is expected because the contrast between air and TiO_2 is higher than contrast between air and glass $n_{TiO2} > n_{glass} > n_{air}$.

Once the thickness of the TiO_2 layer is selected to be 21 nm, we consider the full 3D configuration to evaluate the effects of the polarization states of the probing focused light with and without the particle on it. This leads us to perform 3D rigorous simulation with the incident light focused at the top layer, and including a nanoparticle at the center of the focused beam (see Figure 4.3 B).

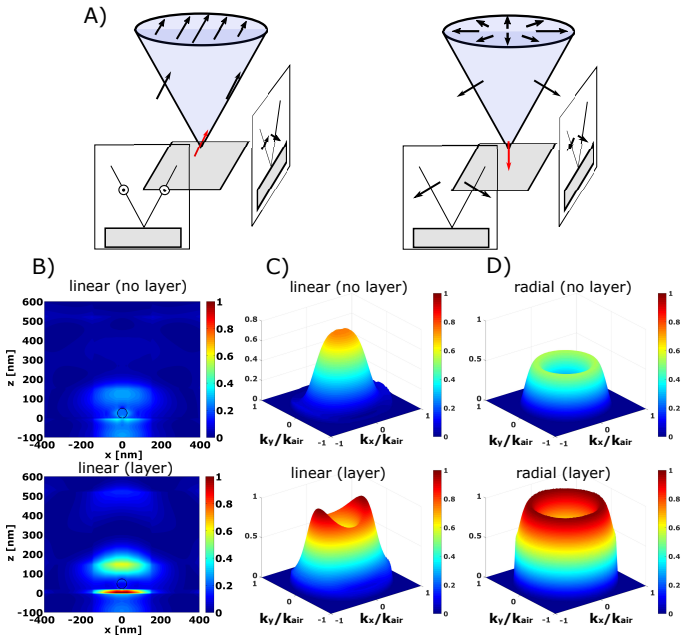


Figure 4.3: A) Linearly (left) and radially (right) polarized light focused on the substrate (electric field in focus in red). B) The near field when linearly, along the x-axis polarized light is focused onto a 50 nm PSL nanoparticle on top of the substrate, without (top) and with (bottom) enhancement layer. C) and D) The far field-maps without (top) and with (bottom) layer for the case of linear polarization and radial polarization, respectively. The enhancement layer is 21 nm of TiO_2, the wavelength is 405 nm and NA = 0.9. The color-code is such that the observed intensity is normalized by the maximum value of the far-field when the TiO_2 layer is present.

The field components as part of the illumination cone interact with both particle and substrate. The diameter of the polystyrene (PSL) nanoparticles with refractive

index $n = 1.57$ at $\lambda = 405$ nm is either 50 or 40 nm. When the beam is focused, all incident angles and polarization components within the numerical aperture of the objective should be considered. We calculate the scattered near and the far-field of the particle on top of the bare substrate and the substrate covered with the thin layer illuminated by the focused laser beam. The scattered field due to the particle is computed by subtracting the portion that is caused by the given incident field when only the multilayer is present from the total field in the presence of the particle. The different polarization states are used: linear, circular, radial, and azimuthal. The boost in the near (all planes) and far field is evident and present in all four cases. The total near field in the xz plane for the case of linear polarization along the x-axis is shown in Figure 4.3 B) and the far field within the NA of the objective for linear and radial polarization is shown in Figure 4.3 C) and D). The color-code is such that the observed intensity is normalized by the maximum value of the far-field when the TiO_2 layer is present. Following the near-field enhancement as in Figure 4.3 B), we observe that linear polarization causes the far field to grow and remarkably changes its angular distribution in 4.3 C). In the case of radial polarization, the far field is also enhanced but does not considerably change in shape (4.3 D)). The central part of the far field map is reduced compared to no layer case and it is biggest in the region where spatial frequencies exceed approximately $k_x/k_{air} > 0.2$, corresponding to acceptance angles higher than 11.5° with respect to the normal. The angular distribution of the far-field can be firstly explained by the shape of the focused spot in the focal plane. The x-linearly polarized beam, left 4.3 A), has the focal spot elongated along the polarization axis, leading to the far field map that is also a bit wider in the x-direction than in the y-direction, hence there is no rotational symmetry. For the focusing of radially polarized light, (4.3 A, right side), the strong longitudinal component of the electric field at the optical axis is responsible for the hollow central part of the far-field map and the rotational symmetry of the spot also translates to the scattered far-field. Secondly, due to the sole p or joint s and p polarization content of the radially and linearly polarized beams, respectively, after the scattering from the particle on the enhancement layer, (4.3 C) and D), bottom), the propagating and evanescent-converted components lead to destructive or constructive interference in different parts of the optical pupil.

Next, to quantitatively evaluate the enhancement of the scattered far field within the NA of the detector that each polarization produces, we introduce the far-field gain factor G. In the equations below Eqs. (4.6-4.8) , we look at two metrics: G_{int} and G_{max}. The G_{int} is the ratio of the total scattered detected intensity in the case of the covered substrate and that of the bare substrate. Similarly, G_{max} is the ratio of the maxima of the scattered intensities over the detector.

$$G_{int} = \frac{I_{Cov_int}}{I_{Bare_int}}, \qquad I = \int_{n=1}^{nrows} \int_{m=1}^{ncol} I(n,m)dndm, \tag{4.6}$$

$$G_{max} = \frac{I_{Cov_max}}{I_{Bare_max}}, \qquad I_{max} = max(I(n,m)), \tag{4.7}$$

$$I(n,m) = (|E_x(n,m)|^2 + |E_y(n,m)|^2). \tag{4.8}$$

The integrals above are calculated by the trapezoidal method. With both metrics that we used, either integral or based on the maximum value, the gain factor expected by the cylindrically polarized light (radial and azimuthal) is superior to conventionally polarized beams, as shown in Table 4.2.

Table 4.2: The gain factors indicate that cylindrically polarized light produces higher gain in the far-field than the conventional polarizations. The material TiO_2 shows better performance than Ta_2O_5. The maximum gain in each column is given in bold. The diameter of the particle is 50 nm.

Polarization	G_{max}		G_{int}	
state	Ta2O5	TiO2	Ta2O5	TiO2
Linearly X	1.13	1.32	1.65	1.97
Radially	2.04	**2.91**	**2.49**	**3.76**
Azimuthally	**2.08**	2.86	2.47	3.53
Cicularly	1.04	1.13	1.65	1.97

For both linear and circular polarization, the central part of the far-field angular distribution is barely larger than at the pupil's edges, whereas for azimuthal and radial polarization, the enhancement occurs at all angles. However, for all polarizations, the position of the maxima of the distribution lies in the high angular range $k_x/k_{air} > 0.2$. This is one reason why we have introduced G_{max} ratio. For example, the G_{max} parameter for the TiO_2 layer and particle of 50 nm is equal to 2.91 for radial polarization. This suggests that for high-enough SNR, applying an annular aperture at the detector plane can yield better sensitivity. In fact, the benefit of the detection with an annular aperture was demonstrated in previous research [20]. In that reference, it was anticipated that the collected far field at higher angles of the aperture are more influenced by the presence of the particle, and this is confirmed here by the far field simulations.

As can be seen from Table 4.2, we also introduce Ta_2O_5 as enhancing layer, which has lower refractive index as compared to TiO_2. Following the same procedure as before we find again that the most suitable thickness is 20 nm. Due to slightly higher absorption, the evanescent amplification that surrounds the particle is lower so both G_{max} and G_{int} are smaller than for the TiO_2 layer.

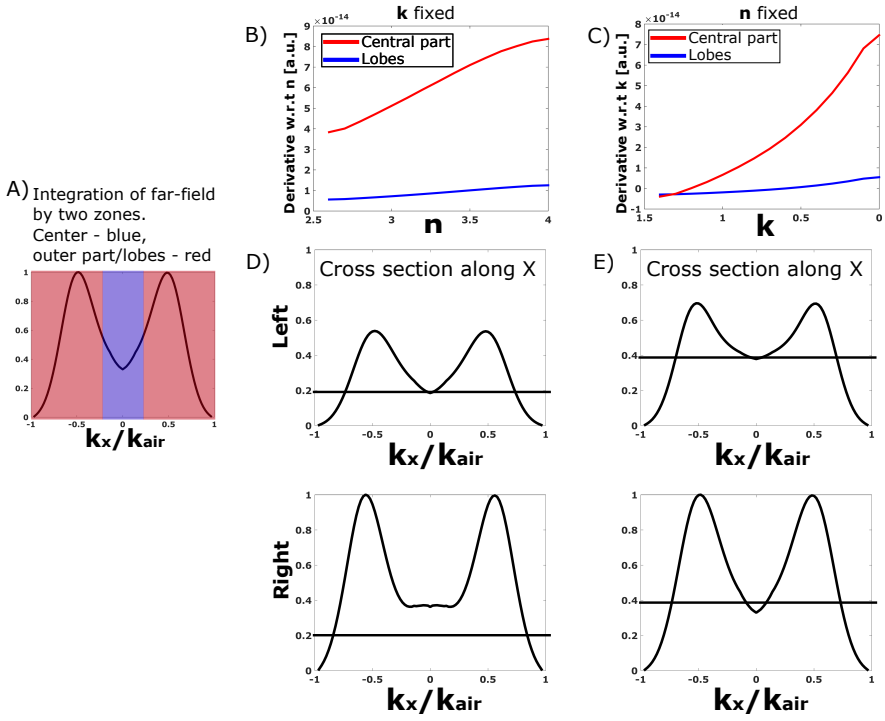


Figure 4.4: A) Cross-section of the far field intensity along the x-axis when the input beam is radially polarized and focused on a 30 nm nanoparticle. The middle zone in blue corresponds to the center of the pupil $|k_x/k_{air}| \leq 0.2$ and the surrounding hollow cone in red corresponds to $0.2 < |k_x/k_{air}| < 0.9$. B) Point per point gradient of the integrated electric field intensity of the blue and red regions of the far field as a function of the real index of refraction with the imaginary part kept at the value of 0.133, C) as function of the imaginary part of the index of refraction with the real part kept at the value of 2.6. D) and E) profile of the field distributions corresponding to the left- and right-most values of the refractive indexes of plots B) and C), respectively. Black lines mark the center of the cross-section $k_x/k_{air} = 0$ that corresponds to the beginning of refractive indexes sweep.

Further, to understand whether the evanescent wave conversion mechanism may explain the superior detection performance, we have studied how the value of the real and imaginary part of the index of refraction of the enhancement layer influences the far-field increase. In order to do that, we perform a parametric sweep. A radially polarized beam focused at the center of the particle of diameter 30 nm, $NA = 0.9$ and enhancement layer with fixed thickness of 21 nm of a fictitious material is considered. We show in Figure 4.4 the far field as the values of the real or imaginary part of the refractive index varies. In Figure 4.4 B) the imaginary part of the refractive index remains fixed as $k = 0.133$ and n changes from 2.6 to 4 in 15 steps. In Figure 4.4 C), the real part of the refractive index is $n = 2.66$ and k changes from 0.004 to 1.5

in 15 steps. Note that for the far-field cross-sections (4.4 D) and E)) the scattered field is cylindrically symmetric (as in 4.3 D)). The cross-section curves D) and E) are normalized to unity by the maximum value of the far field intensity of the corresponding parametric sweeps. For instance when $\underline{n} = 4 + i * 0.133$, that corresponds to the rightmost value in the parametric sweep B), the far-field cross-section (bottom of Figure 4.4 D)) has the maximum value of 1 thus the enhancement is the highest.

In order to emphasize the region of the pupil where either propagating or coupled-evanescent components contributes, we integrate the far-field intensity by two zones: $|k_x/k_{air}| \leq 0.2$ and $0.2 < |k_x/k_{air}| < 0.9$, seen as blue and red color of the cross section in Figure 4.4 A). The integral for total field intensity of both zones is calculated by applying the trapezoidal rule and on top of that we compute the derivative w.r.t n or k. From the profiles B) and C), one can see that the amount of energy that scatters to the outer parts of a pupil is greatly boosted when absorption is low (as can be seen in the region where $k \leq 0.5$), whereas when the real part of the refractive index ($n \geq 2.6$) is increased, the central and side contributions raise almost linearly. The index of refraction properties are varied separately because they have a major influence on either the propagating modes or the coupled-evanescent components. Also, in Figure 4.4 B), the growing behavior of the gradient can be characterized by an almost linear behavior of the intensity in both central (blue curve) and outer part of the pupil (red curve). In contrast, if the absorption decreases from $k = 1.5$ to $k = 0.133$ (see Figure 4.4 E) top and bottom, respectively)), the intensity in the center of the pupil remains almost constant while the slope a curve grows more rapidly at the lobes in lowest absorption value. In other words, allowing for the absorption to be as low as possible is the key mechanism to increase the ratio between the scattering at the higher angles as compared to the lower ones. This behavior seems to be achieved mainly by the evanescent components at the interface converted to propagating modes by the presence of the isolated particle on top of the covering layer. The topographical contrast of optical detection is of significant interest for techniques that rely on differential detection principle, such as the one we propose. Equally important, by considering materials with large real part of refractive index, one can achieve large enhancements of the full far-field distribution. That is important for applications that benefit from low power of illumination.

Finally, we point out that the gradient profiles for the azimuthally polarized light are very close to those of the radially polarized ones. The major differences in directional tendency are observed in the sweep of the real value of the refractive index for both linear and circular polarization.

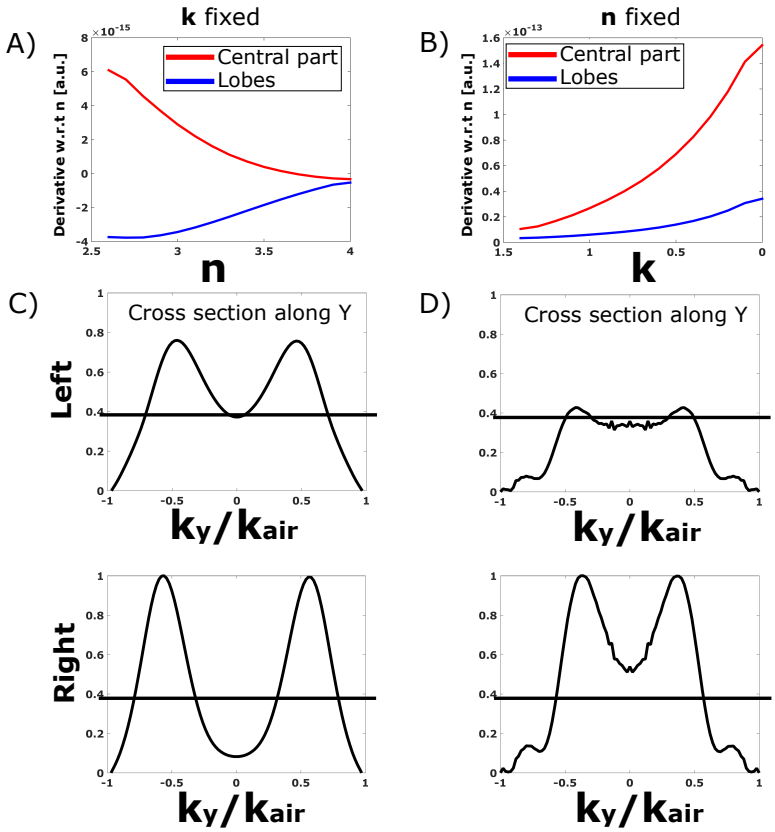


Figure 4.5: Point per point gradient of integrated electric field intensity as function of the real and imaginary parts of the refractive index for the case of linear polarization along the x-direction when A) the real part changes from 2.6 to 4 while the imaginary remains fixed at 0.133, and B) when the real part remains fixed at 2.66 while the imaginary part changes from 0.0004 to 1.4. C) and D) cross-sections of the total far-field pattern corresponding to the maximum and minimum values of the refractive indices of plots A) and B), respectively. The curves are normalized to unity by the maximum value of the distribution belonging to the particular parameter sweep. Black lines mark the center of the cross-section $k_x/k_{air} = 0$ that corresponds to the beginning of refractive indexes sweep.

Further, we calculate the gradient of the integrated far-field intensity by two zones, central and the outer part of the pupil for linear polarization as shown in the Figure 4.5. In the case of the linear polarization gradients with respect to both the real and imaginary parts of the refractive index is different from the gradients in the case of radially polarized light. Also, because of the absence of rotational symmetry, we show in Figure 4.5 C) and D) the profiles in the y-direction. For minimum (top) and maximum (bottom) value of n in C), and maximum (top) and minimum (bottom) k value in D). From the gradients in Figure 4.5 B), one can see that the behavior is such that the

energy grows faster towards the lobes and slower towards the center. But in 4.5 A), when $n \geq 3.2$, it levels out. This can be attributed to the fact that linear polarization has a mixture of s and p components in focus. Similar behavior is observed for circular polarization.

We have also studied the robustness of the enhancement process for variations on the thickness of the single layer. This is an important parameter that should be taken into account since because it is almost unavoidable that the actual thickness will vary as compared to the nominal one. The results of this analysis highlight the advantage of the evanescent wave-amplification method in terms of robustness. In Figure 4.6, the behavior of the total scattered far-field intensity as function of the covering layer thickness is shown (integration over the full pupil according to the trapezoidal method.)

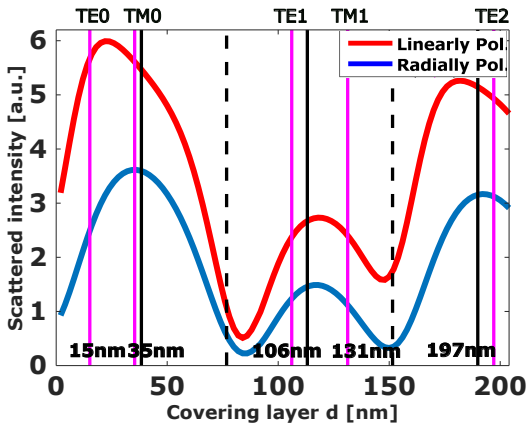


Figure 4.6: The total scattered far-field intensity integrated over the pupil for linearly polarized (red) and radially polarized light (blue). The positions of the minimum thickness for which a particular guided mode appears in the three-layer 1-D slab are indicated in purple, and the positions of maxima and minima in reflection for the normally incident light in black solid and dashed lines correspondingly.

The cover layer material in this case is TiO_2, having refractive index of $n = 2.6632$, $k = 0.00423$ [21] and layer thicknesses that vary from 0 to 200 nm. The diameter of the nanoparticle is 30 nm with the refractive index of $n = 1.58$ for the wavelength of illumination of 405 nm. From Figure 4.6, one can see that the deviation around the maximum of scattering for linear and radial polarization is about 0.84% and 0.71% per nm, meaning that the system is robust for a few nm thickness variations.

The same Figure 4.6 can be used to show the influence of the thickness of the thin layer as it becomes much larger than the values we have considered so far.The maxima and minima fluctuations may be caused by the excitation of modes by the 1D cavity that is formed by the 3-layer system without the presence of the particle. In the vertical lines found in Figure 4.6, we show the positions of the minimum thickness for which either TE modes and TM modes (in purple) appears. We pick truly guided modes, whose spatial frequency k_x/k_{air} is between the two highest indices inherent

to the 3-layer structure. The first two modes, namely TE_0 and TM_0, supported in the waveguide are in good correspondence with the maxima of the far-field scattering for linearly and radially polarized light, respectively. Further, we see the agreement between the minima of the reflected field of the 3-layer stack without a particle (black dashed lines of Figure 4.6) and the far-field scattering profiles in the presence of the particle. The reflectance of the substrate is minimized due to destructive interference in the 3-layer system. As a consequence, the scattered field of a particle reflected by the substrate is also minimized. It becomes clear that for the particular design of the enhancing layer, maximum of scattering can be achieved when the layer supports just one mode that does correspond to the spectrum of excitation (evanescent waves generated from the particle illuminated by the incident light). Further increasing of the thickness of the layer can have a drawback of falling into the local minima of reflectance or having destructive interference upon interaction with particle due to the multiple modes of the waveguide. This simulation allows us to support the idea that the enhancement in scattering is due to the evanescent wave amplification, which becomes possible in the scenario when the probing light excites the dipolar-like response of the sphere at an interface. When exciting with linearly or radially polarized light, the dipole vector will be mainly horizontally or vertically directed, respectively. Coupling is only possible if the radiation of dipole excites the mode supported by the waveguide. If the requirement is fulfilled, modes will propagate in a slab and the evanescent part of the field is maximized, and so it can be re-scattered to the far field by the particle.

Finally, we point out that the shape of the far-field distribution is wavelength dependent, and also the size of the particle influences the angles of the scattering. Therefore, the result cannot be generalized for other wavelength regimes and particle sizes. However, for particles in the regime of $\lambda/9$ and all the states of polarization, the boost related to the evanescent amplification occurs.

4.5. Experimental Results

The enhancement due to the guided mode excitation can be experimentally shown by performing the same experiment with covered and uncovered substrates with the same nanoparticle material and size. In the case of a covering-/no-layer, a relative comparison of scattering intensities is insufficient because only an experiment can reveal whether the measured scattered intensity is significantly higher than the background noise. Thus, it is difficult to predict the ultimate limit of this technique. In any case, in order to verify the validity of the calculations, we carried out several experiments. We use the setup earlier introduced and explained in Figure 2.1. The liquid crystal polarizator converter is placed before the objective to generate either radial or azimuthal polarization. To get the same incident power for the linearly polarized beam the neutral density filter with $\approx 80\%$ transmission is installed. In Table 4.3, the fabricated samples are shown. For each sample, we find an isolated nanoparticle of the nominal size and then collect the far field maps for different polarization states such as linear, radial and azimuthal.

Table 4.3: Summary of the fabricated samples. The thickness of the layers has been measured with an ellipsometer.

Sample #	Design
1	glass + 40 nm spheres
2	glass + 50 nm spheres
3	glass + 20.3 nm Ta2O5 + 50 nm spheres
4	glass + 20.3 nm Ta2O5 + 40 nm spheres
5	glass + 22 nm TiO2 + 40 nm spheres

Each far field map is the result of averaging two repeated scans (Figure 4.7). As the detection is done using the heterodyne technique [22], we modulate our diode laser with a square waveform of either $V_{pp} = 1.7$ V or $V_{pp} = 1.8$ V , so that the power on the substrate is either $P_{low} = 0.02$ mW or $P_{high} = 0.026$ mW, respectively. The higher power is used for the detection of the 40 nm spheres.

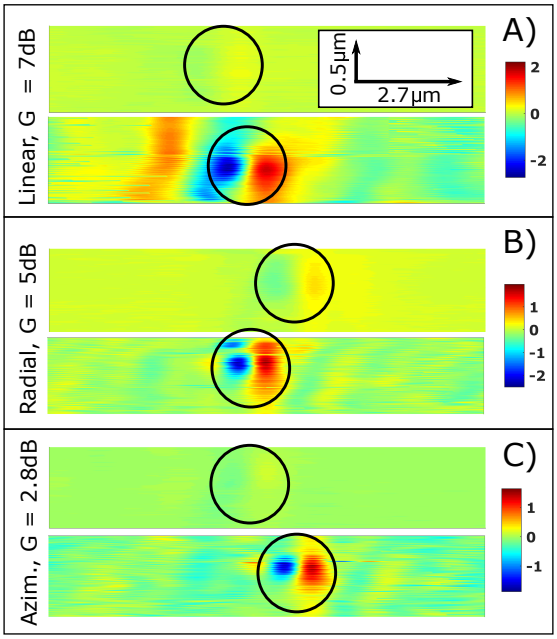


Figure 4.7: The raw far-field signal maps of the glass sample without and with the enhancement layer of Ta2O5. Isolated PSL particle, localized with black circle, of 50 nm and A) linear, B) radial, C) azimuthal input polariziation of the beam. Each map is $2.7\mu m \times 0.5\mu m$.

In order to compare the SNR with and without the thin layer, we define the SNR

gain G (in dB units):

$$G = \mathrm{SNR}_{(\text{layer})} - \mathrm{SNR}_{(\text{no layer})}. \tag{4.9}$$

The SNR (defined in Eq. 3.1 Chapter 3) is estimated based on the profile with the highest amplitude, which corresponds to a probing beam scan across the center of the spherical particle. Table 4.4 summarizes the measured gain enabled by the cover layer for different polarizations. All tested cases show that the enhancing layers do improve the detection of the low refractive index nanoparticles. The gain results are consistent with the simulations done previously that predicted that TiO_2 is a better material to enable evanescent wave amplification.

Table 4.4: SNR gain G [dB] due to evanescent wave amplification for the cases of detection of 50 and 40 nm particles on top of Ta_2O_5 or TiO_2 layers on glass (compared to glass with no layer), for various polarization configurations.

	SNR gain G [dB]		
Polarization State	d = 50 [nm]	d = 40 [nm]	d = 40 [nm]
	Ta2O5	TiO2	Ta2O5
Linearly X	7	11.8	11.5
Radially	5	9.3	6
Azimuthally	2.8	11.1	6.5

There is an evident relationship between the size of the particle and the gain that is achieved with the layer. The smaller particles become visible, compared to the very low signal in the absence of the layer, while the bigger particles have already a SNR > 18 [dB] even without the layer. The reason for that is because the scattering cross-section is larger for 50 nm particles than for 40 nm. Below, we show the measured far-field signal maps, without/with layer of Ta_2O_5 (samples number 2 and 3) for the case of $d = 50$ nm PSL particle and three input polarizations as according to the first column of Table 4.4.

The results show that the cylindrically polarized beams can contribute to a relatively big gain (G of 9.3 [dB] or 11.1 [dB]), yet, it is important to highlight that we have observed experimentally that the use of linearly polarized light is superior to cylindrically polarized light. In Figure 4.8 we show the results for sample number 5. Even though the particle of 40 nm diameter is visible when radially or azimuthally polarized beams are used (Figure 4.8 B) and C)), one can clearly see that for the linearly polarized light, the signal is much stronger (Figure 4.8 A)).

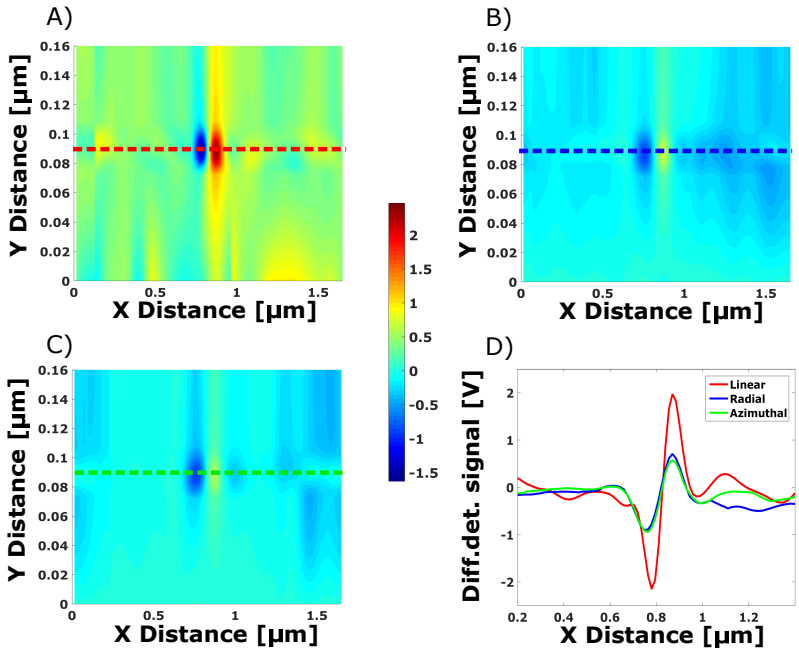


Figure 4.8: SD signal as a function of particle position w.r.t. the focused field as the particle is scanned in the x-direction. The particle is on a glass sample with the enhancement layer of TiO_2 and 40 nm PSL with the A) linear, B) radial, and C) azimuthal polarization of the probing beam. The color scheme is defined by the maximum and minimum of differential signal for the linearly polarized beam. D) The corresponding signal cross-sections are compared.

4.6. Discussion and Conclusion

In this work, the wavelength of the probing light was chosen in accordance with our experimental settings, but in general, this parameter can also be optimized together with the others. Using a resonant wavelength combined with evanescent wave amplification can extend the application of our technique to a wide range of applications. As predicted by numerical computations, our experiments demonstrate gain in the far-field due to the re-scattered evanescent field generated by the particle at a layered structure. We did not find a significant benefit to changing the polarization of the probe to cylindrical polarization, contrary to expectations. We observed a lower SNR as compared to linear polarization, even in the case of no thin layer on top of the substrate. A possible explanation could be the complexity of the experiment. Producing and focusing radial/azimuthal polarization is more difficult than linear polarization. For instance, with radially polarized light, the alignment of the beam w.r.t. the objective as well as the quality of the radial (azimuthal) polarization are extremely critical in order to obtain a very tight symmetrical focused spot.

In this chapter, we studied the enhanced far-field scattering of single spherical PSL particles deposited on top of a glass substrate covered with a thin layer of dielectric material. The key to the far-field enhancement effect is to achieve first substantial

field enhancement in the near-field. By simply looking at the interface between the air and the top surface of the bare/covered substrate, we concentrate our attention on maximizing the spatial frequencies that correspond to the enhancement of evanescent waves. The ultimate sensitivity of the detection is limited by the amount of evanescent waves that are converted to propagating waves.

From the research that has been conducted, a gain factor greater than 5 dB has been experimentally observed for different polarizations (linear, radial, azimuthal). Furthermore, in the experiment, the use of linearly polarized light has shown superior performance for the detection of $40 - 50$ nm nanoparticles. Our analysis has shown that for evanescent wave amplification, the values of n and k of the material are important, with a significant real part of the refractive index and low absorption giving higher enhancements. Our technique is robust to the deviations of the thickness of the cover layer. For the particular design of the enhancing layer, the maximum scattering is achieved when the layer supports just one guided mode. Due to the multiple modes of the waveguide destructively interfering after interaction with particle, or, simply, due to the local minima of reflectance, the high thickness of the covering layer is undesired. Due to the directionality of the scattering from the nanoparticle, applying an annular aperture at the detector can be beneficial.

The proposed technique can be successfully used in applications such as bio-nano detection and contamination detection in the semiconductor or flexible electronics industry. For instance, Si_3N_4 silicon nitride ($n \approx 2$ and almost zero absorption at 405 nm) is present as a top layer of poly-silicon pellicle [23] or in PET plastic substrates [24], if contaminated with small particles could exhibit EWA. Furthermore, there is a potential for this technique to be implemented in interference reflectance imaging techniques such as IRIS and iSCAT. In these label-free techniques, the excellent smoothness and flatness of the substrate are crucial [5]. Finally, one can also mention spinning bio disks for lab-on-chip devices, where the dipolar-like scattering of the particle near the surface is required to be maximized [25].

References

[1] Oguzhan Avci, Celalettin Yurdakul, and M. Selim Ünlü. Nanoparticle classification in wide-field interferometric microscopy by supervised learning from model. *Appl. Opt.*, 56(15):4238–4242, 05 2017.

[2] D Kolenov, HP Urbach, and SF Pereira. Effect of polarization in evanescent wave amplification for the enhancement of scattering of nanoparticles on surfaces. *OSA Continuum*, 3(4):742–758, 2020.

[3] Judith Su, Alexander FG Goldberg, and Brian Stoltz. Label-free detection of single nanoparticles and biological molecules using microtoroid optical resonators. *Light: Science & Applications*, 5:e16001, 01 2016.

[4] A. Salehi-Reyhani. Evaluating single molecule detection methods for microarrays with high dynamic range for quantitative single cell analysis. *Scientific Reports*, 7(1), 2017.

[5] F. Ekiz-Kanik, D. D. Sevenler, N. L. Ünlü, M. Chiari, and M. S. Ünlü. Surface chemistry and morphology in single particle optical imaging. *Nanophotonics*, 6(4):713–730, 2017.

[6] Derin Sevenler, Oğuzhan Avci, and M. Selim Ünlü. Quantitative interferometric reflectance imaging for the detection and measurement of biological nanoparticles. *Biomed. Opt. Express*, 8(6):2976–2989, 06 2017.

[7] Shuming Nie and Richard N. Zare. Optical detection of single molecules. *Annual Review of Biophysics and Biomolecular Structure*, 26(1):567–596, 1997.

[8] Oguzhan Avci, Maria I. Campana, Celalettin Yurdakul, and M. Selim Ünlü. Pupil function engineering for enhanced nanoparticle visibility in wide-field interferometric microscopy. *Optica*, 4(2):247–254, 02 2017.

[9] S. Roy, S. F. Pereira, H. P. Urbach, Xukang Wei, and O. El Gawhary. Exploiting evanescent-wave amplification for subwavelength low-contrast particle detection. *Phys. Rev. A*, 96:013814, 07 2017.

[10] Omar El Gawhary, Nick J Schilder, Alberto da Costa Assafrao, Silvania F Pereira, and H Paul Urbach. Restoration of s-polarized evanescent waves and subwavelength imaging by a single dielectric slab. *New Journal of Physics*, 14(5):053025, 05 2012.

[11] Y. Lu and S. C. Chen. Nanopatterning of a silicon surface by near-field enhanced laser irradiation. *Nanotechnology*, 14:505–508, 05 2003.

[12] David G. Seiler Zhiyong Ma. *Metrology and Diagnostic Techniques for Nanoelectronics*. Pan Stanford, 10 2016.

[13] O. El Gawhary, M. C. Dheur, S. F. Pereira, and J. J. M. Braat. Extension of the classical fabry–perot formula to 1d multilayered structures. *Applied Physics B*, 111(4):637–645, 2013.

[14] Xiuhong Wei, Arthur J. Wachters, and H. Paul Urbach. Finite-element model for three-dimensional optical scattering problems. *J. Opt. Soc. Am. A*, 24(3):866–881, 03 2007.

[15] Stéphane Larouche and Ludvik Martinu. Openfilters: open-source software for the design, optimization, and synthesis of optical filters. *Appl. Opt.*, 47(13):C219–C230, May 2008.

[16] Max Born, Emil Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7 edition, 1999.

[17] W. Sellmeier. About the co-oscillations of the body particles excited by the etheric oscillations and their reaction on the former, especially to explain the dispersion and its anomalies. *Annalen der Physik*, 223(11):386–403, 1872.

[18] Shojan P Pavunny, Reji Thomas, and Ram S Katiyar. Cauchy-urbach dielectric function modeling of amorphous high-k lagdo3 films. *ECS Transactions*, 45(6):219, 2012.

[19] Jan W. Gooch. *Cauchy's Dispersion Formula*, pages 125–125. Springer New York, New York, NY, 2011.

[20] S. Roy, A. C. Assafrao, S. F. Pereira, and H. P. Urbach. Coherent fourier scatterometry for detection of nanometer-sized particles on a planar substrate surface. *Opt. Express*, 22(11):13250–13262, 06 2014.

[21] Thomas Siefke, Stefanie Kroker, Kristin Pfeiffer, Oliver Puffky, Kay Dietrich, Daniel Franta, Ivan Ohlídal, Adriana Szeghalmi, Ernst-Bernhard Kley, and Andreas Tünnermann. Materials pushing the application limits of wire grid polarizers further into the deep ultraviolet spectral range. *Advanced Optical Materials*, 4(11):1780–1786, 2016.

[22] D. Kolenov, R. C. Horsten, and S. F. Pereira. Heterodyne detection system for nanoparticle detection using coherent Fourier scatterometry. In Peter Lehmann, Wolfgang Osten, and Armando Albertazzi Gonçalves Jr., editors, *Optical Measurement Systems for Industrial Inspection XI*, volume 11056, pages 336 – 342. International Society for Optics and Photonics, SPIE, 2019.

[23] Derk et al. Brouns. Nxe pellicle: offering a euv pellicle solution to the industry. *Proceeding*, page 97761Y, 03 2016.

[24] K.S. Kim, K.H. Kim, Y.J. Ji, J.W. Park, J.H. Shin, A.R. Ellingboe, and G.Y. Yeom. Silicon nitride deposition for flexible organic electronic devices by vhf (162 mhz)-pecvd using a multi-tile push-pull plasma source. *Scientific Reports*, 7(1), 2017. cited By 3.

[25] David D. Nolte. Invited review article: Review of centrifugal microfluidic and bio-optical disks. *Review of Scientific Instruments*, 80(10):101101, 2009.

5

Highly-sensitive laser focus positioning method with sub-micrometer accuracy.

In this chapter, we analyze the effect of defocus on particle detection. Firstly, in Section 5.1, we describe the degradation of the particle signal SNR associated with non-ideal focusing. Curiously, the fact that the signal is very sensitive to the focus position leads us to propose a novel method to determine the focus position. Thus, in section 5.2 we describe the method by defining an S-curve for focus determination. We numerically study the phase in the near-field close to the interface and investigate the uncertainty in determining the position of the surface under the particle. In Section 5.3, we investigate the sensitivity of the proposed method and put it in the context of the existing focus-determination techniques. In the last section we present discussions and conclusions.

Parts of this chapter have been published in Meas. Sci. Technol. **31** (6) 064007 (2020) [1]

5.1. Introduction

Focused light plays a vital role in modern technology in the fields of optical lithography, micro-machining, optical data storage, nanostructure characterization and biology [2, 3]. The inability to define and maintain the focal plane position on the sample or surface results in degradation of resolution [4], non-optimal energy use of the focused field [5], damage of the sample [6], failure to store data [7, 8], uncertainty in the localization of cells or lack in reproducibility of biological results [9, 10].

While the problem of determining the focus position in an accurate way is essential for many applications such as the ones mentioned above, there is no single general solution, and in the literature, one finds various techniques that are applied depending on the limitations and other conditions of the experimental setup. These limitations might include sensitivity to vibrations and shocks, the ability to work in-line scan speed, ease of use, etc.

For coherent sources of illumination, monochromatic confocal systems are commonly used to measure the surface position or to simply define the focus position [6, 11]. In a confocal system, the light is focused on the object plane and this plane is imaged on the point detector (usually an optical fiber) or a pinhole.The $z-$scanning should return the peak intensity in reflection when the focal position matches the interface.

This chapter proposes a novel sensitive focus finding method based on the detection of nanoparticles on surfaces using CFS. As a spin-off of the method, the position of the surface under the particle can be determined with sub-micrometer accuracy in a non-contact manner. With the aid of differential signal readout, we can sensitively localize the position of the nanoparticle both in horizontal and vertical planes. While the idea of detecting the focus position on the surface with the aid of differential detection is not new [7, 12], however, to the author's best knowledge, exploiting an isolated particle at the surface to find the surface position has never been proposed. Experimentally, the presented focus-finding technique allows to generate the error S-curve with a linear region that is a few times higher than the corresponding vectorial Rayleigh range $S \approx 3.8Z_r$ for the case of a PSL of 50 nm in diameter deposited on a silicon wafer and a linear region of $S \approx 5.6Z_r$ for a glass substrate containing a 100 nm PSL nanoparticle. Further, three sizes of PSL nanoparticles, namely 200, 100, and 50 nm in diameter are tested, where it was observed that the increase in sensitivity was inversely proportional to the particle size.

5.2. Methods

5.2.1. Measurement Approach

In CFS, the position and size of the nanoparticle can be obtained from the raster scanning procedure over the surface of interest. When the sample on the piezo-stage is scanned in a raster fashion (line scan along the x-axis with an increment in the y-axis between the lines), the scattering from a single isolated particle contributes to the differential signal of the split detector (L-R) over multiple scanning lines (signal group) as shown schematically in Figure 5.1 A).

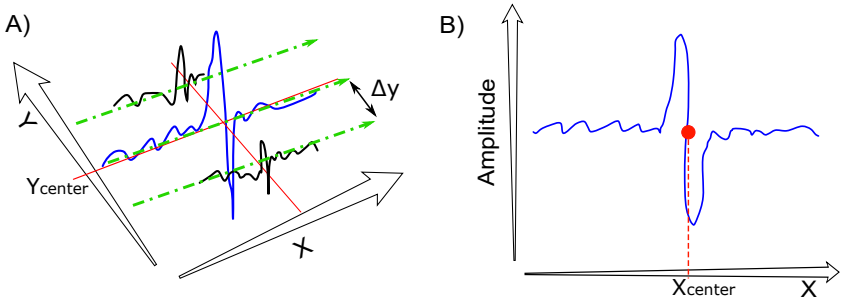


Figure 5.1: A) Schematic showing the raster scanning of the substrate containing one isolated nanoparticle. The scan is line per line, along the x-direction. B) Differential signal of a single scanning line along x through the spherical particle. The zero signal refers to the center of the particle.

The full area of the scan map is $A = X * Y\ [\mu m^2]$, where X and Y are chosen to be a few times larger than the size of the scanning spot. Importantly, a step displacement Δy of the stage along y-direction is recommended to be smaller than the diameter of the sphere. The orthogonal direction, where the sampling of the signal should be high to give a valid representation in the time span, defines the width of the scan in the x-direction. If there is only one of such group of signals in the scanned area A, we can assume that the particle is isolated. Within one scanning map, the Y_{center} position is defined when the signal group amplitude is maximum, and the X_{center} position is attributed to the zero-crossing of the corresponding particle profile (Figure 5.1 A) and B)).

The working principle of the focus determination system is described via flowchart Figure 5.2 A). Firstly, the downward scanning axis z needs to be aligned with the "center" of the particle. The coordinate pairs in the xy plane referring to the center of the particle in the lateral direction is obtained by averaging the coordinate X, Y in all z planes that corresponds to zero intensity at the detector in (see Figure 5.1 B)), i.e., $(X_{ref}, Y_{ref}) = \left(\overline{X_{center}}, \overline{Y_{center}}\right)_{Zj}$. The next two steps are: 1) move on the optical axis to a plane above the particle defined as Z_0 and 2) performing single line scans in the x-direction around the point (X_{ref}, Y_{ref}) for different values of Z. Collecting only one profile at a time allows us to keep the method time-efficient. The stage moves downwards the predetermined amount of lines $nlines$, which has to be big enough to go through the optimal focus. In our experiment, the steps in the z-direction were $\Delta z = 5$nm (closed-loop resolution of P-622.ZCD is 1nm), rendering the complete through-focus distance of $\Delta Z = nlines * \Delta z$. A scheme of this procedure is shown in Figure 5.2 B). Essentially, the scanning is performed in a raster fashion, but now in the plane xz. The obtained differential signal for each scan line is crossing the DC-level (set to zero) when the particle is centered w.r.t. the focused spot and is positive or negative when the particle is either on the left or on the right of the focused spot along the x-axis. It has also been observed that the relative intensity of the differential signal corresponding to the left or right side changes with the chosen Z plane.

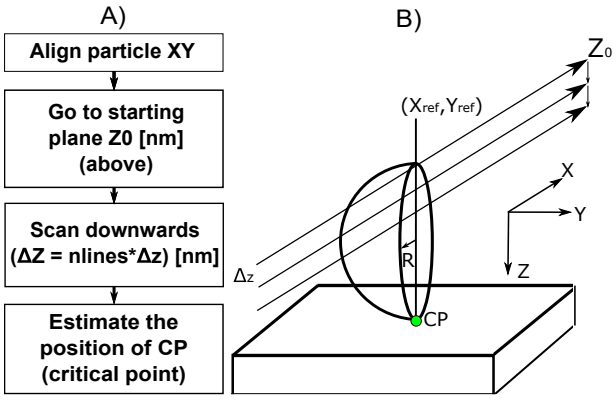


Figure 5.2: A) Flowchart of the proposed focus finding method. B) Spot moves to positive z as the piezo steps in the opposite direction. For each position along z, a line scan is performed along x. The plane xz is at the middle of the sphere of radius R.

At one particular Z plane, the maximum intensities of the differential signal for X positions located on the left and on the right of the center of the particle are balanced ("same" intensity, opposite signal), defined as the CP (critical point) in Figure 5.2 B). At other z planes, the maximum left/right intensities are different, i.e., the signal is unbalanced. The distance from the initial plane to the position of CP can be computed by adding the initial position with the distance to the critical point and including the error.

In Figure 5.3 A) we show schematically these situations (see signals on the 3 squares) as a function of the defocus position. Also in Figure 5.3 A) we show that if we plot the unbalance between the right and left maximum intensities, defined as $|I_{max}| - |I_{min}|$ as a function of the z displacement, one obtains a S-curve.

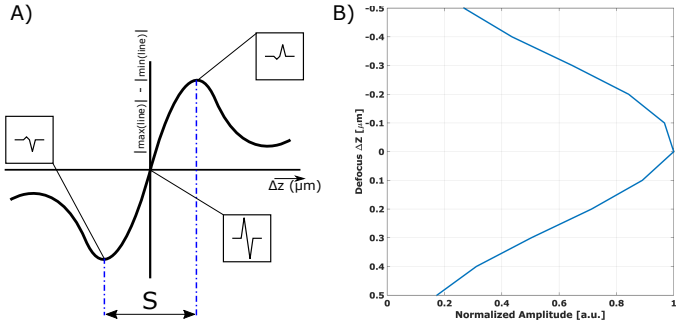


Figure 5.3: A) The measure of unbalance of the lateral scan line as a function of the defocusing ΔZ. S-curve for the through focus error signal. B) The maximum in the amplitude of the simulated through focus differential signal occurs when the focus is located under the spherical particle, i.e., defocus $\Delta Z = 0$ (y-axis).

This focus error signal changes from its maximum negative to its maximum positive value over a distance that typically is a few vectorial Rayleigh ranges and the slope depends inversely on the diameter of the particle.

The far-field nanoparticle scattering is simulated for the defocus distance of 1μm and for linearly polarized focused spot. In Figure 5.3 B), the maximum in the amplitude of the differential signal that is obtained for each x-scan at one fixed defocus position is plotted against the defocus position. In the plot, zero defocus means that the focal plane is set to interface (air/substrate). As it can be seen in the figure, the maximum amplitude of the differential signal occurs when the focal plane is set at the interface, i.e., under the particle. This result suggests that the zero crossing of the S-curve is not only a position to keep the instrument in-focus for nanoparticle detection but also the location of the surface under test in the z-direction. We also observed that the smaller particles are, the steeper is slope of S-curve. More details about the simulations that lead to the result in Figure 5.3 B) are given in the next section.

5.2.2. Position-dependent phase

In order to understand the influence of the focus position in the nanoparticle/interface scattering, we have performed rigorous 3D vectorial simulations. In the first set, we calculated the focused field without a nanoparticle on a interface using the Richards and Wolf integral. The implementation of the spot is explained in details in Ref. [13]. The results shown in Figure 5.4 are calculated for linearly polarized (along the x-axis) uniform amplitude focused field, where on the left colunm we plot the modulus of the amplitude of the $|(E_x(z)|$ and the $|(E_z(z)|$ components and on the right colunm their corresponding phases as the function of $-5\lambda \leq Z \leq 5\lambda$. The numerical aperture is $NA = 0.9$ and the wavelength $\lambda = 405$ nm. The interface is between air $(Z < 0)$ and silicon $(Z > 0)$ and the focal plane is set at the interface $(Z = 0)$. Notice that in the case of the x-linearly polarized focused spot, the E_y component is not shown since this is negligible. To determine the phase change, firstly, we estimate the on-axis (z) positions of the first two minima close to the interface in the amplitude of the focused field components. Secondly, we compute the phase difference between the minima based on the corresponding phase slices. For the E_x component, we estimate that the π phase difference occurs over the distance $\Delta Z = 1.049\mu$m and for the E_z component $\Delta Z = 0.73\mu$m. This phase change suggests that when the particle on the silicon surface will be moved along the z-direction, the collected signal will have the reverse in the polarity over the certain vertical distance. It is important to note that we only show the slices of the phase map in the normal direction to the interface. The far-field result of the near-field focused spot interaction is more complicated and accounts for every angular direction within the numerical aperture.

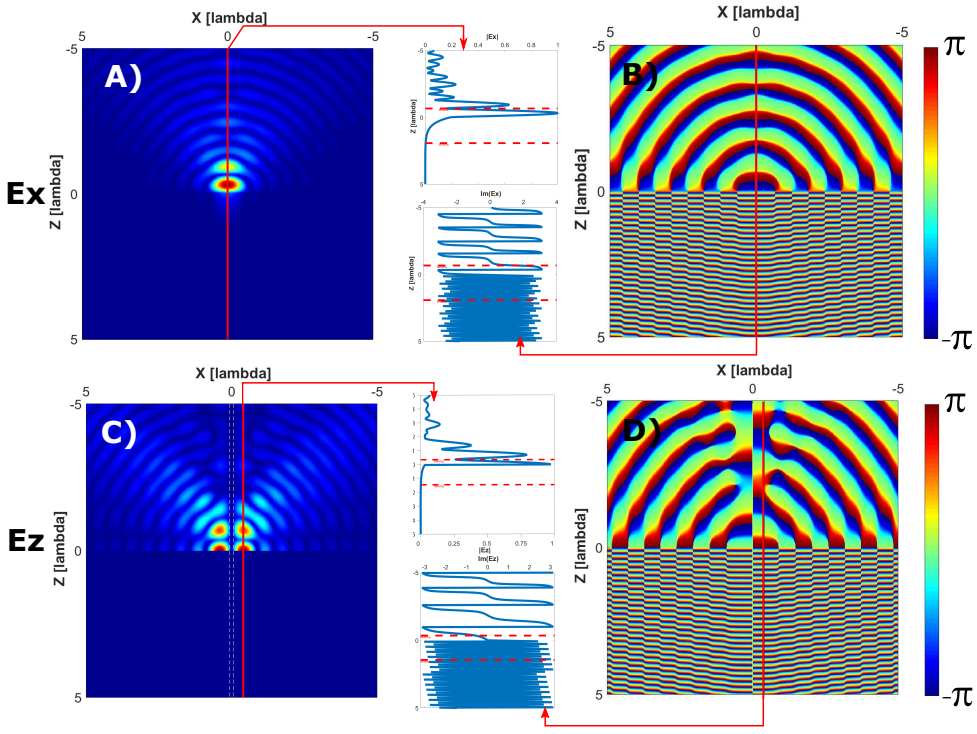


Figure 5.4: Simulated amplitude and phase distributions of the focused field in the xz plane. The middle part of each panel represents the cross-section of either amplitude or phase (red arrows) and dashed lines represent the minima of amplitude. The absolute values of the amplitude E_x in A) and E_z in C), the corresponding phase in B) and D). The $NA = 0.9$, the wavelength $\lambda = 405$ nm, and the interface is between air $(Z < 0)$ and silicon $(Z > 0)$.

Further, to demonstrate the interaction of the particle with the position-dependent phase of the focused field we will consider two situations. Firstly, we set the focal plane right above a 100 nm diameter PSL particle, i.e., at $Z = -0.11\mu$m, as shown in Figure 5.5 A) and C), and secondly, the focal plane is set inside the silicon slab at $Z = +0.11\mu$m in B) and D). Further, we show the total near field for the E_x and E_z components if the nanoparticle is displaced in the x-axis by -0.4μm w.r.t. the center of the beam, which is located at $X = 0$. We are interested in the particle position that corresponds to the maximum in the differential profile such as when $X = -0.4$ μm. Here, the numerical analysis is based on the FDTD method by Lumerical, which allows simulating a spherical particle on a surface, illuminated with a focused beam. As the incident field, we use the focal spot with the linear polarization.

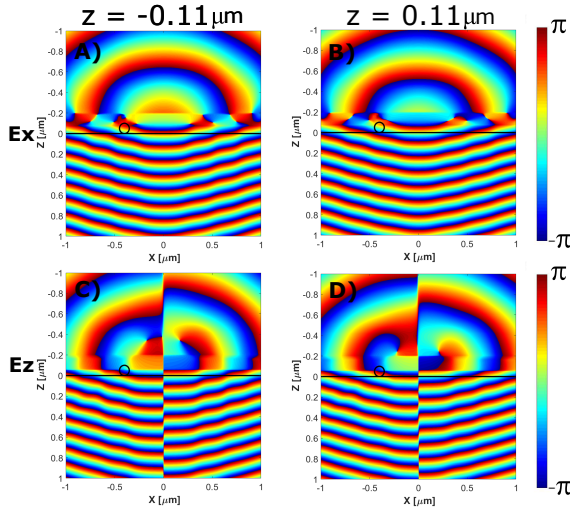


Figure 5.5: The phase distributions in the xz plane of $100n$m PSL particle on top of silicon. Left panel A) and C) the focal plane is set at $Z = -0.11\mu$m. Right panel: B) and D) the focal plane is set inside the slab at $Z = 0.11\mu$m. The particle in the x-axis is positioned at $X = -0.4\mu$m

The opposite (at $Z = -0.11\mu$m and $Z = 0.11\mu$m) phase distributions for the x-component $arg(E_x)$ for the two extrema look similar. The $arg(E_z)$ component, on the contrary, changes both in magnitude and spatial distribution. We can conclude that, presumably, the phase changes in the z-component of the incident field are responsible for the scattering from the symmetric object producing the unbalanced amplitude at the far-field.

We proceed with simulation of an aspherical particle, i.e., an oblate spheroid. Oblate spheroid with half axes of horizontal $r1 = r2 = 0.05\mu$m and vertical $r3 = 0.0125\mu$m made of PSL that is placed on the surface of a silicon wafer.

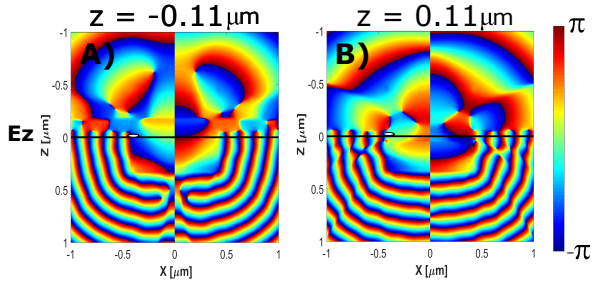


Figure 5.6: The E_z phase distributions in the xz plane of an elliptical PSL particle $r1 = r2 = 0.05\mu$m and vertical $r3 = 0.0125\mu$m on top of silicon. A) $Z = -0.11\mu$m and for B) the $Z = 0.11\mu$m. The particle in the x-axis is positioned at $X = -0.4\mu$m

Similarly to the Figure 5.5 we look at near-field E_z phase distributions above and

below the focus, but now with the particle shape changed to an ellipsoid (Figure 5.6.) Similarly to the Figure 5.5 the phase changes in the z-component are more pronounced than the other polarization components, by looking at different z-positions both in magnitude and spatial changes of the phase of the near field are observed. If compared to the results from Figure 5.5, the change of particle shape renders a non-negligible effect on the phase distribution. Since the proposed focus-finding technique relies only on the difference between the peaks in the measured signal $|I_{max}| - |I_{min}|$ as a function of the z-displacement, non-spherical particles could also be considered in our technique.

5.2.3. Uncertainty Estimation

The distance from the initial position, somewhere above the surface, to the final position, at the surface, is defined from the zero-crossing of the S-curve. The exact location of this point depends on multiple sources of measurement uncertainty. The process of calculating the uncertainty U includes two steps. Firstly, we determine directly measured uncertainties and also take into account the uncertainties given by manufacturers. Secondly, we combine those individual uncertainties in root sum of squares. We use the type A evaluation process from guide to the expression of uncertainty in measurement (GUM) [14] to estimate the axial shifts along x, y, and z of the zero crossing position Z_{cp}. To find the "center" of the sphere, the maps over multiple vertical positions are acquired and the average reference position that corresponds to the maximum amplitude of the differential signal is stored. First, two uncertainties associated with this process are standard deviations σ_x and σ_y of separately X and Y coordinates, that describe the fluctuation versus the reference pair (X_{ref}, Y_{ref}). The σ_x has no contribution to final uncertainty as long as $\sigma_x < \Delta X/2$, where ΔX is total width of the scan along x (typical number is $\Delta X = 20\mu$m.) Hence, for our method, the effect of misaligning the particle with respect to the focused spot in the horizontal plane is described by σ_y. Deviation from the particle's true center described by σ_y translates to the apparent vertical off-center distance to the surface that is shorter than if it would had been estimated exactly through the center. Next, based on the repeated one line scan measurements along the z-axis line, from the resulting S-curve, the σ_z can be estimated. For the final calculation of U, the σ_z corresponding to the zero of curve Z_{cp} is used. Both σ_x and σ_y are calculated by averaging the $\left(\overline{X_{center}}, \overline{Y_{center}}\right)_{zj}$ coordinates of ten signal maps across the total vertical distance of one micrometer, starting close to the focus position, with steps of 100 nm. The standard deviation σ_z is computed for the zero-crossing points of the S-curves by averaging 4 repeated profiles, each with a step of 5 nm between parallel scanning lines over the distances that are different for different particle diameters, ranging from $1.5\,\mu$m to $2.6\,\mu$m. Following that, from the data sheet for the z-translator stage (P-620.ZCD), we take the approximated accuracy error on the linearity of the z-displacement as double. For example, a 0.02% linearity for the full range of $50\,\mu$m is a 10 nm maximum deviation. Hence, the approximated accuracy error is 20 nm pk-pk or $\sigma_{s_z} = 0.02\,\mu$m. In order to build up the error-curve profiles, the raw sampled data from the split detector is interpolated using Friedman's algorithm [15] with high coefficients of determination for all the studied particles diameters of 200, 100 and 50 nm ($R^2_{100\,and\,50} > 0.95$ and $R^2_{200} > 0.85$).

Moreover, as a sample is fixed on top of the stack that is comprised of a z-translator stage attached to a xy-translator stage, it is important to check the correlation between the axial contributions of uncertainty. We choose a DIC (Digital Image Correlation) method [16, 17] applied to 2D CCD images of the reflected focused spot at different axial positions, with these being close to the true focus. The core of the algorithm is to take two closely separated ($\Delta z = 5$ nm) images and take the normalized cross-correlation between the subsets of corresponding numerical arrays A and $A^{\Delta z}$. The raw intensity images at the CCD 1024×1280 size are cropped to the central pupil part of 550×641 or generally $k \times m$ pixels. For a single reference in-focus cut-out image I the $1 \times (n+1)$ vector $\boldsymbol{r_{yz}}$ is computed based on cross-correlations with shifted cut-outs $I^{\Delta z}$ in $y-$direction (circular shifts of array along first dimension). The dimensions of correlated cut-out images are $n \times n$ and the range of y- displacement is $[-n/2, n/2]$, where $n = 20$. The normalized cross-correlation (*corr2* function in MatLab) is defined as

$$r = \frac{\sum_i \sum_j \left(I_{ij} - \bar{I}\right)\left(I_{ij}^{\Delta z} - \overline{I^{\Delta z}}\right)}{\sqrt{\left(\sum_i \sum_j \left(I_{ij} - \bar{I}\right)^2\right)\left(\sum_i \sum_j \left(I_{ij}^{\Delta z} - \overline{I^{\Delta z}}\right)^2\right)}}. \tag{5.1}$$

The shifted cut-outs are necessary to check whether the correlation throughout pairs of subsets grows with respect to the non-shifted position. If this is the case, the vertical displacement of the wafer surface by Δz translates to the displacement Δy. To stay on the conservative side of the uncertainty estimation the maximum value of the cross-correlation vector of the single reference cut-out image $r_{yzmax} = max\{\boldsymbol{r_{yz}}\}$ is stored in the cross-correlation vector $\boldsymbol{c}$ over complete field of view and that is formed based on comparing total of $t = (\lfloor 2 * k/n \rfloor - 1) * (\lfloor 2 * m/n \rfloor - 1)$ reference cut-outs with their y-shifted counterparts. Finally, we compute the correlation coefficient R_{yz} as :

$$R_{yz} = \left(\overline{\operatorname{arctanh}(\boldsymbol{c})}\right)^{-1} = \left(\overline{\operatorname{arctanh}\left(r_{yzmax}^{t}\right)}\right)^{-1} \tag{5.2}$$

Based on the two images of a large random defect $> 5\,\mu$m in the silicon wafer, we estimate the cross-correlation between the y- and z-displacement as $R_{yz} = R_{zy} = 0.6331$.

We put together the uncertainty contributions according to the following equation for the combined standard uncertainty

$$U_{partc} = \sqrt{\sigma_y^2 + \sigma_z^2 + 2R_{yz}\sigma_y\sigma_z + \sigma_{s_z}^2}, \tag{5.3}$$

In addition to the standard uncertainties reported with the partial correlation U_{partc} between y and z contributions, we include more conservative estimates assuming the

full (positive) correlation [18].

$$U_{posc} = \sqrt{(\sigma_y + \sigma_z)^2 + \sigma_{s_z}^2}. \tag{5.4}$$

Finally, to establish a more reasonable accuracy of the technique, we use the expanded uncertainty U given by $U = k * U_{partc/posc}$, where k is coverage factor and $U_{partc/posc}$ is the combined standard uncertainty. We take the value of the coverage factor to be $k = 2$, which is equivalent to approximately 95% coverage for a normal distribution. Importantly, to collect the data, we have used two settings of optical power; the peak power at the objective (before being focused on the substrate) is either $P_{low} = 0.58$ mW or $P_{high} = 1.64$ mW. For the cases of 200, 100 nm PSL particles on silicon and 200 nm PSL particles on glass, we have used the power P_{low}, while for the 50 nm PSL particles on silicon and 100 nm PSL particles on glass, we have used P_{high}. For most of

Table 5.1: Expanded uncertainty contributions for computing the surface position. Upper panel: PSL particle on silicon wafer and Lower panel: PSL particle on glass.

PSL diameter [nm]	$2 * U_{partc}$ $[\mu m]$	$2 * U_{posc}$ $[\mu m]$
	Silicon surface	
200	± 0.7392	± 0.8110
100	± 0.2301	± 0.2494
50	± 0.0864	± 0.0936
	Glass surface	
200	± 1.007	± 1.1833
100	± 0.3067	± 0.3374

the demonstrated cases, at wavelength of $\lambda = 405$ nm, we estimate sub-micrometer accuracy in finding the position of the surface (Table 5.1). The exception is the 200 nm PSL particle on top of the glass surface, where the uncertainty is slightly above 1 μm. It is thus not recommended to use such a big diameter for focus-searching purposes with materials of low reflection. The obtained results are comparable to focus searching methods used in micro-machining applications such as the one based on a nonlinear harmonic generation [19].

5.3. Results

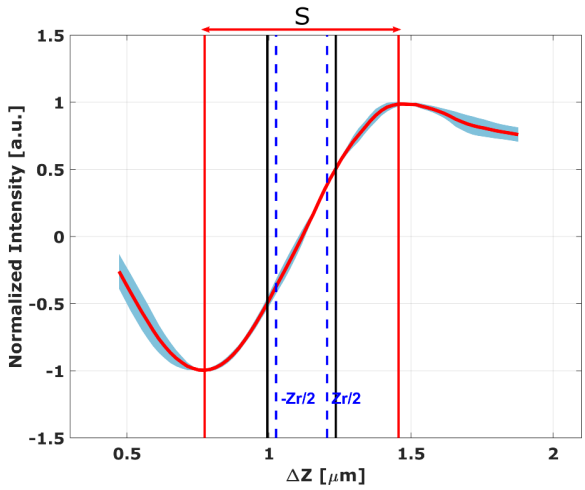


Figure 5.7: Measured mean S-curve of the through focus scan of a (nominal) 46 nm PSL sphere on top of a silicon wafer (red curve). Sky blue color indicates one standard deviation with respect to the mean over 4 data sets. Black vertical lines define the region where 50% Pk-Pk of the curve lies and red vertical lines the region over which the slope S is estimated as compared to the one Rayleigh range (blue dashed lines). For this data, the slope $S \approx 3.8Z_R$.

Figure 5.7 shows the S-curve obtained from the through focus signal for a sample of monodisperse polystyrene calibrated spheres (PSL) that have been spin coated on top of a silicon wafer. For this plot, the scanning with the focused spot is done through an isolated sphere of nominal diameter $d = 46 \pm 2$ nm. For the acquired S-curve, 100% (peak-to-peak) of the slope lies within $S \approx 3.8Z_r$, where Z_r is the Rayleigh range calculated from the vectorial diffraction theory according to the formula $Z_r = \lambda/4[1 - cos(\alpha)]$, where $\alpha = asin(NA)$. N.B. This Rayleigh range is almost three times smaller than paraxial one (Section 2.3 of this thesis) [20]. Notice how the middle 50 percent of the curve is slightly bigger than the Rayleigh range. Remarkably, in the right half of the S-curve, the region after its maximum positive value decreases slower than the corresponding region in the left half of the S-curve. The most likely explanation for this result is the scattering from the background that increases balance in the collected signal as we scan below the focus, deeper under the particle. Further, we investigate the effects of the particle diameter on the sensitivity of the acquired S-curve (see Table 5.2). The motivation for this is to show that this method is compatible with several particle sizes as it is not always possible to add a particle with a known size. Secondly, as our prototype includes custom-made electronics for the detector, which allows us to detect particles of < 80 nm (low dark noise and with a design to minimize pick-up from environment), in contrast, when using an off-the-shelf photodetector, there will be considerably more noise and it will be hard to detect small scattering signals from the spheres < 100 nm in diameter. In this case, slightly larger particles could be used.

Table 5.2: Comparison between 200, 100 and 50 nm PSL nanoparticles studied in the through focus measurements. The middle 100% of the S-curve for each measurement fall within $S[\mu m]$.

PSL diameter [nm]	The linear region of S-curve $S \pm \sigma$ at 100% Pk-Pk $[\mu m]$	
	Silicon surface	Glass surface
200	1.1618 ± 0.0167	1.2817 ± 0.0429
100	0.8016 ± 0.0104	1.0116 ± 0.0033
50	0.68273 ± 0.0088	

For the sensitivity comparison, the S-curves were acquired for the same experimental conditions of sampling, scanning speed, vertical and horizontal displacement steps. Each profile is based on the averaging of at least four data sets for each particle size. Evidently, the slope of the S-curve increases when smaller particles are studied. For a sphere of 50 nm diameter, as compared to other samples, the variation of the linear region extent, as projected to the vertical axis $\Delta Z \pm \sigma$, the mean distance S is the smallest. The main reason is shorter acquisition time needed to record multiple data sets for averaging, during which, the whole system is less perturbed as compared to longer scans for larger particles. Next, the decrease of the particle diameter does not linearly improve the sensitivity of the acquired curve. As from the electromagnetic point of view, we are not dealing with physical objects but rather with collection of emitting sources. The behavior can be better understood from a perspective of scattering by a sphere on a substrate [21].

Table 5.3: Sensitivity comparison Ratio G $= T'/T$ of the presented method to various implementations of the known astigmatism method. T',T are the ratio of the linear region of each S-curve to the corresponding Rayleigh range for each implementation. The reference T used for the comparison is the value obtained in this chapter for a 50 nm particle, which is $T = 3.8Z_r$.

Linear region S-curve S $[\mu m]$	Details	Rayleigh range Zr' [nm]	Ratio T' = S/Zr'	Ratio G = T'/T	Reference
6	$\lambda = 633nm$, NA = 0.6	791.25	7.6	2	[22]
4	$\lambda = 405nm$, NA = 0.65	421.7	9.48	2.5	[23]
3	$\lambda = 405nm$, NA = 0.85	214	14.02	3.67	[24]
2	$\lambda = 405nm$, NA = 0.85	214	9.34	2.45	[25]

For a horizontal dipole-like particle at interface, the slope in the scattering response

as a function of the particle size is fairly small for the particles in the domain $< \lambda/4$. At the same time, a change in slope is rather drastic for the bigger particles $> \lambda/4$. Thus, in order to benefit from tiny particles in focus searching purposes, the sensitivity of the detector should be suitable to record small changes in the scattered intensity. In addition, we observed that the slope of the S-curve for a 50 nm particle is close to the slope of a 100 nm sphere $S_{100nmPSL} = 0.8$, $S_{50nmPSL} = 0.68$. Thus, we expect that the sensitivity will not improve dramatically if < 50 nm particles would be used. It follows from Table 5.3 that the sensitivity of the proposed technique outperforms a set of results demonstrated by several implementations of one of the most commonly used techniques for focus control, namely, astigmatism.

For the published results mentioned in Table 5.3, the extents of the linear region S for the corresponding S-curves (first column) were taken directly from the publication, while the corresponding Rayleigh ranges were computed by the authors of this chapter. Since the benchmark results were based on systems with $NA \geq 0.6$, the formula for the Rayleigh range was also computed according to the vector diffraction theory.

5

5.4. Discussion and Conclusions

This chapter shows that high sensitivity in focus positioning can be achieved using the scattering from a nanoparticle on a surface. An immediate consequence is that the user might need to deposit a particle on the surface of interest, which may be unwanted. However, it is important to highlight that many application areas may not require depositing the isolated spheres but rather may use nanoparticles that could already be present in the sample. For example, in the domain of the semiconductor manufacturing, silicon wafers that come immediately from the producer are not completely clean and do contain particles in the size domain from 30 to 100 nm. Additionally, during the lithography process, contamination is unavoidable [26]. If the assumption that the contamination has a shape close to a sphere or an oblate spheroid can be made, the technique is directly applicable. In the case of biological samples, metal particles are sometimes present to stimulate resonance, so the smaller features of the neighboring molecules or viruses become visible [27]. Next, in the fine machining stages of laser material processing, the surface will have low roughness, but locally tiny scratches, pits or dust [28]. All these little isolated features could be used in our method.

We emphasize that in order for the technique to work, the particle diameter should be smaller than the depth of the focus of the given optical system. In our case, the $DOF \approx 360$ nm whereas the maximum diameter of the studied particle was 200 nm. The DOF is defined as double the Rayleigh range $DOF = 2Z_r$. On the other hand, the smallest suitable particle is defined by both resolution of the vertical translation and the detection sensitivity of the technique itself. In order to build up the error curve, the differential signal should be recorded at least, at three vertical positions, when in practice one or two orders higher sampling is necessary. Further, the studied sphere upon interacting with the focused spot should produce sufficient scattering to overcome the inherent experimental noise, specifically the noise floor at the detector. The strength of the collected signal, given a constant power of illumination and fixed particle diameter, depends both on the reflectivity of the sample and the material of the particle.

This technique could be extended to determine the tilt of a large surface such as a wafer by having a few nanoparticles distributed on the surface (spherical or elliptical), and determining the position of the surface at each point. In principle, to determine the 2D tilt, three spheres distributed as a triangle would suffice.

We observed that the asymmetry in scattering from the nanoparticle that occurs when the focal plane is moved away or towards the surface could be used to generate the S-curve error signal and consequently the position of the focal plane w.r.t. the reflecting surface. According to rigorous numerical simulations, we found that the zero in the S-curve occurs when the focal plane is set at the plane of the surface, i.e., directly under the nanoparticle. The obtained experimental results indicate that method is suitable for both silicon and glass surfaces. The reflectance of the surfaces, estimated for the contributing angles within the NA, varies between 20% and 72% for the glass and silicon respectively. The achievable accuracy of the method is in the sub-micrometer range with major contribution due to the translation stage (reported as the 2σ measurement uncertainty.) Moreover, several comparisons between the sensitivity of the presented method and various implementations of the astigmatism method for an error-curve generation, have been reported. We believe that the proposed method can find applications in microscopy, micro-machining, laser writing, optical lithography and highly sensitive alignment systems.

References

[1] Dmytro Kolenov, Peiwen Meng, and Silvania F Pereira. Highly-sensitive laser focus positioning method with sub-micrometre accuracy using coherent fourier scatterometry. *Measurement Science and Technology*, 2020.

[2] Xinbing Liu. Method and apparatus for determining focus position of a laser, 10 2001. https://patents.google.com/patent/US6303903B1/en.

[3] Derin Sevenler, Oğuzhan Avci, and M. Selim Ünlü. Quantitative interferometric reflectance imaging for the detection and measurement of biological nanoparticles. *Biomed. Opt. Express*, 8(6):2976–2989, 6 2017.

[4] Chris A. Mack. Understanding focus effects in submicrometer optical lithography: a review. *Optical Engineering*, 32:2350–2362, 1993.

[5] Ilya Alexeev, Ji Wu, Michael Karg, Zeev Zalevsky, and Michael Schmidt. Determination of laser beam focus position based on secondary speckles pattern analysis. *Appl. Opt.*, 56(26):7413–7418, 9 2017.

[6] Binh Cao, Munju Bae, Hyonkee Sohn, Jiyeon Choi, Youngduk Kim, Jeng-o Kim, and Jiwhan Noh. Design and performance of a focus-detection system for use in laser micromachining. *Micromachines*, 7:2, 01 2016.

[7] G Bouwhuis, Joseph Braat, A Huijser, J Pasman, G van Rosmalen, and Kees Schouhamer Immink. *Principles of Optical Disc Systems*. 01 1987.

[8] K. J Stout and W. P Dong. *Three dimensional surface topography : measurement, interpretation, and applications : a survey and bibliography*. London ; Bristol, Pa. : Penton Press, 1994.

[9] Ingrid Spadinger, Steven S. S. Poon, and Branko Palcic. Effect of focus on cell detection and recognition by the cell analyzer. *Cytometry*, 11(4):460–467, 1990.

[10] Alex D. Herbert, Antony M. Carr, and Eva Hoffmann. Findfoci: A focus detection algorithm with automated parameter training that closely matches human assignments, reduces human inconsistencies and increases speed of analysis. *PLOS ONE*, 9(12):1–33, 12 2014.

[11] Robert H Webb. Confocal optical microscopy. *Rep. Prog. Phys.*, 59:427–471, 1996.

[12] E. Mainsah, J.A. Greenwood, and D.G. Chetwynd. *Metrology and Properties of Engineering Surfaces*. Optoelectronics, Imaging and Sensing Series. Springer US, 2001.

[13] A. S. van de Nes, L. Billy, S. F. Pereira, and J. J. M. Braat. Calculation of the vectorial field distribution in a stratified focal region of a high numerical aperture imaging system. *Opt. Express*, 12(7):1281–1293, 4 2004.

[14] Joint Committee for Guides in Metrology. Jcgm 100: Evaluation of measurement data - guide to the expression of uncertainty in measurement. Technical report, JCGM, 2008.

[15] J. H. Friedman. A variable span smoother. Technical report, Laboratory for Computational Statistics, Department of Statistics, Stanford University, 1984.

[16] Daniel et al. Claus. Large-field-of-view optical elastography using digital image correlation for biological soft tissue investigation (erratum). *Journal of Medical Imaging*, 4:029801, 06 2017.

[17] H. A. Bruck, S. R. McNeill, M. A. Sutton, and W. H. Peters. Digital image correlation using newton-raphson method of partial differential correction. *Experimental Mechanics*, 29(3):261–267, 9 1989.

[18] Ian Farrance and Robert Frenkel. Uncertainty of measurement: A review of the rules for calculating uncertainty components through functional relationships. *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, 33:49–75, 05 2012.

[19] Ilya Alexeev, Johannes Strauss, Andreas Gröschl, Kristian Cvecek, and Michael Schmidt. Laser focus positioning method with submicrometer accuracy. *Appl. Opt.*, 52(3):415–421, 1 2013.

[20] J. Braat and P. Török. *Imaging Optics*. Cambridge University Press, Cambridge, U.K, 2019.

[21] P.A. Bobbert and J. Vlieger. Light scattering by a sphere on a substrate. *Physica A: Statistical Mechanics and its Applications*, 137(1):209–242, 1986.

[22] En-Te Hwu, Hsien-Shun Liao, Filippo Bosco, Ching-Hsiu Chen, Stephan Sylvest Keller, Anja Boisen, and Kuang-Yuh Huang. An astigmatic detection system for polymeric cantilever-based sensors. *Journal of Sensors*, page 580939, 2012.

[23] Y. Chiu, H. Shih, J. Chiou, S. Cheng, K. Hung, F. Tseng, and W. Fang. Design and fabrication of a small-form-factor optical pickup head. *IEEE Transactions on Magnetics*, 45(5):2194–2197, 5 2009.

[24] Yuan-Chin Lee, Shiuh Chao, Chun-Chieh Huang, and Kuen-Chiuan Cheng. A compact optical pickup head in blue wavelength with high horizontal stability for laser thermal lithography. *Opt. Express*, 21(20):23556–23567, 10 2013.

[25] Yoshinori SatoKoji Noriyuki Kawano, Tetsuya Nishiyama. Optical head device and optical recording and reading system. 10 2008. https://tinyurl.com/y6oaldrw.

[26] Contamination Blog. Classification & analysis on metallic contamination in semi-conductor manufacturing industry. https://goo.gl/MkduXD.

[27] Jeong-Eun Park, Keunsuk Kim, Yoonjae Jung, Jae-Ho Kim, and Jwa-Min Nam. Metal nanoparticles for virus detection. *ChemNanoMat*, 2(10):927–936, 2016.

[28] L.A. Cacace. *An optical distance sensor : tilt robust differential confocal measurement with mm range and nm uncertainty*. PhD thesis, Department of Mechanical Engineering, 2009.

Intermission

So far, we have shown that the improvements in the detection system using heterodyne techniques, optimization of the input light polarization and analysis of the defocusing effect can boost the performance of CFS for sensitive nanoparticle detection. The reason for the improved performance is the adoption and development of hardware solutions and measurement approaches that rely on the numerical simulations and analytical studies.

After mitigating the limitations emerging from experimental noise, non-optimal illumination and defocus, in the second part of the thesis, we start to investigate the constraints of the algorithms and software that extracts the relevant information from the scatterometry data. We concentrate on developing a robust framework for semi-automatic processing of the surface scan maps that contain the particle signals in the presence of background signals and artefacts in the data. We applied novel techniques known from computer vision to discriminate particle size-types and particle counting. The concepts and validation experiments are the themes of the following chapters of this thesis.

6

Efficient signal processing with the aid of Machine Learning.

Detection, classification, localization, and counting particles based on the scattered maps of CFS are complicated processes because of the problem of particle clustering, the variety of particle sizes and densities. In this chapter, we present the pre-processing of the data and a supervised approach for automatic detection of the differential voltage particle signal in the 2D raster scan dataset.

The chapter is organized as follows. In Sections 6.1-6.3, we explain the potential causes of data distortion in off-line and in-line measurements obtained with the particle scanner, the creation of a CFS dataset, and the required stages for the data processing workflow. Then, in Section 6.4, we propose digital filters that minimize the measurement-induced artifacts in the scatterometry results, develop a signal-search algorithm that targets datasets with a large density of particles, and finally modify the clustering algorithms and analyze their precision, computational complexity and robustness for datasets with a different size of the particles and distorted signals. The experimental findings and conclusions are accordingly summarized in Sections 6.5 and 6.6.

Parts of this chapter have been published in Optics Express **28**, 19163-19186 (2020) [1]

6.1. Introduction

For the detection of very small particles using CFS, it is crucial to optimize the entire system. In this regard, we mean both improvements directed to the system itself as well as advancement to the data processing framework.

As we have shown in earlier chapters, in order to validate the results of the particle detection, reliable numerical tools need to accompany the measurements process. An example here can be rigorous simulation using full electromagnetic solvers or tools that process the experimental data. However, scatterometry data become less powerful if the algorithm that treats the data cannot effectively discriminate between different sizes of the particles present on a particular surface. One of the complicating factors is that, besides the inherent noise related to the detection of light, in a production environment, data can be corrupted with several other sources of noise and artifacts. The presence of extensive size-range contamination severely complicates the analysis of individual particles. In the worst-case scenario, if the measured data are examined in the wrong way, this can lead, for example, in the case of lithography, to a drop in system productivity. Finally, taking into account the growing amount of data, the techniques such as CFS lack the tools of being able to process raw datasets semi-automatically and effectively. Recently, to overcome the challenges of detection and classification of smaller particles, machine learning methods, including the regularized matrix-based imaging framework [2], principal component analysis [3], and convolutional neural networks [4] were applied to image-based defect detection.

The objective of this chapter is to develop a full framework for particle size classification in scatterometry data that includes the pipeline of pre-processing, signal search and histogram formation with an algorithm that can be directly targeted at data that are corrupted with noise and drift, as well as including mixed-size particles per sample. For this framework, we relied on the established noise-removal and unsupervised clustering techniques and adapted them to detect the nanoparticles. We developed a parameterized search by thresholding that picks the differential signal shape (raw data from the scatterometer) and relates it to the detection information (size distribution and location of the particles). By using these techniques, we show that nanoparticles could be accurately quantified, even in the case of high densities. With sufficient resolution, an experimental sample containing a mixture of nanoparticles with $60, 80$ and 100 nm was distinguished in conditions where the dataset had a lot of noise and drift due to the scanning of the surface. The framework enables the demanded automatic analysis of the scatterometry data and facilitates the validation of the detection results.

The summary of the functions used in this chapter is given in Table 6.1.

Table 6.1: Glossary of the main functions used in this chapter.

name	explanation	mathematical description
detrend	Trend can be modeled and removed from the time series	minimize $J = \sum_x [y_x - (ax + b)]^2$ where measured data values y_x in time x and a, b are chosen to minimize J
abs	Modulus of the real number	$\lvert y \rvert = \begin{cases} y, & \text{if } y \geq 0 \\ -y, & \text{if } y < 0 \end{cases}$
ceil	Round up or round towards plus infinity	$y = ceil(x) = \lceil x \rceil = -\lfloor -x \rfloor$
dist	The Euclidian distance between a point x_n and μ	$\text{dist}(x_n, \mu) = \sqrt{(x_n - \mu)^2}$
ind1 *ind2*	Sampling points of the signal at which the amplitude is maximized/minimized	$\text{argmax}_x\, y(x)$ $\text{argmin}_x\, y(x)$
size	Size or cardinality of a set is a measure for the number of elements of the set n	$card(y_x) = n$

6.2. Methods

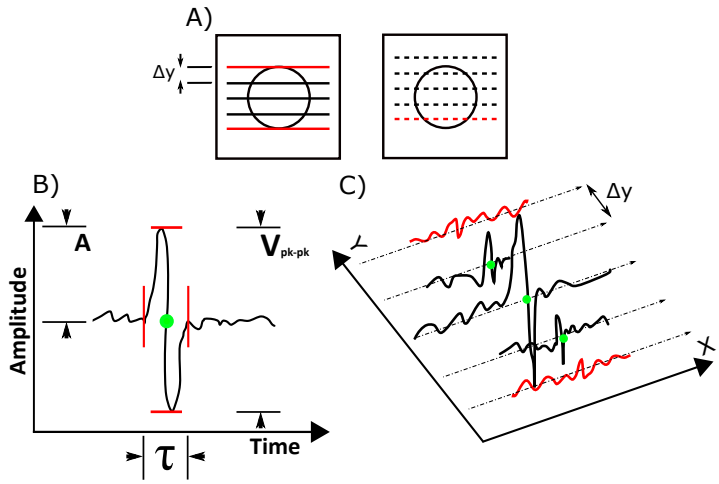


Figure 6.1: A) 2D raster scanning showing scan lines in the x-dir. separated by $y = 10$ nm in y-dir. The geometrical size of particle is $d = 50$ nm. On the left - the first scanning line coincides with the edge of the particle, and consequently the differential signal will appear in 5 consecutive lines, with the red lines providing small amplitude of the signal. On the right - if there is an offset between the first scanning line and the edge of the particle, the signal due to this particle will be spread in fewer lines. B) The recorded diff. signal as one line containing the particle is scanned - first maximum and then minimum when the positive lobe is equal to the negative lobe (balanced signal). Red lines constrain the features of width τ and amplitude V_{pk-pk} of the diff. signal. C) A particle response as a collection of similar patterns is called a scattered map.

In CFS the recorded signals from the photodetector are the basis for the signal maps (one voltage value (photocurrent to voltage) for every scanning position reshaped to 2D xy distribution). When the probe is focused on the interface, scanning of the spherical nanoparticle on the surface will render the so-called balanced pulse (positive lobe is equal to the negative lobe of the signal). The zero-crossing of this pulse refers to the perfect alignment between the center of the nanoparticle and the focused spot (green point in Figure 6.1 B)). The effect of the defocus produces unbalance of the signal as well as a drop in the SNR, as demonstrated in Chapter 5. When analyzing the surface in a raster scanning fashion, one needs to choose the proper Δy displacement step between the parallel lines of scanning. The bigger the step, the lesser the time it takes to cover the complete area, but the downside is that particles can be missed. The simplified picture showing the relationship between the scanning step Δy and the particle size is shown in Figure 6.1 A). Experimentally, if one wants to detect subwavelength contamination of e.g. 50 nm, the step between lines of the scanning should be set lower than the particle diameter, for instance $\Delta y = 10$ nm. For relatively big particles, one can expect that every time the probe interacts with the particle, a high-enough scattering will be produced and thus the estimate for the amount of lines where the particle is visible equals to $n = d/\Delta y$, with d being the diameter of the particle. N.B. the SNR is the defining factor for the effective amount of lines and the influence area of the particle is bigger than the physical size as according to its scattering cross-section. Yet, the outlined picture highlights the idea that there is a certain minimum and maximum expected number of beneficial signals that would emerge from using the different settings of Δy. For the tiny particles < 80 nm with $\lambda = 405$ nm, one can expect that the scanning lines that would go through the edges of the particle have much less of an SNR (see Figure 6.1 A) and C)), because the amplitude of the scattering becomes small as the probe position goes away from the center of the particle. Furthermore, mismatches between the position of the particle and the scanning step might occur (dashed lines Figure 6.1 A)) and in this case, even fewer lines containing signals due to the particle are made. The rule of thumb is to have at least two signals that come from an isolated particle that is distinguished from the background. Finally, the nanoparticles are generally classified based on their dimensionality, where the size of the calibrated sphere is associated with the features of V_{pk-pk} amplitude or time-width τ of the measured differential signal (Figure 6.1 B)).

6.3. Sub-problems

The task of detection and classifying the particles using scatterometry data can be split into sub-problems. In this section, we discuss these sub-problems: pre-processing of data, finding the particle-like signals, estimation of the width and cluster assessment.

6.3.1. Pre-processing

The goal of the pre-processing task is to prepare the raw sampled data for further steps. Commonly, a DC bias and sometimes baseline fluctuations in the signal at the detector can occur due to vibrations and other experimental factors. For the removal of the various electronic noise, low-pass (LP), notch filtering and wavelet-based subtraction were applied.

6.3.2. Selection of suitable Amplitude and Width

We use two parameters for the object detection: the A amplitude ($V_{pk-pk}/2$) and the τ width of the complete differential signal due to a single particle in the time-domain (see Figure 6.1 B)). We look for an algorithm that is robust to non-particle signals that can be present in the data. Examples of such signals include environmental vibrations or large defects present on the surface of the material, which we can consider as false-detections.

In the scan direction x, multiple particles may be present on one scan line since the density of particles can be high in some areas of the surface. Multiple maxima and minima needs to be determined on a scan line, sorted and the relative distance between pairs needs to be determined. Each "transition" between maxima and minima is associated with the corresponding zero-crossing position at the middle of the signal. Finally, fine adjustment is needed to define the width of the particle-like signal accurately; this is done by parameterizing it such that it can be distinguished from noise or another signal in the dataset. Next, one needs to take care of the particle signal appearing at the border of the scan line. In this case, if the signal was sampled for one of the two lobes (positive or negative), the algorithm should estimate the complete width of the pulse.

6.3.3. Multiple line particle detection identification

A particle-like signal is distinguished from a false detection if the centroid of the signals (position of the zero between maxima and minima) has the same X position over multiple lines (see Figure 6.1 C)). A false detection is identified when the particle-like signal is observed in only one scan line, and is further removed from the data. Finally, per cluster, the most clear particle-like signal and its features (see Figure 6.1 C)) are stored for the histogram. The pulse with the biggest V_{pk-pk} is a good representative because it corresponds to the center of the particle in the x and y directions.

There are many well-known algorithms for cluster determination, such as hierarchical clustering, K-means, DBSCAN [5–7]. Almost every clustering algorithm can be tuned to penalize one error more than the other according to the requirements. For instance, we can use the predefined vertical step of Δy and set the expected number of zero-crossings associated with a single particle. Additionally, the clusters of zero-crossings (pair of x, y coordinates) can have a characteristic spread of σ.

6.4. Algorithm

In this section, we show the specific tools and algorithms we have used to solve the sub-problems shown above. We also mention the computational efficiency and some other aspects of the algorithms.

6.4.1. Pre-processing

Among various noise sources that might be present in our experiment [8], the power line interference and the baseline wandering can strongly affect the further detection and classification of particle signals. The 50 Hz local power-line frequency (bandwidth of < 1 Hz) can be removed digitally using the notch filter. However, the baseline wandering is not easy to be suppressed by analogue circuits. Hence we take the

notch-filtered waveform and subtract the wavelet decomposed version of the same signal to recover the clean particle signal. This step effectively introduces the point by point correction to the wandering profile. Finally, an average filter (LP) is applied to remove glitches. This approach can be considered as more rigorous because it relies on the sampling frequency used in the experiment. Therefore, the input-output description of the digital filter operation on an input signal vector $x(n)$, where n is the number of samples, can be expressed in the form of the difference equation:

$$a(1)y(n) = b(1)x(n) + b(2)x(n-1) + \ldots + b(n_b+1)x(n-n_b) - a(2)y(n-1) - \ldots - a(n_a+1)y(n-n_a), \tag{6.1}$$

where n_a is the feedback filter order, and n_b is the feed-forward filter order. We design a second-order notch digital filter, thus $n_a = n_b = 2$ and Eq. 6.1 becomes:

$$a(1)y(n) = b(1)x(n)+b(2)x(n-1)+b(3)x(n-2)-a(2)y(n-1)-a(3)y(n-2), \tag{6.2}$$

with the notch at frequency 50 Hz and a bandwidth at the -3 dB level (q-factor of 35), we have notch frequency $W = 50/(f_s/2)$ (specified as a positive scalar in the range $(0.0, 1.0)$, where 1.0 corresponds to π radians per sample in the frequency range) and bandwidth $BW = W/35$. Thus, e.g. for sampling frequency $f_s = 3$ kHz, the coefficients are $\mathbf{b} = [0.998, -1.986, 0.998]$ and $\mathbf{a} = [1, -1.986, 0.997]$. The local power line frequency is removed from the data set and $y_{notch} = a(1)y(n)$. Further, we subtract the wavelet decomposed version of the signal y_{wd} from the filtered waveform y_{notch} effectively introducing the point by point correction to the profile. The discrete wavelet transform (DWT) of signal $y_{wd}(n)$ is defined as a combination of a set of basis functions:

$$y_{wd}(n) = \sum_{k=-\infty}^{\infty} c_j(k)\phi_{j,k}(n) + \sum_{j=1}^{J}\sum_{k=-\infty}^{\infty} d_j(k)\psi_{j,k}(n), \tag{6.3}$$

Where

$$\begin{aligned} \phi_{j,k}(n) &= 2^{j/2}\phi(2^j n - k), \\ \psi_{j,k}(n) &= 2^{j/2}\psi(2^j n - k). \end{aligned} \tag{6.4}$$

In Eq. 6.3, $\phi_{j,k}(n)$ is the scaling function, $\psi_{j,k}(n)$ is the wavelet function, $c_j(k)$ are the scaling and $d_j(k)$ are detailed coefficients. In this chapter, the Daubechies 6 scaling and wavelet functions were used because it has been proved to be excellent in analysis of signals that contain baseline wandering [9, 10]. For computing the $c_j(k)$ and $d_j(k)$ coefficients, the low-pass (LP) and high-pass (HP) filters are being recursively applied to a signal. When the signal is processed for the first time, the HP filtered data gives the details and LP filtered data gives the scaling coefficients at level 1. The more times the filters are applied, the more detailed levels of the signal representation can be achieved. In this chapter, we have used the decomposition level of $j = 10$ and have applied the translation factor of $k = 8$ for the scaling and wavelet function. The baseline wandering is removed and $y_{bcor} = y_{notch} - y_{wd}$.

Finally a simple moving average filter is applied according to Eq. 6.5.

$$y_i' = \frac{1}{M}\sum_{j=0}^{M-1} y_{bcor}[i+j]. \tag{6.5}$$

The output signal $y_i^{'}$ is a result of averaging the points in the input signal y_{bcor}, and $M = 5$ is the number of points used in the moving average. A less accurate way of dealing with the offset in the data can be MatLab's *detrend* function that removes the best straight-fit line from the data in a vector of the sampled points.

6.4.2. Selection of suitable Amplitude and Width

Hyperparameters:

$$A, \tau, NullingR, Window_y$$

These are user-defined parameters of expected threshold amplitude A and width τ. Since the multiple expected amplitudes and widths are passed iteratively, the results from the previous search should not translate to the consecutive one. Let's consider the 2D measured data $\boldsymbol{I}_{ij}$ with each row representing a single scan line (y and column representing the sampling point over the width (x of the Figure 6.1 C). The differential signal at the detector for the $i = 4$ scanning lines and with $j = 4$ samples in horizontal direction of scan is given in Eq. 6.6a.

$$\boldsymbol{I}_{4x4} = \begin{bmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \end{bmatrix}, \tag{6.6a}$$

$$\boldsymbol{I}'_{4x4} = \begin{bmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & 0 & 0 & 0 \\ I_{31} & 0 & 0 & 0 \\ I_{41} & 0 & 0 & 0 \end{bmatrix}. \tag{6.6b}$$

The parameters of $NullingR$, $Window_y$ represent the half-width and length of the region to be zeroed w.r.t reference sampling point. Hence, for measured data, if I_{33} is the reference position (center of the particle), the $NullingR = 1$ and $Window_y = 3$ dataset becomes Eq. 6.6b

Thus by, $NullingR$ and $Window_y$, the user can zero the lines that are close to the reference detected particle. Per line, the algorithm looks for the multiple peaks and minima and checks whether their absolute values fall under the amplitude A. The retrieval of secondary peaks and minima allows increasing the overall accuracy of the algorithm. By default, we assume every particle-like looking signal to be a forward signal (Figure 6.1 B)). The reverse signal can be stored separately or included in the estimation process. Some key reasoning are highlighted in the following bullet-points and also shown in Figure 6.2.

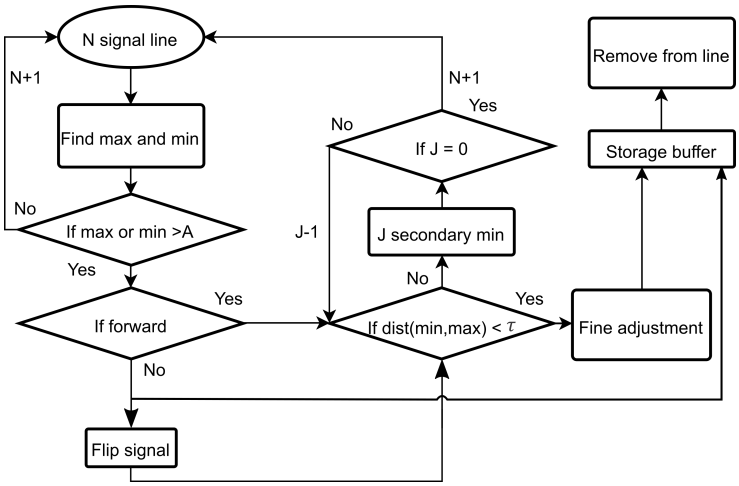


Figure 6.2: Block diagram of the signal search algorithm that starts with the N signal line.

- Find and store the values and indices of global line maxima $ind1$ and global line minima $ind2$. Next, check the amplitude condition $abs(max1) > A\ OR\ abs(min1) > A$. Store True or False for the first condition.

- Define whether the signal is forward (maxima appears before minima), choose between ignoring or flipping the reverse pulses. Check whether the distance between 2 indices $abs(ind2 - ind1) < \tau$ fits the condition of the time-width. Store True of False for the second condition.

- When both conditions are true, a particle is roughly detected. We calculate the position of the particle's signal zero-crossing by taking the average between maxima and minima position $middle = ceil(abs(ind1 + ind2)/2)$ (ceil function rounds towards plus infinity). Perform the fine adjustment (next section) and remove the signal from the dataset. The $NullingR$ is global parameter that represents half-distance in indices to replace with zeroes about the $middle$ of the signal. The rule of thumb, in this case, is that region to be zeroed should not exceed the time width of the particles you are looking for.

- If only the amplitude condition is satisfied, the indices of multiple minima (above threshold) that belong to the current line are checked to fall closer to the $ind1$ than $ind2$. If other minimum falls closer, reassign the $ind2$ and repeat the width check. If both conditions are satisfied, remove the signal from the dataset and apply the $NullingR$

- The multiple particle search routine is to find numerous maxima (above threshold), sort them in descending order, and, maximum by maximum, follow the steps outlined previously. If there are multiple particles on a single line, the algorithm returns x's corresponding to the particles middles.

Throughout this chapter, we will use the terms "zero-crossing" and "middle" interchangeably, following the variable name of $middles$ as defined in the MatLab software.

6.4.3. Fine adjustment for the boundaries of particle's signal

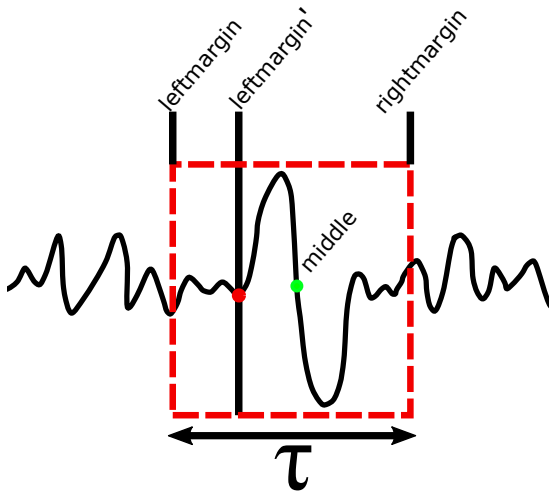


Figure 6.3: The sketch demonstrating the margins of the signal separating it from the background. The fine adjustment algorithm is to go from $leftmargin$ to $leftmargin'$.

Fine adjustment is the part of the search process right after the width, and amplitude conditions are satisfied. We assume that from the $middle$ position, the particles' signal occupies the same amount of samples on both sides of the signal (spherical object). The initial guess for the left margin for the signal is $leftmargin = middles_{ind} - \tau/2$. To make sure that signal does not go outside the indexing in Matlab, i.e., $middles_{ind} - \tau/2 <= 0$, we rewrite the left margin as index 1. In this case, in the procedure that follows, we should rely on the $rightmargin$ to be defined accurately and then $leftmargin$ is recomputed based on it. Analogously the right margin is calculated as $rightmargin = middles_{ind} + \tau/2$. If a signal appears close to the right border, we rewrite right margin as the last index of the sampled voltage vector. The crucial part of the fine-adjustment step is to cut-out the region of signal for zoom-in study, i.e. from $leftmargin$ to $rightmargin$. The secondary minima of the cut-out region are checked to fall closer to the middle position, compared to the initial $leftmargin$. If there is a closer point, it is redefined as $leftmargin'$. The reason for this is an observation that typically there is a small dip in the signal preceding the quick rise in amplitude of the particle pulse. Further, we reassign the $rightmargin$ of signal to be the same separation as to the left $rightmargin = middles_{ind} + abs(middles_{ind} - leftmargin')$. Finally, the $signalSize = abs(rightmargin - leftmargin')$, and one can notice that in our procedure the estimator can generalize outside the original size of the sampled vector.

6.4.4. Clustering of data from one single particle

The steps of the algorithm presented previously result in an array of coordinate pairs for $(\boldsymbol{X},\boldsymbol{Y})$, that correspond to the position of the $middles$ (zero-crossings) of each signal and has dimensionality $2 \times N$, where N is the number of $middles$. Since one particle results in a few signals at consecutive lines of the scan, one should recognize a group of the particle-looking signals as a centroid that represents this particle. In this way, the particle-size distribution histogram will identify one particle on the sample corresponding to one cluster of signals. The centroid of the cluster will correspond to the line with the highest V_{pk-pk} of the cluster, and consequently the particle's center.

We modify a well-known machine learning algorithms of K-means and DBSCAN to recognize the clusters of the particle-looking signals and use prior information that can help to spot the isolated particles. One complicating factor that can be present in the data is the random drift between the lines when sampling. The drift manifests itself in the shift of the signal zero-crossing position in the x-direction between consecutive lines (see Fig. 6.1 C). Now, we will define several algorithms that can account for the drift in the data. We will compare the computational complexity of the modified K-means to the algorithm of DBSCAN. Besides, we will highlight the sensitivity of algorithms to initialization parameters.

Modified K-means, DBSCAN and comparison.

Given, for instance, the $middles$ coordinates $(\boldsymbol{X},\boldsymbol{Y})$, K-means clustering, can converge to a K amount of clusters, among which per cluster we know the distance between each point and the position of the cluster centroid mean μ. The first two algorithms are used to treat the outliers in the clusters K, where n being the number of points in a cluster with indices iii:

Algorithm 1 Sort in descending order

procedure sort(X_n) $\quad Where\ x_1, x_2, ..., x_{iii} \in \mathbb{R}$
  $inversions = 0$
  for $iii\ in\ 1 : n$ **do**
    for $jjj\ in\ 1 : iii$ **do**
      if $X(iii) > X(jjj)$ **then**
        $inversions = X(iii)$
        $X(iii) = X(jjj)$
        $X(jjj) = inversions$
      end if
    end for
  end for
return X
end procedure

Algorithm 2 Remove if σ improves

Input: $(X,Y)_{iii} \in K$ In cluster each *middle* (X,Y) is associated with *dist* to μ
Output: Nremove
$set\ Nremove\ to\ zero\ (return\ if\ iii < 2)$
$Sort\ the\ points\ by\ dist$ Descending order Alg. 1
for $b \leftarrow 2$ to iii_{dist} **do**
  $m \leftarrow Eq.(6.7)$
  if $m < 0.1$ **then**
    $Nremove \leftarrow b - 1$
    return $Nremove$
  end if
end for

Where the metric m, standard deviation σ, and average μ are computed according to Eqs. (6.7 - 6.9). For a random variable vector M made up of S_o scalar observations

$$m = \frac{\sigma(X,Y)_{iii}}{\sigma(X,Y)_{iii-1}} - 1, \quad (6.7) \qquad \sigma = \sqrt{\frac{1}{S_o - 1}\sum_{i=1}^{S_o} |M_i - \mu|^2}, \quad (6.8) \qquad \mu = \frac{1}{S_o}\sum_{i=1}^{S_o} M_i. \quad (6.9)$$

Example of such an algorithm (Alg. 2) applied to an arbitrary cluster is shown in Figure 6.4 A). The idea is to remove the points that are too far from the mean, and we use the constant of 10% decrease in standard deviation (STD) to reject the outliers. The initial 8 points in cluster K are sorted in descending order by the distance from the mean μ. When removing the first two points, the metric $m > 0.1$, but not when we remove the third, $m < 0.1$ thus cluster will be reduced to the most packed 6 points.

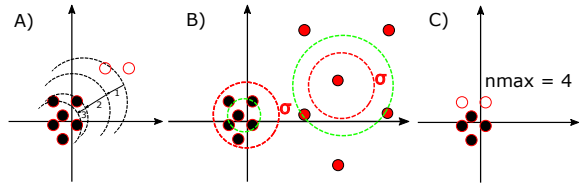


Figure 6.4: A) Per cluster the outliers are removed according to predefined 10% decrease in STD. Only the first two points will be moved to the outliers because the 3rd point is close together with the other points; B) Global parameter of σ_{thresh} in red and, per cluster, the estimated σ in green, is either below or outside the expected range; C) Maximum number of points per cluster n_{max}, and $n_{pk} < n_{max}$ is the number of points in cluster

When the outliers are removed we want to reject the clusters that are overly spread for instance due to the drift. In our approach, the spread of particular cluster has to be below $\sigma_K < \sigma_{thresh}$ and it is computed according to Algorithm 3.

Algorithm 3 Compute spread

1: $\mathbf{X} \leftarrow \mathbf{X} - \mu(\mathbf{X})$
2: $\mathbf{Y} \leftarrow \mathbf{Y} - \mu(\mathbf{Y})$ $cluster \leftarrow centertoZero(cluster)$
3: $\sigma_K = \sigma(K(:))$ $Eq.\ 6.8\ operating\ on\ clusterK$

The idea here is that the user selects a single cluster that with a big confidence corresponds to an isolated particle and passes the corresponding recommended limit of σ_{thresh}. The example of the thresholding by spread 6.4 B) shows that such a limit will be met by the set of black points but not by the red set. Finally, based on the geometric considerations outlined in Section 6.2 of this Chapter, we add a prior on the resultant amount of points that contribute to a single cluster. For a target particle diameter of the d, the amount of the zero-crossing points $(X,Y)_n \leq \frac{d}{\Delta y}$.

Algorithm 4 K-means with prior

procedure Kmeansp($passp, stdLimit, thresh$)
$passp : input\ points\ (X,Y), stdLimit : \sigma_{thresh}, thresh : n_{max}$
  while $size(passp) > 0$ **do**
    $K \leftarrow optimalK$ Defined by elbow method
    $kmeans(passp, K)$ Apply classical K-means
    for $1 : K$ **do** For each cluster
      if $Nremove \neq 0$ **then** Remove outliers in cluster, Algorithm 2
        $set\ sp = sortedPoints(1 : Nremove)\ false$ Defining outliers
        $NoOutliers \leftarrow reverseSort(sp)$
      end if
      if $spread(cluster) > stdLimit$ **then** Find sparse clusters
        $highSpread \leftarrow (spread(NoOutliers) > stdlimit)$
      end if
      if $size(cluster) > thresh$ **then** Find dense clusters
        $ManyPoints \leftarrow (size(NoOutliers) > thresh)$
      end if
      $goodPoints \leftarrow\sim NoOutliers(highSpread$ AND $ManyPoints)$
      $badPoints \leftarrow NoOutliers(highSpread$ AND $ManyPoints)$
      if $size(goodPoints \leq 1)$ **then** Remove single/zero point clusters
        $Remove\ isolated$
      end if
      $passp \leftarrow badPoints$ Send not suitable points to new iteration
    end for
  end while
end procedure

Hyperparameters: n_{max}, σ_{thresh}
We modify the K-means [6, 11] such that it can accurately establish the isolated particles. For validation purposes, the density of spheres is crucial, thus the algorithm needs to overcome its inherent tendency to overestimate clusters. One difference to

the original K-means is that we introduce the outliers. The outliers include: A) clusters with single particle; B) empty clusters; C) distant points previously included in a cluster. The option C) is treated by Algorithm 2. The second difference is conditioning of the assigned clusters. Clusters are considered to be valid if $n_{pk} < n_{max}$ (Figure 6.4 C)) and $\sigma'([X,Y]) < \sigma_{thresh}$, where $\sigma'([X,Y])$ is $spread$ computed for the set of points, with a mean moved to the zero and the distance normalized to unity (Algorithm 3). After K-means convergence (one epoch), if there are points that fail on both conditions n_{max} and $\sigma'([X,Y])$, they are passed through the K-means again. The algorithm stops if all points are assigned to either cluster or an outlier. In every epoch of the K-means, the optimal amount of clusters is defined from the elbow method based on the average of 3 random initializations. Hence, in our implementation, the K-means is described via Algorithm 4.

DBSCAN

Hyperparameters: n_{min}, ϵ

The density-based clustering algorithm (DBSCAN) [7, 12] has a straightforward advantage in taking care of the obscure points, such that all points that are not reachable from any other point are outliers or noise points. The two hyperparameters are inclusion radius ϵ and minimum number of points in the cluster n_{min}. The input for the n_{min} is straightforward, such that it can be any number between $1 < n_{min} < n_{max}$. For the ϵ recommendation, we use the following routine:

- Normalize the complete data set of $middles$ to the unity, such that $norm\mathbf{X} = \frac{\mathbf{X}}{max(\mathbf{X})}$ and $norm\mathbf{Y} = \frac{\mathbf{Y}}{max(\mathbf{Y})}$.
- Select the set of points that with a high confidence forms a cluster, via visual inspection, $confX = \{norm\mathbf{X}\}$ and $confY = \{norm\mathbf{Y}\}$. Center this cluster to the zero $centertoZero(confX, confY)$ (first two lines of Algorithm 3)
- Determine the average distance between points. Includes computation of Euclidian distance between each pair of observations in separately **X** and **Y** and taking average of each vector.

As a result of K-means or DBSCAN, one can pick up the converged clusters and either: A) Pull the features of $signalSize$ specifically for the highest V_{pk-pk} from corresponding pulses in a cluster; B) Average of the corresponding time-spans $\bar{\tau}$ of points in clusters (Eq. 6.10); C) The full amplitude of a signal itself V_{pk-pk}.

$$\bar{\tau} = \frac{\sum_{i=1}^{n_c} \tau_i}{n_c}. \tag{6.10}$$

Where n_c is number of points assigned to the cluster. Correspondingly, the centroids of clusters are stored for the mapping of the particle positions.

On computational complexity of algorithms. Comparing two algorithms

The classical K-means algorithm has a complexity $O(TKn)$, where n is the number of input points, K is the desired number of clusters, and T is the number of iterations needed for convergence. It is also observed that approximately $T \propto n$ [13]. Hence, the effective time complexity becomes $O(n^2)$. The K-means is a greedy algorithm since it can produce both empty and over-populated clusters. Another drawback is the large

dependence on the initialization of cluster centers. As according to the quadratic time complexity, it should not be used in extremely large data applications [14]. Implementation of the K-means with prior in this chapter has $O(n^3)$ time complexity or worst time complexity of higher-level polynomial $O(n^c)$.
In the DBSCAN implementation, for each of the points of the input data, we have at most one region query. Thus, the average run time complexity of DBSCAN is query of log n times the amount of points n, $O(n \cdot logn)$ or worst time complexity of $O(n^2)$.

Sensitivity of K-means and DBSCAN algorithms

Sensitivity analysis was used to explore how the accuracy of algorithms would change with slight variations in the hyperparameters. The green point in every plot represents the most preferred initial value that yields the highest accuracy, while the offset from this point defines the sensitivity (see Fig. 6.5).

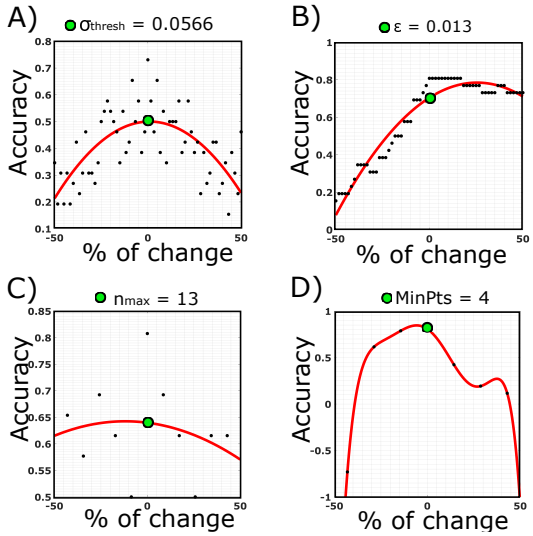


Figure 6.5: Sensitivity analysis of isolated changes of hyperparameters for K-means and DBSCAN algorithms. The accuracy changes as a function of A) σ_{thresh} and B) ϵ, with fixed $n_{max} = 13$ and $MinPts = 4$ and, as a function of C) n_{max} and D) $MinPts$ with $\sigma_{thresh} = 0.056$ and $\epsilon = 0.013$ in C) and D).

6.5. Results

Throughout this section we experimentally study three different samples of the PSL particles spin-coated on the silicon wafer. The first sample includes particles with diameters of 50 nm, the second 100 nm, and the third a mixture of 60, 80 and 100 nm. The details on sample preparation are outlined in the Appendix C.

6.5.1. Pre-processing and search.

In the IC manufacturing double side polished wafers are used. The block of pure crystalline Si is diced and polished right before the resist deposition. Due to the lack

of precision in the wafer holder, unstable rotation and heat deformation, the polishing can affect the flatness of the wafer. Additionally, the thickness of the wafer is not uniform across the sample [15]. This effect mostly occurs at the edges of the wafer. Nevertheless, the scanners need to provide information over the entire wafer under the study. For sensing or particle detection applications using CFS, the probing light should be focused on the interface between air and top surface.

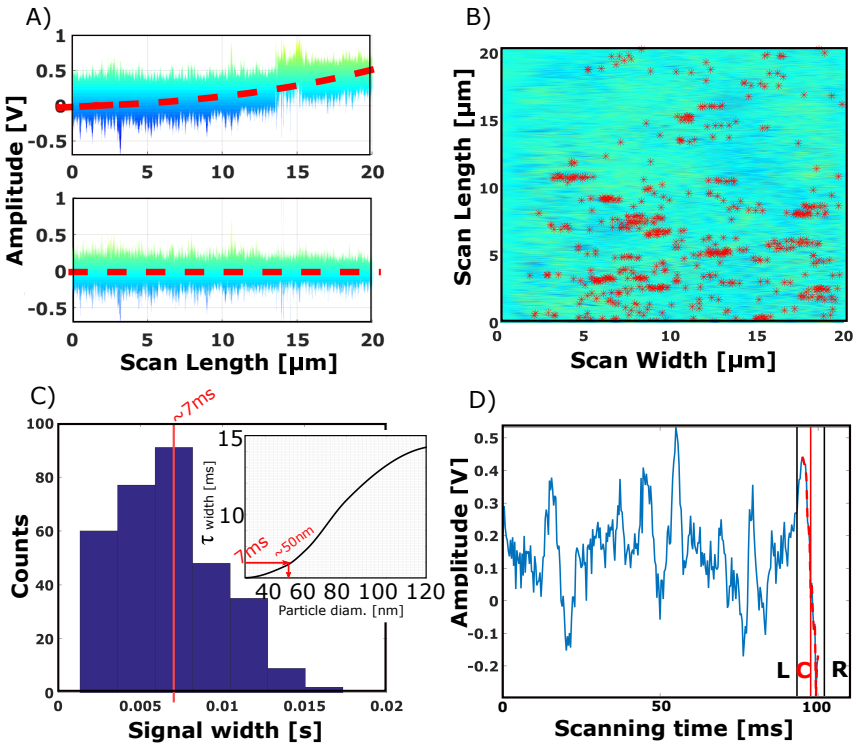


Figure 6.6: A) Top - the side view (along y of the raw sampled data wherein the baseline wandering is present. Bottom - the corresponding data after the baseline wandering is removed. B) Top view on the same scan with the red points representing the detected zero-crossings. C) Histogram representing the particle size distribution, based on time width τ from detected pulses. The inset shows the calibration of size of the particle as a function of the time width of the signal. D) Example of the line from the data set, the dashed line is an initial guess for the time-width and left - L and right - R boundary is returned from the fine-adjustment step. The scan width of 20 microns corresponds to 100 [ms].

Due to several experimental factors during the scanning, the baseline (differential signal when no particle is present) may fluctuate or drift from the expected zero value. Hence, occasionally, the data set might include DC offsets mixed with low frequency noise (baseline wandering) [16, 17]. This problem can be corrected as shown in the data presented in Figure 6.6 A) raw data (top), and with baseline correction (bottom).

Further, the scattered map from the bottom Figure 6.6 A) is analyzed with the search algorithm (Section 6.4.2) to produce the particle-mapping data that is seen on the 6.6 B). Here we analyzed a random area from the calibrated sample and the histogram nicely peaks at the position of the $\tau = 7.05$ [ms] that corresponds to PSL particle with 50 nm in diameter as according to the recipe of the first sample. For the area that contained only a few particles, one can notice a relatively high amount of counts, and this is because all the localized zero-crossings contribute to the output histogram. The SNR ratio for this dataset is low $SNR = 7.14$ [dB] while the algorithm can still localize the particle detections, including the one that resides at the border of the scan, thus generalizing beyond the input data. N.B. The particle classification in CFS is based on the width of a time-domain particle signal. The quantitative limit of the post-processing framework for discrimination between the different-size particles is defined by the accuracy of the fine-adjustment routine of Section 6.5.1. More specifically, in the ability to find the minima closest to the rising edge of the differential signal. If we assume the infinite sampling of the signal and low noise, there are virtually no limitations on how accurate the position of minima can be defined, aside from those emerging from the numerics or computational effort [18]. On the practical side, there is a limitation in the manufacturing of the monodisperse PSLs. The target size of the particle diameter has the uncertainty in the range of $1 - 2$ nm [19].

6.5.2. Comparing the accuracy of clustering routines on a dataset with drift.

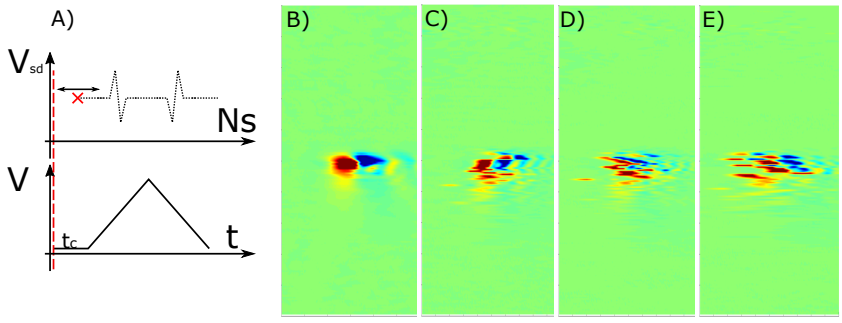


Figure 6.7: A) The start sampling point (red cross) has fluctuation in time - above (asynchronous). The primitive of the voltage waveform for moving the piezo along one axis forth and back - below. The time-constant tc tries to match the start of the piezo movement (uprising edge of the waveform) with the sampling start. Example of the isolated nanoparticle with diff. amount of drift present. The non-drift image B), an increasing amount of drift from 100, 90, 50 ms scanning time per line, C), D) and E) correspondingly.

The source of the drift originates from the sampling at the detector being an asynchronous process with respect to the piezo stage movement. When the piezo controller passes the initialization signal to the computer, the jitter and USB connection produce a random time delay before the sampling will actually start. One can mitigate the prob-

lem by introducing the constant wait time tc (empirical estimate) at the piezo before the voltage will be increased (Figure 6.7 A)). Yet, the random nature of the delay will not be equal to the introduced tc. When faster scanning is performed, the drift in the dataset gets worse. Figure 6.7 B) - E) shows the same isolated nanoparticle scanned at a different speeds: 100, 90 and 50 ms per line, demonstrating an increasing amount of the distortion in the dataset.

We take the data corrupted with the drift and compare the accuracy (Eq. 6.11) of two clustering algorithms as the average result of 100 random initializations (not to be confused with the classification accuracy of the next Chapter.)

$$Accuracy = 1 - \left| 1 - \frac{N_{det}}{N_{true}} \right|. \tag{6.11}$$

where N_{det} is a number of detected clusters, hence isolated particles, and N_{true} the actual amount of particles on the sample. This formula ignores the difference between the over- and underestimate in the N_{det}.

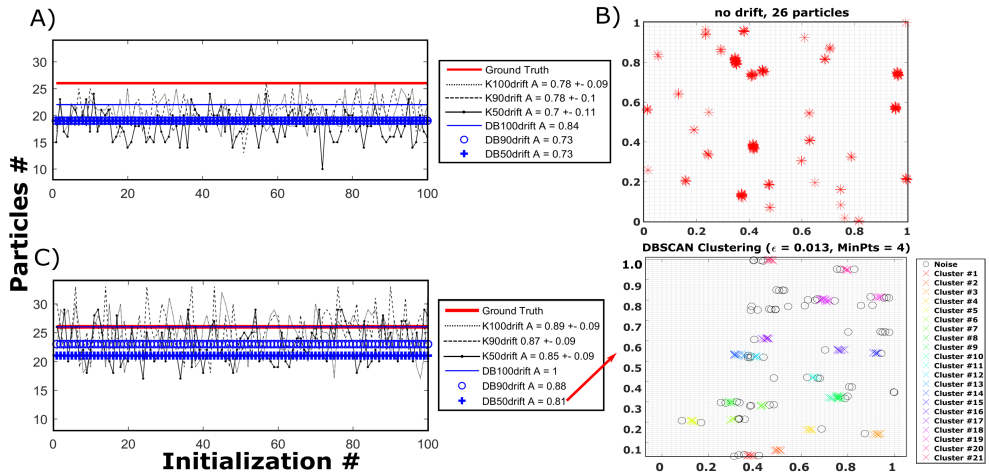


Figure 6.8: The true number of particles (ground truth) in red. Comparison of the DBSCAN (in blue) and modified K-means algorithm (in black) for the three levels of drift. 100 drift represents the least distorted dataset, 90 drift dataset with average distortion, 50 drift is the dataset with severe distortion. A) Recommended and C) tuned hyperparameters. B) Result of 81% accurate convergence by the DBSCAN for the case of 50 drift present.

We will use the non-drift corrected "image" as a ground truth for this comparison providing us the number for N_{true}. The non-drift "image" is achieved by establishing a new synchronization approach with the trigger pulse generated at the piezo controller through analogue output upon each beginning and end of the scanning line.

The first test is to use some of the global parameters such as $n_{min}, \epsilon, n_{max}, \sigma_{thresh}$ according to the reasoning outlined in the Section 6.4.4. Recommended hyperpa-

rameters come from showing the program once what the "good cluster" looks like. A number of 100 random initializations were needed to get an idea of how the K-means algorithm will suffer from random initialization, specifically the starting number of clusters K and their positions are randomly initialized. On the contrary, the DBSCAN, regardless of initialization, always converges to the same result (see Figure 6.8). This test reveals that both algorithms can achieve relatively high accuracy $> 70\%$. As expected, accuracy on the data that contains less drift is higher and contains less uncertainty. On average, accuracy does not exceed 84% for the case of the DBSCAN and the algorithm produces the same amount of clusters at every iteration. When the input data is shuffled, the only "non-deterministic" behavior is in the label for the cluster being assigned, but not the composition of the cluster itself. The behavior was firstly highlighted in the original paper of DBSCAN [7] where the authors claimed that convergence result is independent of the order in which the points of the database are visited expect the "rare" situations. This "rare" situations occur when border points belong simultaneously to two clusters. This border point will be assigned to the cluster that is considered first to avoid the overlap. In other words, there is always the same amount of density-reachable points from a reference point, hence the same amount of the assigned clusters is constant.

In the next test for both algorithms, the global parameters were manually adapted to yield higher accuracy (Figure 6.8 C)). The adjustments to the K, the desired number of clusters in K-means can be set higher than the elbow method recommends, and for the DBSCAN algorithm the ϵ parameter is crucial. This test demonstrates that with the aid of completely manual tuning, higher accuracy $> 80\%$ for any type of dataset can be achieved. Even more, the ϵ parameter in the DBSCAN can be chosen to recover the 100% accuracy on the dataset with the minor drift. N.B. The average convergence time for the DBSCAN algorithm is 0.01 second and for the K-means algorithm 47 seconds on laptop Dell Inspiron 7577.

6.5.3. Benefit of the centroids re-assignment

For the domain of the semiconductor industry, specifically for the lithography process, it is crucial that cleaning can be performed if contamination above a certain size is present on the sample. In this way, for instance, the very small particles are of minor importance for the pellicle layer above the UV mask, and only if the bigger particles are present, cleaning action needs to be taken. In the absence of the pellicle, on the contrary, one should only be concerned with the small contamination landing on the mask [20]. The quantitative description of the surface, provided by the surface scanner in this regard becomes very crucial. The confusion between the different sizes of the scatterers on the sample should be minimal. For our system, the width of the signal changes as scanning through the spherical particle is performed (Figure 6.9).

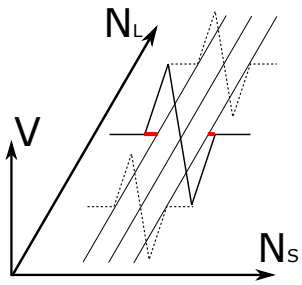


Figure 6.9: Sketch of the signal from an isolated spherical particle visible over three consecutive line scans. The red region represents the increase in the τ width of the signal when the scan line passes through the center of the particle as compared to other consecutive lines $\pm\Delta Y$ (signals as dotted lines).

In the first approximation, all the detected signals can be fed to the histogram as it was done in Section 6.5.1. This approach would work properly if the dataset includes a single particle size or if the contamination is reasonably different. Realistically, samples contain a wide range of particle sizes.

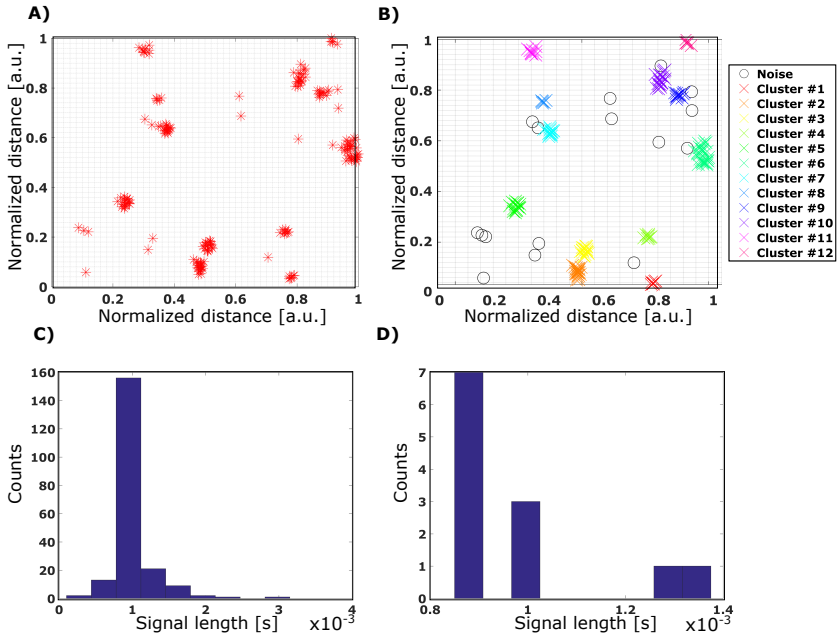


Figure 6.10: A) Zero-crossings of the diff. signals by the search alg. and B) the corresponding isolated particles by converged DBSCAN. The dataset includes minor drift where one scanning line of $\Delta X = 25\mu m$ takes 100 ms. C) Histogram includes all particle-looking signals and D) when the signals corresponding to one particle are clustered, and only one signal per particle (highest V_{pk-pk}) is taken.

If the pulses on the edges of the particle scan are included in the estimation histogram, they will contribute to the interclass confusion.

In the Figure 6.10 we demonstrate the outputs from the signal search algorithm and corresponding clusters defined by the DBSCAN algorithm. This algorithm was chosen since the convergence time is faster than the modified K-means, and it had achieved higher accuracy at the previous test. The region of the sample under study is a good representative of the multi-class sample where additionally to the nominal 60 and 80 nm PSL particles, there are isolated particle-looking signals that are treated as outlier by the algorithm as well as the contamination of bigger particles ≈ 100 nm in diameter. The first approximation histogram includes the side detection from the class of the 100 and 80 nm contributing to the class of 60 nm as well as features between the classes and it seems that there is only a single class present in the data (see Fig. 6.10 C). When all signals that corresponds to one particle are clustered and only the highest V_{pk-pk} pulses from each cluster is assigned as being one particle (Fig. 6.10 B), we observe three separable classes in the histogram (Fig. 6.10 D), showing that this strategy solves the problem of particle size confusion.

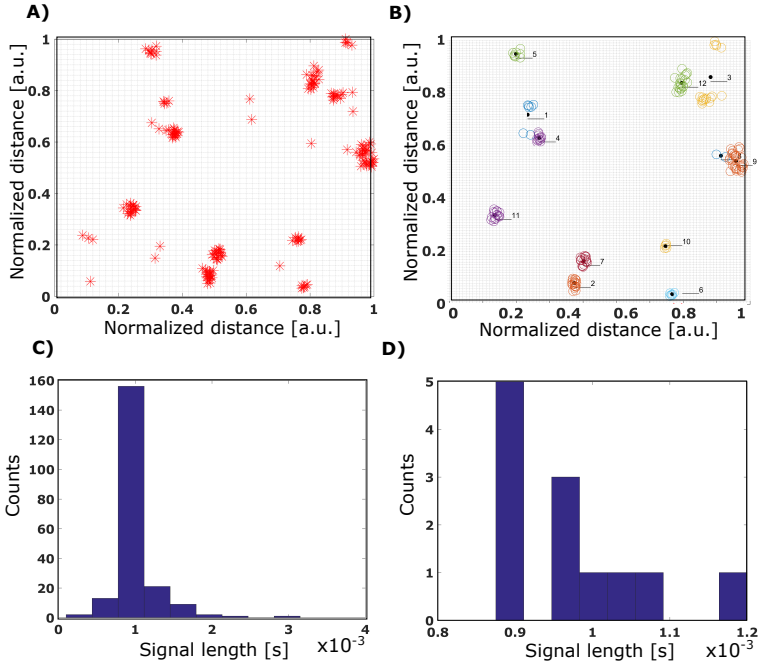


Figure 6.11: A) The zero-crossings of the differential signals by the search algorithm and B) the corresponding isolated particles (after clustering) obtained by converged adapted K-means algorithm. The data set includes minor drift where one scanning line of $\Delta X = 25\mu m$ takes 100 ms. Histograms obtained when C) all the particle-looking signals contribute to the histogram, and D) when the signals are clustered and only one centroid (with maximum V_{pk-pk}) is assigned to represent the particle.

In addition to the better performing algorithm of DBSCAN presented above, we demonstrate the output of the modified K-means algorithm 6.11 B). The difference with the DBSCAN algorithm is a tendency to merge the clusters that would easily be separated by the human eye. Such cluster can be seen as cluster n. 3, which contains two separable groups of points. The solution for this problem can be to re-initialize the algorithm multiple times until the cluster is assigned correctly. Nevertheless, it is more informative to present an average initialization result. The algorithm is capable of separating three classes of particles, as shown in 6.11 D), which is much better than the result of using all the detected signals 6.11 C).

6.6. Discussion and Conclusions

The approach of clustering the data has a downside, namely, the risk of losing the beneficial signals that correspond to very tiny particles. These particles may produce only single scan line with a sufficient SNR, if the selected step between the scan lines Δy is too big. To improve the sensitivity of the algorithm even further, a separate routine could reconsider the outliers. This step could include adding a collection of matched filters operating in the time domain to filter out signals with the expected duration. Alternatively, one can try to establish spectral differences between the particle and non-particle signals (multiple wavelength approach).

While this study considered the detection of polystyrene particles, the technique could also be applied to extract features from a measurement of particles of different materials. Scatterometry is not an imaging technique, and some other features (such as material) can be recovered if one can model them and obtain more diversity in the experimental data. For example, instead of only looking at the time span of the particle signal (related to the size of the particle), one can add its magnitude, which is proportional to the diameter and material of the particle [21]. Another example is the work of Potenza et al. [22] where using a similar technique, they were able to recover the complex index of refraction of the particles, and in this way, reveal their material.

DBSCAN can yield higher accuracy than the K-means subroutine in the case of when the scale of the data is well understood. Also, the convergence of DBSCAN algorithm is fast. Nevertheless, there is still room for implementing the K-means routine because the sensitivity to the hyperparameters is much higher in the case of the DBSCAN, including the complete failure in defining the clusters from the initial data. The K-means, on the contrary, can be considered as a more robust algorithm that yields relatively high accuracy, and at any initialization will always define a certain amount of clusters. The K-means algorithm is scalable to large datasets while the DBSCAN can suffer from the curse of dimensionality [23]. A final point to consider is when working on datasets with severe drift and wide range of the particle sizes, the DBSCAN can fail to cluster data with significant differences in densities [24].

Throughout the IC manufacturing process, large amounts of data need to be mined in a fully automated mode [25]. With a growing amount of the data, we can envision that the line-by-line analysis of the dataset can become computationally slow. Also, the total amount of hyperparameters is significant. Search and clustering routine in total has up to 6 parameters fed by the user. Hence, in the next chapter, we will explore the potential of methods for handling big data, such as deep learning and CNN [26, 27].

In this chapter, we have developed a generalized framework that accurately extracts features of the differential signal produced by the scattering of a nanoparticle and uses these features for particle location and size determination. We have combined pre-processing with search algorithms based on thresholding, such as peak-to-peak amplitude, and the width in time of the signal. The proposed method makes use of unsupervised clustering techniques to separate particles with high density on the samples. We adapt algorithms of DBSCAN and K-means and use them together with the simple prior.

We have tested the framework for the dataset with high density of the particles, presence of large experimental noise, and drift present in the dataset. The accuracy of the algorithm resulted in the 84% for the hyperparameters set semi-automatically, and the 100% accurate result for manually-tuned parameters. The algorithm of DBSCAN is a go-to solution because it works much faster than K-means. However, the latter is more robust because it is less sensitive to the change on the input parameters. While we tested the framework for the particular case of experimental data obtained with CFS, this method can be generalized to other experiments that involve measurements with differential detection, such as coherent time-addressed optical CDMA systems [28] and ferromagnetic resonance spectrometers (VNA-FMR) [29]. In these techniques, the data set might include mechanical vibrations or other experimental fluctuations, similar to the drift studied in this chapter. We believe that the proposed framework is an essential addition to the nanoparticle detection experimental community.

References

[1] D. Kolenov and S. F. Pereira. Machine learning techniques applied for the detection of nanoparticles on surfaces using coherent fourier scatterometry. *Opt. Express*, 28(13):19163–19186, Jun 2020.

[2] Jinlong Zhu, Renjie Zhou, Lenan Zhang, Baoliang Ge, Chongxin Luo, and Lynford L. Goddard. Regularized pseudo-phase imaging for inspecting and sensing nanoscale features. *Opt. Express*, 27(5):6719–6733, Mar 2019.

[3] Sanyogita Purandare, Jinlong Zhu, Renjie Zhou, Gabriel Popescu, Alexander Schwing, and Lynford L. Goddard. Optical inspection of nanoscale structures using a novel machine learning based synthetic image generation algorithm. *Opt. Express*, 27(13):17743–17762, Jun 2019.

[4] Mark-Alexander Henn, Hui Zhou, and Bryan M. Barnes. Data-driven approaches to optical patterned defect detection. *OSA Continuum*, 2(9):2683–2693, Sep 2019.

[5] Joe H. Ward Junior. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[6] Stuart P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28:129–137, 1982.

[7] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han, and Usama M. Fayyad, editors, *KDD*, pages 226–231. AAAI Press, 1996.

[8] D. Kolenov, R. C. Horsten, and S. F. Pereira. Heterodyne detection system for nanoparticle detection using coherent Fourier scatterometry. In Peter Lehmann, Wolfgang Osten, and Armando Albertazzi Gonçalves Jr., editors, *Optical Measurement Systems for Industrial Inspection XI*, volume 11056, pages 336 – 342. International Society for Optics and Photonics, SPIE, 2019.

[9] Murugappan Murugappan, Subbulakshmi Murugappan, and Bong Siao Zheng. Frequency band analysis of electrocardiogram (ecg) signals for human emotional state classification using discrete wavelet transform (dwt). *Journal of physical therapy science*, 25:753–759, 07 2013.

[10] Khaled Daqrouq. Ecg baseline wander reduction using discrete wavelet transform. *Asian Journal of Information Technology*, 4:989–995, 01 2005.

[11] Orlando Ramirez Barron. k-means clustering, 2019. https://tinyurl.com/y2gsmhoo.

[12] S. Mostapha Kalami Heris. Implementation of dbscan clustering in matlab, 2015. https://tinyurl.com/y6te7247.

[13] M. K. Pakhira. A fast k-means algorithm using cluster shifting to produce compact and separate clusters. *International Journal of Engineering, Transactions A: Basics*, 28(1):36–45, 2015.

[14] Xin Jin and Jiawei Han. *K-Means Clustering*, pages 563–564. Springer US, 2010.

[15] William Andrew. *Handbook of Silicon Based MEMS Materials and Technologies*. Elsevier Inc., 2010.

[16] Jianning Yang, Jun Sun, and Jun Ni. Removing DC offset and de-noising for inspecting signal based on mathematical morphology filter processing. In Yetai Fei, Kuang-Chao Fan, and Rongsheng Lu, editors, *Fourth International Symposium on Precision Mechanical Measurements*, volume 7130, pages 1146 – 1152. International Society for Optics and Photonics, SPIE, 2008.

[17] K. Brzostowski. An algorithm for estimating baseline wander based on nonlinear signal processing. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–5, Sep. 2016.

[18] Timothy Sauer. *Numerical Analysis*. Addison-Wesley Publishing Company, 2nd edition, 2011.

[19] Zhemin Chen, Pengbing Hu, Qingqiang Meng, and Xinyong Dong. Novel optical fiber dynamic light scattering measurement system for nanometer particle size. *Advances in Materials Science and Engineering*, 2013, 11 2013.

[20] Derk Brouns et al. NXE pellicle: offering a EUV pellicle solution to the industry. In Eric M. Panning, editor, *Extreme Ultraviolet (EUV) Lithography VII*, volume 9776, pages 567 – 576. International Society for Optics and Photonics, SPIE, 2016.

[21] S. Roy, K. Ushakova, Q. van den Berg, S. F. Pereira, and H. P. Urbach. Radially polarized light for detection and nanolocalization of dielectric particles on a planar substrate. *Phys. Rev. Lett.*, 114:103903, Mar 2015.

[22] Marco Potenza, Tiziano Sanvito, and Alberto Pullia. Measuring the complex field scattered by single submicron particles. *AIP Advances*, 5:117222, 11 2015.

[23] Machine learning course Google Dev. K-means advantages and disadvantages, 2019. https://tinyurl.com/ydjqw4d2.

[24] Hans-Peter Kriegel, Peer Kröger, Jörg Sander, and Arthur Zimek. Density-based clustering. *WIREs Data Mining and Knowledge Discovery*, 1(3):231–240, 2011.

[25] V. Sze, Y. . Chen, T. . Yang, and J. S. Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329, 2017.

[26] N. G. et al. Orji. Metrology for the next generation of semiconductor devices. *Nature Electronics*, 1(10):532–547, 2018.

[27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[28] Sun-Jong Kim et al. All-optical differential detection for suppressing multiple-access interference in coherent time-addressed optical cdma systems. *Opt. Express*, 12(9):1848–1856, May 2004.

[29] S. Tamaru, S. Tsunegi, H. Kubota, and S. Yuasa. Vector network analyzer ferromagnetic resonance spectrometer with field differential detection. *Review of Scientific Instruments*, 89(5), 2018.

7

Convolutional Neural Network applied for nanoparticle classification

In this chapter, we investigate the effectiveness of the novel concepts from the domain of computer vision, such as deep neural networks and anomaly detection when they are applied to the datasets of CFS.

We encourage interest in deep learning techniques in Section 7.1. In Section 7.2, the architecture of the multi-output classification of the convolutional neural network is proposed and configured. Furthermore, in Section 7.3, we practice and examine the ability of the trained model to refuse unknown inputs if we change the penultimate layer of the network. We further conduct a comparative test with the ML method of Chapter 6. Finally, we provide Conclusions and Discussion in Section 7.4.

Parts of this chapter have been published in Applied Optics **59**, 8426-8433 (2020) [1]

7.1. Introduction

Recently, there has been a growing interest in deep learning which has demonstrated its feasibility to significantly improve optical microscopy, enhancing its spatial resolution over a large field of view and depth of field [2], analysis of medical images [3], analyzing Through-Focus Scanning Optical Microscopy (TSOM) images of nanostructures [4], detecting and localizing holographic features [5] and many other application areas in optics and physics [6, 7]. Deep learning algorithms are part of a broader family of machine learning algorithms, which can be considered as a network consisting of multiple neural layers with the idea to progressively extract higher level features from the raw input, otherwise known as learning on the representation of the data [8]. Examples include: deep neural network DNN [9], recurrent neural network RNN [10, 11], long short-term memory LSTM [12] and convolutional neural network CNN [13, 14]. The feasibility of CNN has been demonstrated by wafer map defect pattern classification using simulated wafer maps (synthetic data) [15], relying on scanning electron microscope (SEM) images for classification of defects and contamination [16] and recently defining the chemical composition of particle defects on semiconductor wafers by merging the SEM image data with Energy Dispersive X-rays Spectroscopy (EDX) spectral data as input [17].

In the previous chapter, the $2D$ scattered maps generated by the CFS technique have been studied with line-by-line search algorithms resulting in histograms that rely on the features of characteristic electronic signals that are generated when a particle is detected. The cumbersome search routines are associated with the hyperparameters defined by the user for each specific input dataset, e.g. an expected amplitude and width of the characteristic signal, density and number of points in a cluster of isolated particles, and the zeroing parameters for different iterations of the search [18]. There is still a need for discussion on minimization of user-algorithm-interaction [19, 20]. However, it is true that considering a growing amount of data and pressure to inspect data more quickly, the line-by-line analysis of the data set can become computationally slow. The additional challenge is the precise categorization of killer-particles using automated contamination or defect classification. In other words, the confusion between the different sizes of the particles on the sample should be minimal. Moreover, while there are solutions to all of these problems, there is a cost associated with each of them. For instance, the application of neural networks requires re-training the network when the physical parameters change, which will consume a lot of resources and time. On the other hand, when pre-trained and deployed, the classification runs almost at no time and virtually no a-priori parameters nor additional tuning of the network are required.

In this chapter, we propose a method to classify the scattered maps of isolated nanoparticles using CNN. We utilize calibrated samples of PSL spin-coated on wafer, with diameters ranging from 40 to 80 nm to collect the training data. Polystyrene particles are standard for the calibration of surface inspection tools because they have well-characterized optical properties (low index of refraction, thus most challenging to detect) and a very tight monodisperse size distribution [21]. Furthermore, for the classification, we study the areas of the wafer where the particle is absent, contributing to the "background" class. We also target at a novelty detection, by looking at ways for the network to separate the "unknown" class from the input data, i.e., classes that

have been unseen in the training. In order to do that, we rely on a simple approach of baseline and also a more sophisticated approach of introducing the OpenMax layer [22]. We carried out an experiment by adding noise to the scattered maps (degrading thus the SNR), fooling images as well as some unfamiliar (reversal) images to the network. As one of the main goals for the CNN is to accurately discriminate between classes, we study not only the monosized but also samples that contain multiple classes of PSL particles with diameters of 40 and 50 nm, 50 and 60 nm, 60 and 80 nm (see preparation of samples at Appendix C). The results show that our model can successfully discriminate between the proposed 5 classes with an accuracy up to 95%. By providing the samples that were unseen during the training, our results for the first time highlight the importance of the novelty detection to capture the confusing inputs in a contamination detection problem. The results show that the proposed method has superior capabilities compared to the classification with the traditional search algorithm [18]. The dataset and the codes used to generate the typical results of this chapter are available online [23, 24].

7.2. Method

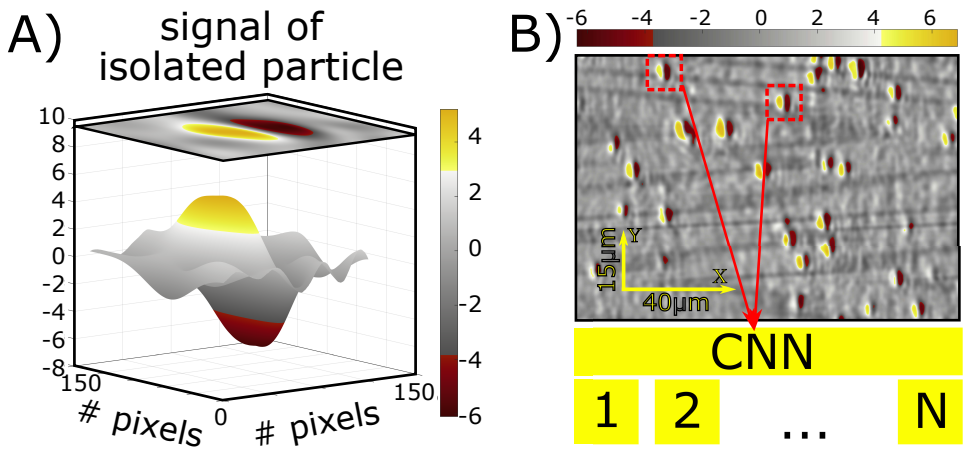


Figure 7.1: A) Example of 2D map with the signal that is obtained when an isolated particle is recorded by the balanced detector. B) Schematic process of the CNN-based classification. The dotted red boxes indicate that the images are cut out. The input to the CNN are cut-out far-field maps containing the detection of nanoparticles; the output of the CNN is the label of the particle diameter $1, 2, ...N$. The example image size is $40x15\ \mu m^2$ represented by arrows.

The proposed CNN algorithm takes a set of "images" (signal intensity maps) as the network input and outputs the class labels (see Figure 7.1 B). The discretization of the data is due to the sampling speed of the NI 5922 acquisition board and selection of the scanning step between lines. The scale in the Figure 7.1 A) (150x150) is given in pixels with each pixel corresponding to 2 nm in x-direction and 4 nm in the y-direction. The intensity maps (2D scan maps) generated from the data points arranged line per

line, according to the raster scan pattern.

The raw data yields the detection of numerous isolated particles. The density of deposited particles has been chosen in order to have a considerable number of detected particles in a scan area of e.g. $40x15\ \mu m^2$ (Figure 7.1 B). Other signals are also present, corresponding to clusters of particles, particle deposition residues and possibly cross-contamination. Since directly using all particle-like detections from a sample are not always possible and would not give a high-quality dataset, we performed manual labeling. The type of the particle signal that is present with the highest density inside the global scan area is the representative for the nominal size. The bounding box is placed such that the particle signal is fully visible in the region of interest (the dotted red boxes Figure 7.1 B). This square cut is centered about the position of the maximum amplitude of the differential signal. For the background class, the particle signal pattern (positive and negative amplitude) of the particle should be absent. We have grouped the images into the classes of 40, 50, 60 and 80 nm particles and the "background" class that corresponds to the areas of the sample without particles (see Fig. 7.2 A)). We created a class-balanced dataset (see Table 7.1) with roughly 260 images per class. The total amount of 1302 images are fed to the network, and we use the $60-20-20$ split for training, validation and testing. Here we ensure that all three sets contain representative examples by randomly splitting data from each class into three parts and then merging to form the unbiased sets.

Table 7.1: The amount of images per class (original dataset).

Class	40 nm	50 nm	60 nm	80 nm	background	total
# of images	254	253	276	272	247	1302

We use the holdout method for validation, meaning that after each epoch, the validation dataset is passed through the network. When the training is complete, we show the test set to the network for a single time. The amount of images required for the training of the network, contrary to the expectations, turn out to be relatively small, presumably due to the simple pattern of the particle signal. We did not apply any geometric transformation to the experimental data, thus the input data only contains the diversity of examples due to the inherent experimental conditions. As shown in Fig. 7.2 B) we use the network architecture where no manual feature selection is necessary. The simple deep neural network was composed of repeated units of convolutional layers, which number and sizes are chosen to have a balance between speed and low $error\ rate = (1 - Accuracy)$ (Figure 7.2 C).

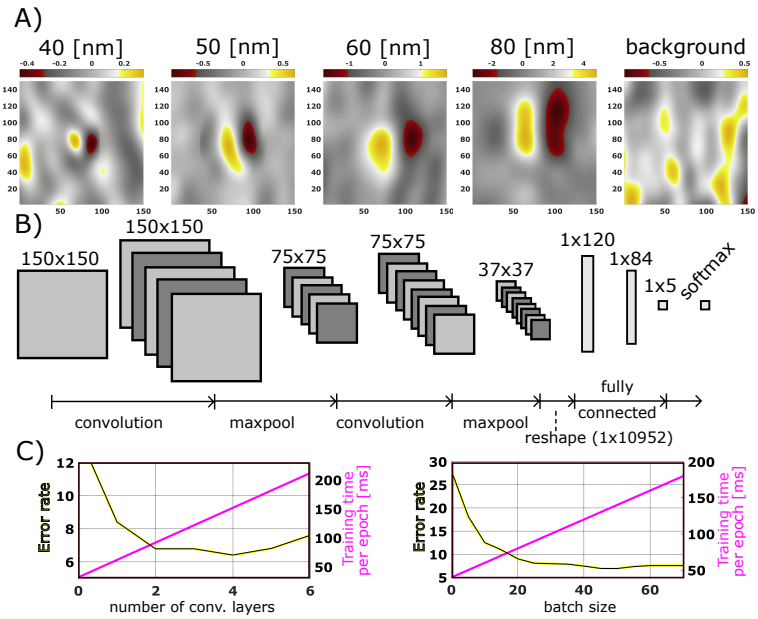


Figure 7.2: A) Examples of the five output classes. B) The architecture consists primarily of convolutional layers capable of extracting relevant features of the input samples. Three fully connected layers at the end serve as a decision layer, mapping the automatically extracted features to the desired output class. C) Error rate as a function of the number of convolution layers and batch sizes. From the plots, we see that 2 convolution layers and batch size of 15 are the optimal choice for the architecture, since they introduce a good balance of training time and low-enough error rate

The final architecture includes input size of 150×150 pixels and two convolution layers operating with filter (kernel) sizes of 5×5 pixels. The amount of filters in the first convolutional layer is 5 and in the second is 8, stride is 1. In between and after the convolutional layers, we have inserted two max-pooling layers with size of 2×2 pixels and stride 2, effectively reducing the image resolution by a factor of 2 at each step. The purpose of these layers is to reduce computation for consecutive layers and to provide a form of translation invariance. All convolutional layers have rectified linear unit (ReLU) activation. Each ReLU in the network is followed with the Batch normalization [25]. The final maxpool layer is fed into three fully connected layers of size 120, 84 and 5 respectively. Final layers are necessary to learn the relationship between the learned features and the sample classes, which is in our case is five. Finally, the logits are converted to the probability scores by the SoftMax function. The network's output is used to compute the mean-square error between the true label and the predicted label, also known as the cross entropy loss. We used the Adam optimization scheme with a global learning rate of 0.001 to minimize this loss function. The total number of weights in the network that are being updated during the training process is 1326087. We built a Pytorch [26] implementation and move it

to the GPU (NVIDIA GTX1050 Ti) calculation with tensors. More information about the implementation can be found in Ref. [23, 24].

7.3. Results

7.3.1. Closed set classification

The best model has to be selected based on the Accuracy metric calculated on the test data. For the closed set of five classes, as according to Table 7.1, we found that after approximately 12 epochs (12 times through all the training examples), the loss no longer decreased significantly (Figure 7.3 A). The top performing network ($Accuracy = 95\%$) was stored to be used in further tasks.

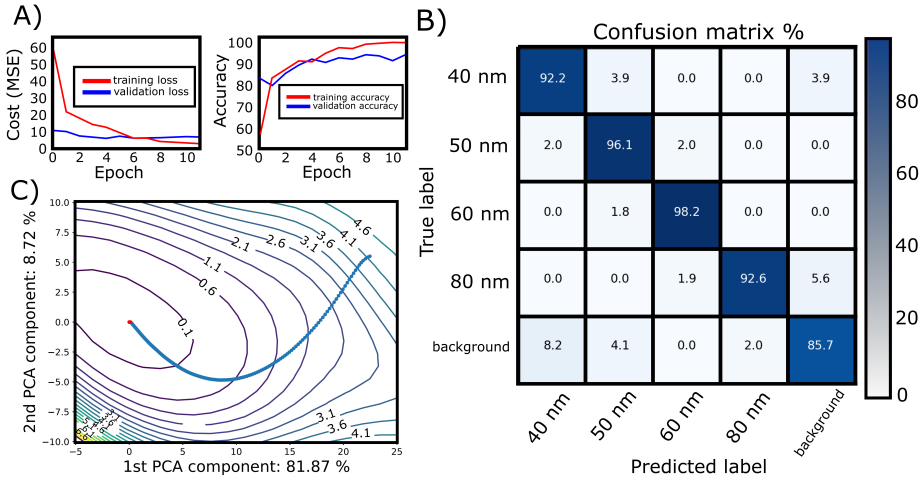


Figure 7.3: A) The training and validation loss - left and accuracy - right, is evaluated across different number of epochs based on the optimal parameters (Figure 7.2 B) architecture of the CNN model. B) The accuracy for the test set in the confusion matrix. C) 2D visualization of the loss surface of the CNN model with the projected learning trajectories using normalized PCA direction (batch size of 20, Adam optimizer and 15 epochs of training).

In order to see which classes the network struggles to distinguish and to what degree, we built the confusion matrix (Figure 7.3 B). The horizontal axis represents the particle classes predicted by our model, and the vertical axis represents the true input image labels. For example, the 80 nm row (the 4th row in the matrix) indicates that 92.6% of the images labeled with 80 nm are correctly predicted as 80 nm; 1.9% of them are incorrectly predicted as 60 nm; 5.6% of them are predicted as "background". Our experiment shows that 50 is often confused with 40 nm. The reason is that there is a small difference between the scattering cross-sections generated by these two particle sizes (the scattering varies by the sixth power of the diameter [21]). It is clearly visible that most misclassification involves the background class. Finally, we built the landscape [27], where we demonstrate the convergence to minima as our learning procedure follows the loss in a gradual manner. The projected learning

trajectory is estimated using normalized principal component analysis (PCA) directions. The squared nature of the loss function leads to mostly convex loss landscape (Figure 7.3 C).

7.3.2. Comparison with thresholding classification method

We compared the performance of our CNN classifier, pre-trained on the five classes (Section 7.3.1) with a method of the previous chapter. We did that on new test sets of separately 40, 50, 60 and 80 nm particles, with roughly 40 cut-out images per class. The thresholding classification method can be summarized as:

- Search line-by-line for signals that have characteristic shape (positive-negative pulses) and that are close to the expected amplitude and time-width of the particle in question;
- Use the Density-based spatial clustering of applications with noise (DBSCAN) algorithm to define the group of signals that are attributed to a single scatterer and return the estimate of the time-width from centroid. By group of signals, we mean that the signal should repeat itself at the same X position in a few consecutive scan lines (in the y-direction);
- Use a calibration curve based on the time-width of the signal as a function of the particle size to return a class label for the particle.

This method operates on the reference positions of the cut-out images from the corresponding raw scan maps. We keep the number of the output classes equal to five to provide a fair comparison hence in the thresholding method instead of the background, the class of 100 nm particle is present.

Table 7.2: Comparison of accuracy per class between the proposed CNN and method based on thresholding and search.

	40 nm (35 images)	50 nm (37 images)	60 nm (37 images)	80 nm (46 images)
Thresholding	0.37	0.43	0.63	0.82
CNN	0.97	0.94	1	1

In Table 7.2, we present the classification performance of thresholding and CNN approaches on the four test sets. From the results, we can see that the classifier based on the neural network achieves better performance as compared to the classical search routine. Both approaches perform very accurate on the data of 80 nm particle class, but when reducing the size of a particle, the accuracy drops much faster in the case of the thresholding method.

It is critical to note that both approaches can consider the 2D local information inherent to our measured data. In the case of the thresholding approach, positive-negative signals that are present in consecutive scan lines at the same X positions are clustered and considered as a single particle (see Fig. 7.1). In the case of CNN, the convolution filter can extract the spatially-connected information by walking over the

image. It is unlikely thus that improved classification accuracy is due to the 2D nature of convolutional kernels. The essential difference is the ability of CNN to extract the higher-level representation by cascading the filters and learning on these representations. On the contrary, classification based on the calibration curve (signal feature as a function of a particle diameter) always relies on the representation of the data that are manually engineered.

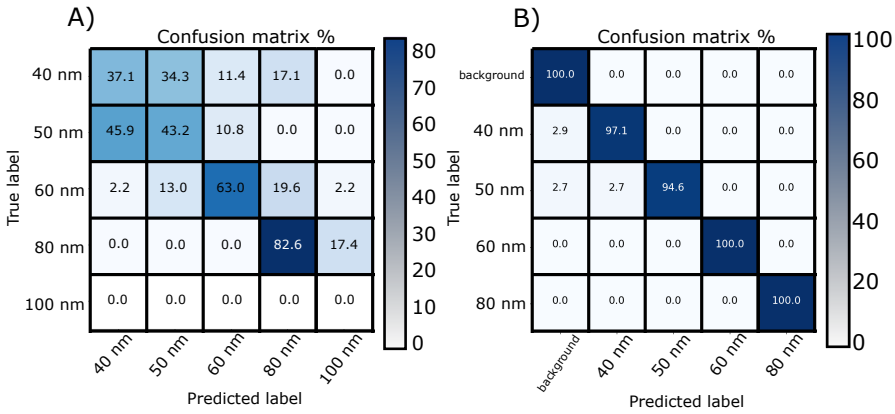


Figure 7.4: Confusion matrices comparing classification tendencies between predicted and true labels by A) thresholding and B) CNN approach. Squares are colored based on the value of the cell, with darker colors indicating more matches. Values along the diagonal of each confusion matrix represent the images classified correctly, while values in off-diagonal regions represent blurring between types of classes.

To address further the classification tendencies as they appear in methods under comparison, we demonstrate the confusion between the classes. The clustering method performed a lot worse, where the overall accuracy was 56%, with relatively accurate results for the 60 and 80 nm classes yet with a lot of confusion on the 40 and 50 nm classes. Evidently, in both approaches, the confusion between the neighboring classes is present (see Figure 7.4). For the thresholding method, it is clear that the steeper the calibration curve would become, the less confusion would be present. Inherently, this reference method relies on the time width, that showed to be a sensitive parameter for the case of particles $\geq$ 100 nm [28]. We point out that classification of the single particle image takes 0.89 [s] for the case of the thresholding algorithm and 0.02 [s] for the pre-trained network of this chapter. Generally speaking, other features can be selected in order to allow for better discrimination, which addresses the topic of feature engineering. At the same time, not so deep CNN is enough to pick up on patterns in many different features of an input. Features extracted by the network extend far beyond these that make sense to the human eye, such as, maxima, minima or the pattern size. A CNN trained to recognize particles might find other features such as patches of color, topography or background, which can become even stronger predictors.

7.3.3. Towards Multi-class Open Set Classification

Making alterations to the regular input data even in the form of tiny changes that are typically invisible to humans can mislead the best neural networks. These problems are not easy to solve because CNNs are fundamentally fragile. As shown in the previous two sections, the accuracy in classification is very high, but this is possible until networks are taken into unfamiliar territory where they can break in unpredictable ways. To bring a spotlight on the problem of confusion by the so called "adversarial examples" [29], scientists have evolved images that look like an abstract pattern but which the DNNs see as familiar objects [30, 31]. In the context of CNN applied for the classification of the different particle size-contamination, we should envision that distorted measurement data or other types of untrained particles could also be spotted by the proposed network.

The output layer of the original architecture at Figure 7.2 B) is SoftMax layer that contains the vector of probabilities

$$P(y = j|\mathbf{x}) = \frac{exp(\mathbf{v}_j(x))}{\sum_{i=1}^{N} exp(\mathbf{v}_i(x))}, \tag{7.1}$$

where x is the sample image and **v(x)** is the corresponding activation vector. The number of classes is $j = 1, ..., N$ and i is the index that goes over the classes. Due to the summation in the denominator, the probabilities are normalized and sum up to one. We want to build a $(N+1)$-classifier $f(x)$ with the classes $C = \{d_1, d_2, ..., b, rejection\}$. The most straightforward approach of a novelty detection is to introduce the baseline value for the scores of the SoftMax layer. If the probability of the output classes is not high-enough, the input image is assigned with the unknown label. The novelty score NS is defined as

$$NS = 1 - max(P(y = j|\mathbf{x})). \tag{7.2}$$

The procedure of computing the OpenMax probabilities includes 4 steps:

1. For each class $C = [c_{j,...,N}]$, the mean activation vector is computed $MAV = [\mu_{j,...,N}]$, where $\mu_j = mean(v_j(x_{i,j}))$ and $x_{i,j}$ represents the correctly classified sample.
2. Per class, fit the Weibull model with parameters $p_{c_j} = (t_{c_j}, \lambda_{c_j}, k_{c_j})$ to the distance between the input sample and the mean of the set of η number of outlier examples of class j. t_{c_j} is used for shifting the data, λ_{c_j} and k_{c_j} are the scale and shape parameters derived from the training data of the class c_j and control the cumulative density function (CDF). For more details on Weibull distribution and Extreme value theory, see the Reference [32].
3. Estimate the Weibull CDF probability on the distance between sample x_i and known class' mean activation vector: $MAV[\mu_{j,...,N}]$ defined as w(**x**). Re-calibrate activation vector by $\hat{v}(x) = \mathbf{v(x)} \circ w(\mathbf{x})$. To allow the novelty detection, augment output to $N + 1$ classes by $\hat{v}_{N+1}(x) = \sum_i \hat{v}_i(x)(1 - w_i(x))$.
4. To support explicit rejection, pseudo-probability of an unknown class is estimated

from known class's activation scores

$$\hat{P}(y = j|\mathbf{x}) = \frac{exp(\hat{\mathbf{v}}_j(x))}{\sum_{i=0}^{N} exp(\hat{\mathbf{v}}_i(x))}, j = 1, ..., N + 1. \quad (7.3)$$

The third way of dealing with the unknown input is similar to OpenMax; however, it is much simpler and essentially relies on the mean activation vector:

1. Calculate the MAV for the correct classifications of each class.

2. For each image x in the train and validation sets, obtain the activation vector $v(x)$ and predicted class $c(x)$. Then, calculate the distance to the MAV with $d = ||v(x) - MAV_{c(x)}||$. Save values of d separately for correct and incorrect classifications.

3. For each image in the test set, calculate d in the same way, and if it is above some threshold, reject the classification (thus classifying it as unknown).

Thus, we utilize and compare three approaches of "Baseline", OpenMax, and distance to MAV approach in order to catch open set examples each time showing the unseen images to the network without additional training. We introduce an input sample with high noise, where the Gaussian and $1/f$ noise were added to every image. For instance, samples that corresponds to the 80 nm (see Figure 7.5 A) were modified such that the SNR decreased by -9.7 dB (1.5 times), fooling images of an elephant from $Animal - 10$ dataset [33], and finally, the mirrored image along x axis of the 2D scattered maps with detected particles, which is representative for the image of a defect such as a small pit. It is not rare that studies on detecting fooling/adversarial images have a narrow focus on optimizing the output or penultimate layer of the network such that the probability for an unknown image is low. This is something of a pitfall because it is possible to optimize the score of rejecting the unknown class with the cost of losing a significant number of the correct images. With this intention, when making the comparison, we apply the network with different output layers for both the original and fooling dataset. In different scenarios, we observe similar behavior. The vast majority of the original images lie in the shallow uncertainty region, which is an indication of a highly accurate network. However, the incorrect ones span nearly the same range as the correct ones, meaning we cannot completely dismiss incorrect classifications by thresholding uncertainty. No matter what threshold we would set, we would always have some incorrect classification. Further, we chose this threshold such that we reject approximately 10% of standard data. This is an arbitrary value, since the acceptable maximum rejection of standard data would depend on the application, and on how frequently images appear in the data that should be rejected.

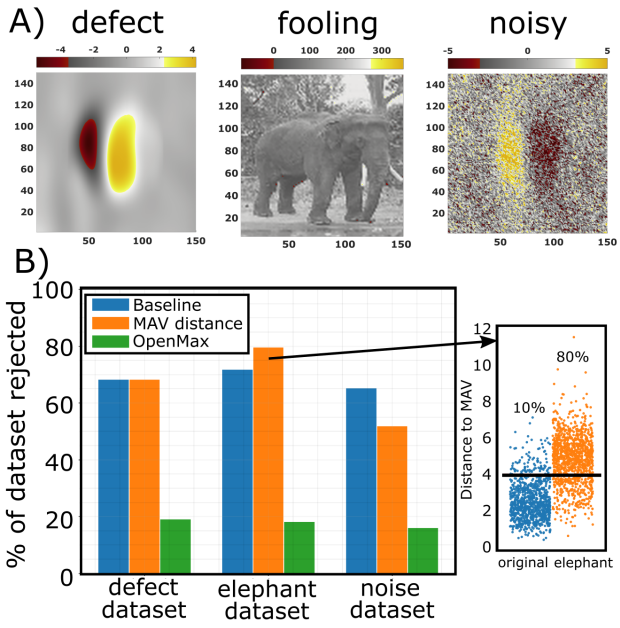


Figure 7.5: A) Three types of open-set examples: image of defect, the fooling image of the elephant, the noisy particle image. B) Summary of comparison between three unknown detection methods as applied to the open-set examples ≈ 1090 images per type. Inset: Example of the best performing case, where the "Distance to MAV" method is applied to the fooling set of elephants. Blue points represent the complete original set passed through the network and the orange points represent the fooling set. If the threshold (in red) is set such that only the 10% of the "good" images are dropped, then the same network can capture 80% of the unknown images.

As a result, we compare the amount of rejection in open-set examples by the three supervisor approaches in Figure 7.5 B). In particular (inset figure), for the fooling dataset, the best performing algorithm is the one based on the distance to MAV, where, we can see how it possible to separate the 80% of the fooling images. We pass the entire original dataset (blue points) and the whole fooling elephant dataset (orange points) through a trained network and set the threshold based over the information from the mean activation vectors. The additional information gained from using the entire vector of outputs rather than just the maximum (novelty score) helps with rejecting unknown inputs there. For the defect set, it performs no better than the Baseline approach, and on the noise set, it performs significantly worse. Finally, the OpenMax gave us poor results on all sets.

7.4. Discussion and Conclusions

The Weibull model is the core of the OpenMax approach. As an essential part of the algorithm, the method selects the m highest activations per activation vector. Our data

has only 5 classes, meaning our activation vectors have only 5 entries. There is not much selection possible in this case. The original paper of OpenMax [22] visualizes activation vectors for a 450 class system and provides some intuition about where the information resides that is used in OpenMax. It is thus clear that, with low amount of classes, the CDF distribution would be very discrete, and a lot less information could be gained from it.

Instead of having the supervisor in the network, such as baseline or MAV, one can include the fooling images as a part of the training data; in particular, to regularly expose the network to problematic cases. In this form, the output layer would explicitly contain the desired class. However, training a network to withstand one kind of "unknown" image could weaken it against others [34].

In case the network performs poorly, inspecting the images that contribute to the off-diagonal elements of the confusion matrix allows us to study the data better and, if required, remove inputs from the dataset to get to the higher accuracy. There are also existing approaches where the subsets of outlier images are removed from the training or test data. These are the interactive learning-based methods for curating datasets using user-defined criteria [35, 36].

To deploy the network in the in-line fab inspection scenario, more output classes should be provided, approaching the real-world situation with a variety of particle sizes on the surface. The synthetic data could replace the types of particles not available experimentally for calibration.

In this chapter, we have applied CNN to 2D maps obtained using coherent Fourier scatterometry for nanoparticle detection and classification. We trained a convolutional neural network to recognize four classes of nanoparticles and surface background class. Based on a total of 1302 experimental images rather than synthetic data, with a simple CNN with two convolutional layers and batch normalization, we demonstrated 95% accuracy on the test data. The proposed approach outperforms existing algorithm for the analysis of the scattered maps that is based on thresholding and search [18]. For relatively small particles, with diameters (classes) of 40 and 50 nm, the accuracy has been improved by a factor of 2. Besides, when studying the amount of misclassification presented by both methods, we see that the CNN can cope better in separating the nanoparticles that produce very similar scattering cross-section (such as particles with diameters of 40 and 50 nm). The demonstrated increase in accuracy and minimized confusion could be attributed to the fact that CNN automatically extracts the features from the proposed data, while the search approach only looks at manually engineered features. Further, we experimented with selecting the best approach to capture the images unseen during the training of the network. The output layer of the proposed CNN can be augmented with a method based either on mean activation vectors, OpenMAX or simple baseline approach. Depending on the need, the threshold value for the uncertainty of the unknown images can be introduced such that as we show experimentally, e.g. 80% of the fooling images of an elephant can be neglected at the cost of only dropping 10% of the particle-type dataset. We believe that the proposed CNN is an essential addition to the nanoparticle detection and classification.

References

[1] D. Kolenov, D. Davidse, J. Le Cam, and S. F. Pereira. Convolutional neural network applied for nanoparticle classificationusing coherent scatterometry data. *Appl. Opt.*, 59(27):8426–8433, Sep 2020.

[2] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, Nov 2017.

[3] S. Kevin Zhou, Hayit Greenspan, and Dinggang Shen. *Deep Learning for Medical Image Analysis*. Elsevier Inc., 1 2017.

[4] Yufu Qu, Jialin Hao, and Renju Peng. Machine-learning models for analyzing tsom images of nanostructures. *Opt. Express*, 27(23):33978–33998, Nov 2019.

[5] Mark D. Hannel, Aidan Abdulali, Michael O'Brien, and David G. Grier. Machine-learning techniques for fast and accurate feature localization in holograms of colloidal particles. *Opt. Express*, 26(12):15221–15231, Jun 2018.

[6] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.

[7] Kavitha Kuppala, Sandhya Banda, and Thirumala Rao Barige. An overview of deep learning methods for image registration with focus on feature-based approaches. *International Journal of Image and Data Fusion*, 0(0):1–23, 2020.

[8] Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[9] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[10] Michael I. Jordan. Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 531–546. Hillsdale, NJ: Erlbaum, 1986.

[11] Barak A. Pearlmutter. Learning state space trajectories in recurrent neural networks. *Neural Computation*, 1:263–269, 1989.

[12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[13] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980. Cited By :1721.

[14] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, Dec 1989.

[15] Takeshi Nakazawa and Deepak Kulkarni. Wafer map defect pattern classification and image retrieval using convolutional neural network. *IEEE Transactions on Semiconductor Manufacturing*, PP:1–1, 01 2018.

[16] Shota Monno, Yoshifumi Kamada, Hiroyoshi Miwa, Koji Ashida, and Tadaaki Kaneko. Detection of defects on sic substrate by sem and classification using deep learning. In Fatos Xhafa, Leonard Barolli, and Michal Greguš, editors, *Advances in Intelligent Networking and Collaborative Systems*, pages 47–58, Cham, 2019. Springer International Publishing.

[17] Jared OrLeary, Kapil Sawlani, and Ali Mesbah. Deep learning for classification of the chemical composition of particle defects on semiconductor wafers. *IEEE Transactions on Semiconductor Manufacturing*, PP:1–1, 01 2020.

[18] D. Kolenov and S. F. Pereira. Machine learning techniques applied for the detection of nanoparticles on surfaces using coherent fourier scatterometry. *Opt. Express*, 28(13):19163–19186, Jun 2020.

[19] F. Hutter, J. Lücke, and L. Schmidt-Thieme. Beyond manual tuning of hyperparameters. *KI - Kunstliche Intelligenz*, 29(4):329–337, 2015. cited By 33.

[20] Y. Sun, B. Xue, M. Zhang, and G. G. Yen. Completely automated cnn architecture design based on blocks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(4):1242–1254, 2020.

[21] Elton Williams Howard R. Huff, Randal K. Goodall. Measurement of Silicon Particles by Laser Surface Scanning and Angle-Resolved Light Scattering. *Electrochem. Soc.*, 144:243–250, 1997.

[22] Abhijit Bendale and Terrance E. Boult. Towards open set deep networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1563–1572, 2015.

[23] Davy Davidse and Dmytro Kolenov. Cnn for nanoparticle. https://github.com/ddavidse/CNN_for_nanoparticle, 2020.

[24] Dmytro Kolenov and Davy Davidse. Training and test data for the preparation of the article: Convolutional neural network applied for nanoparticle classification using coherent scaterometry data. https://doi.org/10.4121/uuid:516ab2fa-4c47-42f8-b614-5e283889b218, 2020.

[25] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 448–456. JMLR.org, 2015.

[26] Adam Paszke and Sam Gross. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[27] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets. In *Neural Information Processing Systems*, 2018.

[28] S. Roy. *Sub-wavelength metrology using Coherent Fourier Scatterometry*. PhD thesis, TU Delft, Prometheusplein 1, 2628 ZC Delft, 9 2016. doctoral thesis.

[29] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *International Conference on Learning Representations*, 2014.

[30] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, October 2019.

[31] Douglas Heaven. Why deep-learning ais are so easy to fool. *Nature*, 574:163–166, 10 2019.

[32] Walter J Scheirer, Anderson Rocha, Ross J Micheals, and Terrance E Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1689–1695, 2011.

[33] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. SELFIE: Refurbishing unclean samples for robust deep learning. In *ICML*, 2019.

[34] Daniel Kang, Yi Sun, Dan Hendrycks, Tom Brown, and Jacob Steinhardt. Testing robustness against unforeseen adversaries, 2019.

[35] R. Tous, O. Wust, M. Gomez, J. Poveda, M. Elena, J. Torres, M. Makni, and E. Ayguadé. User-generated content curation with deep convolutional neural networks. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 2535–2540, 2016.

[36] Wenjie Ye, Yue Dong, and Pieter Peers. Interactive curation of datasets for training and refining generative models. *Computer Graphics Forum*, 38(7):369–380, 2019.

8

The limits of detection

In this chapter, we analyze the limits of CFS for nanoparticle detection. We will demonstrate that in standard optical laboratory conditions it is possible to detect a low-index particle as small as $\lambda/14$ in diameter.

After a short introduction, in Section 8.2, we present the experimental detection results of the fabricated samples containing nanoparticles made of resist. Further, in Sections 8.3 - 8.6, we analyze background surfaces with AFM and CFS; numerically estimate the roughness contribution to the signal; analyze the electronic noise from the detector circuit and the influence of environmental vibration. In Section 8.7 we analyze the limits of detection and in Section 8.8 we present results of the pellicle sample. In Section 8.9, we present the conclusions.

Parts of this chapter have been published in Optics Express 29.11 (2021): 16487-16505 [1]

8.1. Introduction

In previous chapters, the samples were made of spin-coated particles. However, to investigate smaller particles, there are limitations to this sample-fabrication approach. An important limitation is the clustering of particles when they are < 40 nm in diameter. Namely, the mean distance between isolated particles is less than $1\ \mu$m [2, 3]. Further, the full traceability of the results for such samples is difficult, because particles are randomly distributed on the surface. In order to be sure about the detection and localization of all individual particles, one would like to compare the scanned area with the corresponding image obtained by a high-resolution imaging technique, such as SEM.

To address the issue of producing low-index < 40 nm particles with mean separation of $\geq 1\ \mu$m, we fabricated resist-made nanoparticles using electron-beam lithography (EBL). We combine the analysis of such samples with > 40 nm particles made in spin-coating. After reaching the detection of $\lambda/14$ particle for $\lambda = 405$ nm at $SNR \approx 4$ dB, we present a thorough analysis of the CFS scatterometer. We show a study of the substrate roughness for both the spin-coated samples and the roughness of chemically developed silicon surface from the EBL sample. Furthermore, we analyze the electronic noise emerging from the circuit of the differential detector, and investigate the vibration sources that distort the measured data. Finally, we fit the experimentally acquired signals for the studied particle sizes and extrapolate towards the detector noise level to indicate that the limit of the CFS setup was reached.

8.2. Experimental results on the minimum size limit of particle detection

As discussed above, in order to be sure about the detection and localization of all individual particles, one would like to compare the scanned area with the corresponding image obtained by a high-resolution imaging technique, such as scanning electron microscope (SEM). To address the issue of producing low-index $< \lambda/10$ particles with a mean separation of $\geq 1\ \mu$m, we have fabricated resist nanoparticles on silicon wafers using EBL. We used negative resist samples made with EBL because it allows for the precise control of the particle size as well as the separation between the isolated nanoparticles. We write the enclosed reference structure and the particle array of tiny pillars with two spot sizes to improve upon the writing speed. After the surface scan of the reference structure with CFS is performed, the sample is analyzed with a SEM, to provide the comparison. The array of 4×4 resist particles on top the silicon wafer is shown in SEM image (Figure 8.1). The size of the manufactured array is a compromise between the time it takes to fabricate/analyze the sample and the number of available target-size particles. Depending on the magnification of the SEM, it is easy to locate the rectangular edge structure Figure 8.1 A) or the star marker C). Performing ellipsometric measurements before developing the resist, we measured the height of the resist giving as result $h_{resist} = 25$ nm.

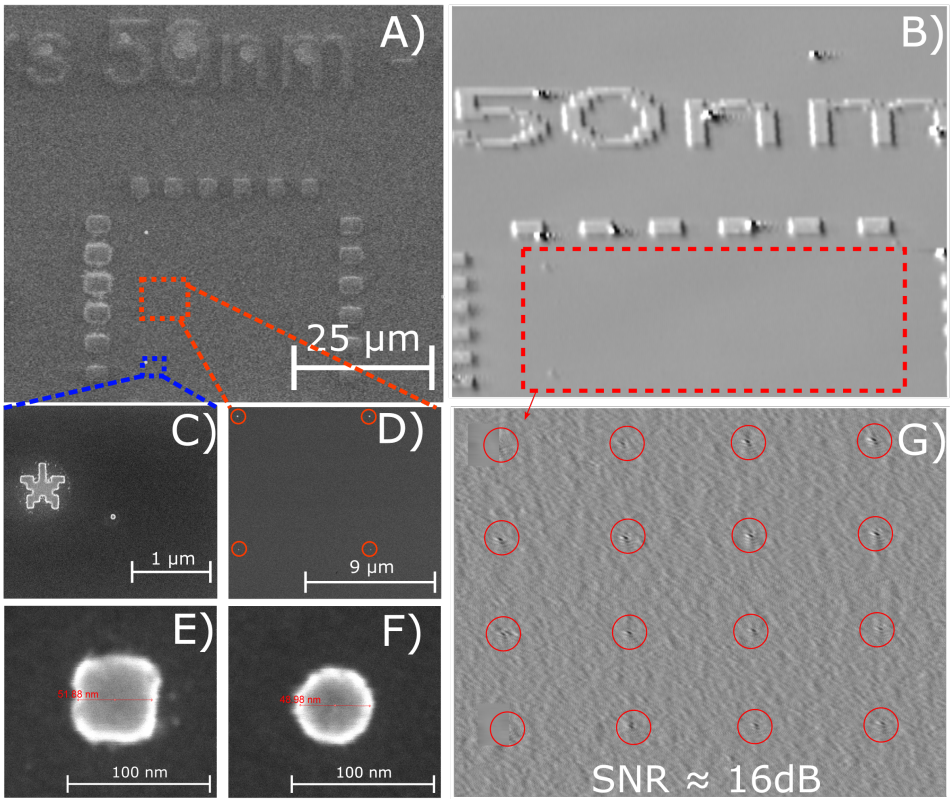


Figure 8.1: A) Scanning electron microscope (SEM) image of the reference structure. B) The CFS scattered map of the reference region with the close-up G) of all 16 particles that have been detected in a single scan of $34.6 \times 37\ \mu m$. C) An SEM image of the star (marker) of ≈ 700 nm next to the 50 nm cylinder. The nanoparticles separated by 9 μm with the square prisms and cylinders in the upper and bottom rows D). SEM images of the 50 nm square and cylinder E) and F), respectively.

Figures 8.1 E) and F) show the SEM images of the isolated resist nanoparticles with either square prism or cylindrical shapes. The Latex Sphere Equivalent (LSE) of these particles are 46 and 42.5 nm for nominal 50 nm square prisms and cylinders, respectively. The resulting differential signal maps obtained with CFS show that all particles within the reference region have been detected (Figure 8.1 B and G). The $SNR \approx 16$ dB is sufficiently high to localize all 16 particles. The noise level is computed based on a differential signal of no particle region scanning.

In order to check the ability of the CFS system to detect smaller particles, we fabricated another sample containing guiding lines of $4 \times 0.5\ \mu m$ and 25 nm resist cubes. The lines made of resist facilitate the localization of the nanoparticles both with SEM and CFS.

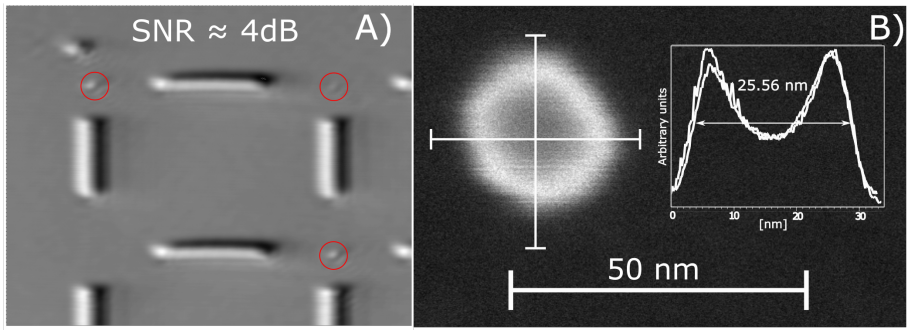


Figure 8.2: A) CFS scattered map of the reference region where the three cube nanoparticles are visible ($17.2 \times 17.2\ \mu m$). B) Scanning electron microscope (SEM) image of the ≈ 25 nm cube nanoparticle. In the inset, we see the cross-sections of the particle image in two orthogonal directions

Figure 8.2 A) displays the measurement with CFS where the three cube nanoparticles have been detected. The exact size of these nanoparticles is confirmed with the SEM image shown in Figure 8.2 B). With this measurement, we validate the detection of nanoparticles of diameter $LSE \approx 29$ nm (resist cube of $25 \times 25 \times 25$ nm). For an incident power level of $P \approx 20\ \mu$W, a $SNR \approx 4$ dB has been obtained. We remind that the wavelength used in these measurements was $\lambda = 405$ nm.

8.3. AFM for background surface and no particle CFS scans

The summary of three studied surfaces is shown in Table 8.1.

Table 8.1: The summary of AFM Park NX20 scan for three types of surfaces: blank silicon wafer and no particle regions of spin-coated and developed EBL silicon wafer.

Silicon surface type	Combined measurement area [μm^2]	h_{rms} [nm]
blank	2.43x2.43	0.4959 ± 0.0016
spin-coated	1.8x1.8	0.6920 ± 0.0006
developed EBL	11x11	0.2524 ± 0.0002

For each sample we rely on the region where the target particles are absent. Further, looking at the three types of processed silicon surfaces (Figure 8.3) indicates that none of the samples is completely clean.

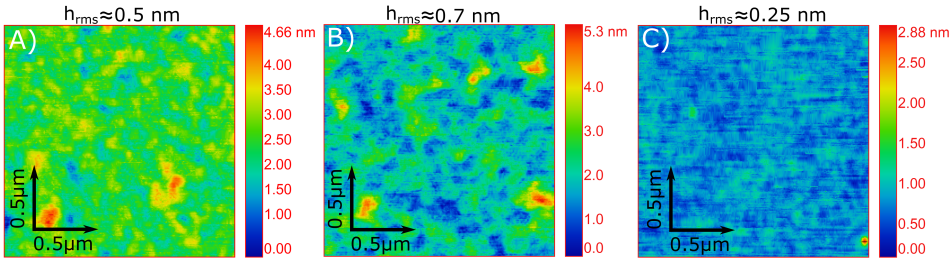


Figure 8.3: A) The AFM Park NX20 scan of a random 0.5×0.5 μm region from three types of samples. The $h_{rms} \approx 0.5$ nm of blank silicon wafer, B) the $h_{rms} \approx 0.7$ nm of no particle region from spin-coated sample, and the C) $h_{rms} \approx 0.25$ nm of no particle region from EBL fabricated sample.

The bumps or defects in the size range from 12 nm to 80 nm are present. With the smallest $h_{rms} \approx 0.25$ nm demonstrated for the sample made with EBL (Figure 8.3 C) we approach the surface roughness of the high-end blank wafers used in industry (≈ 0.1 nm) [4]. The possible reason why this sample contains the unwanted particles of ≈ 12 nm transverse dimension, bottom right in Figure 8.3 C), is the remainder of resist after the development stage. It is also clear that after the blank wafer (Figure 8.3 A) is spin-coated (Figure 8.3 B) there is contamination due to the dilution with DI water and the solution in which the PSL is contained. Hence one gets the increased from $h_{rms} \approx 0.5$ nm to $h_{rms} \approx 0.7$ nm levels of roughness and the amount of unwanted particles.

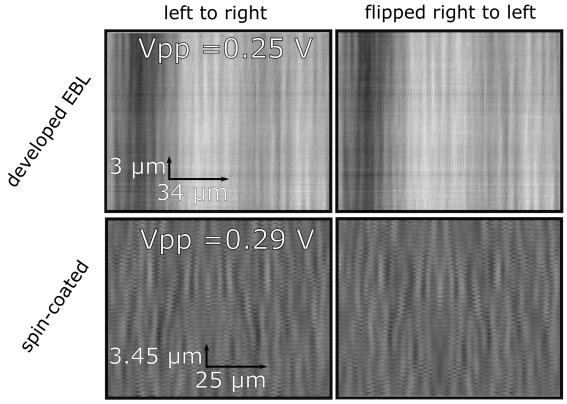


Figure 8.4: The scattered maps of no-particle regions. The same region is revisited in left to right and right to left scanning. Developed EBL and spin-coated sample in upper and lower panel respectively.

Further, we analyze "clean" surface regions of no target particle for samples of Figure 8.3 C) and B) type. For the incident power level $P \approx 20$ μW we look at the

region of 34×3 μm for developed and the region of 25×3.45 μm for spin-coated sample (Figure 8.4 upper and lower panel respectively).

It is evident that the surface roughness difference for two types of samples $h_{rms} \approx 0.25$ nm and $h_{rms} \approx 0.7$ nm does not render a substantial change to the signal level $V_{pp} = 0.25$ V and $V_{pp} = 0.29$ V respectively. Here, we performed the scanning slowly (1 second per 25 micrometers of scan). The important observation is the repeated shape of the signal surface upon the scanning in two opposite directions upper or lower panel of Figure 8.4. This observation is also valid when the stepping mode in scanning is performed over another region of spin-coated sample Figure 8.5.

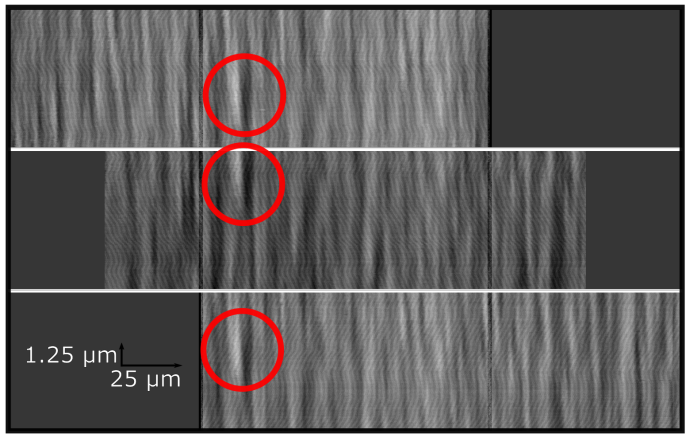


Figure 8.5: The scattered maps of no-particle regions from spin-coated sample the three left to right scans each 25×1.25 μm^2 are performed in stepping mode.

In Figure 8.5 the top scan has starting coordinates of $X = 745$, $Y = 750$, middle scan has starting point $X = 750$, $Y = 749.5$, and bottom scan has staring coordinates $X = 755$, $Y = 750$. With the red circle of Figure 8.5 we mark the defect on a studied region. One can see that there is a returning pattern, visible at the three scans, indicating that specific no-particle structures are measured at the scanning epoch. A possible explanation for this is the micro-scale elongated structures of the polished silicon, as an example demonstrated in [5], systematic error in the system or the speckle effects.

8.4. Effects of the roughness of the surface

The differential detection mode is sensitive to irregularities in reflection from the surface of the substrate. These irregularities can come in the form of, e.g. surface roughness, contamination or defects around the particles. For the same optical properties of surface and particle (height of the surface roughness reaches the diameter of the particle $h_{RMS} \approx d$) the signal from the particle will be in the order of the signal from the surface. For the case of particles or contamination with low refractive index on a high refractive index silicon substrate, the reflectance can be monitored. Compared to the perfectly flat interface, the reflectance will drop if multiple particles or roughness

are present.

In order to understand the influence of roughness, we have performed some rigorous simulations. Within the FEM solver [6], two-dimensional rough surfaces can have variations along the lateral directions. The rough surface height profile is based on a centralized Gaussian distribution in space given by:

$$g(x,y) = e^{-\frac{1}{2}(\frac{x}{c_x})^2 - \frac{1}{2}(\frac{y}{c_y})^2}, \tag{8.1}$$

which is randomly displaced while summed and normalized. This operation can be written in terms of a 2D forward $\mathcal{F}$ and inverse $\mathcal{F}^{-1}$ Fourier transform:

$$h(x,y) = C\mathcal{F}^{-1}\left[\mathcal{F}[g(x,y)]e^{i\phi(k_x,k_y)}\right], \tag{8.2}$$

where the scaling factor C is determined such that the

$$h_{RMS}^2 = \lim_{l\to\infty} \frac{\int_{-l}^{l}\int_{-l}^{l} h^2(x,y)dxdy}{\int_{-l}^{l}\int_{-l}^{l} 1dxdy}. \tag{8.3}$$

The displacement is according to the deterministic pseudo-random number generator that affects the phase function $\phi(k_x, k_y)$. The resulting roughness of the scattering structure is statistical in nature; thus, the average results of the simulations should be considered.

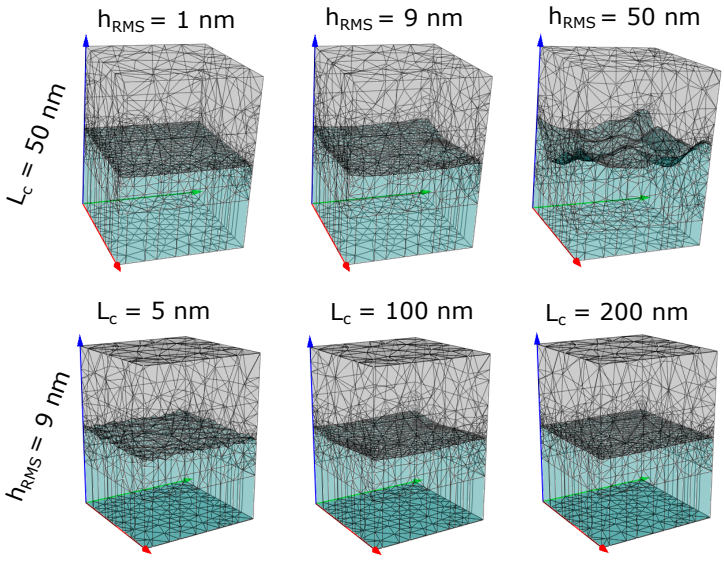


Figure 8.6: 3D FEM simulation volume with dimensions of $600 \times 600 \times 800$ nm. The interface between two media is defined as the rough surface with two parameters: height h_{RMS} and correlation length L_c. Fixed $L_c = 50$ nm with varying h_{RMS} in the upper panel, and fixed $h_{RMS} = 9$ nm with varying L_c in the bottom panel.

The effect of varying one of the surface parameters $L_c = [c_x, c_y]$ (correlation length) with the other being fixed h_{RMS}, and vice-versa is displayed in Figure 8.6. The variation of h_{RMS} can be intuitively understood, while the surface correlation describes the statistical independence of two points on the surface and increases with the correlation between two neighboring points. For a smooth surface $L_c = \infty$.

Further, periodic boundary conditions are necessary for the horizontal directions to account for the scattering out of the sides of the unit cell. Another reason is to mitigate the drawbacks related to the finite surface size such as the reduction in the angular resolution of scattered fields, as well as the potential influence of scattering and diffraction from the surface edges [7].

We estimate the influence of h_{RMS} on the reflection from the silicon surface assuming an refractive index of $\underline{n_{Si}} = 5.43 + i * 0.34$ at 405 nm. The rough silicon is normally illuminated by the plane wave with s and p polarization. For periodic cells, as one in Figure 8.6, the Fourier transform post-process yields the discrete diffraction modes (amplitudes of the reflected diffraction orders). The amplitudes of the electric plane wave $E(k)\exp(ikx)$ and the magnetic plane wave $H(k)\exp(ikx)$ can be converted to the the power flux density according to $P = \frac{1}{2}\mathrm{E} \times \mathrm{H}^* = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}||\mathrm{E}||^2\mathrm{k}^*/||\mathrm{k}||$. Further, the power fluxes in discrete directions of the reflected fields P_r can be divided by the power fluxes of the incoming plane wave P_i to obtain the reflectance $R_s \equiv \frac{P_r^s}{P_i^s}$ and $R_p \equiv \frac{P_r^p}{P_i^p}$. The total reflectance is obtained from the average of both polarizations $R = \frac{R_s + R_p}{2}$.

8

The real lattice vectors defined for our cell with period $\Lambda = \Lambda_x = \Lambda_y = 600$ nm are $\mathrm{a}_1 = [\Lambda, 0, 0]$ and $\mathrm{a}_2 = [0, \Lambda, 0]$. The corresponding reciprocal vectors b_1, b_2 are defined such that $b_i \cdot a_j = 2\pi\delta_{ij}$. Hence the reciprocal vectors are $\mathrm{b}_1 = [\frac{2\pi}{\Lambda}, 0, 0]$ and $\mathrm{b}_2 = [0, \frac{2\pi}{\Lambda}, 0]$, such that $\mathrm{a}_1 \cdot \mathrm{b}_2 = 0$ and $\mathrm{a}_1 \cdot \mathrm{b}_1 = 2\pi$. The Fourier modes available for the observation are linked to the reciprocal grid in k-space, where the transversal components of the k-vector in the pupil are defined as $\mathrm{k}_{\perp,n_1,n_2} = n_1\mathrm{b}_{\perp,1} + n_2\mathrm{b}_{\perp,2}$. The remaining normal components k_z are determined by $k_z = \sqrt{k_m^2 - |\mathrm{k}_\perp|}$, where the wave number in the material is given by $k_m = k_0 n_{Si}$ and free-space wavenumber $k_0 = 2\pi/\lambda$. To compute the total number of orders, one needs to determine all the integers n_1, n_2 such that the corresponding k-vectors are propagating $Re\{k_z\} > 0$. Alternatively, if the pupil plane is limited by the NA, the following criteria should be satisfied: $Re\{|\mathrm{k}_\perp|/k_0\} \leq NA$. It is evident that the only five modes will propagate within the NA (see Figure 8.7 A), such that the following combinations of the n_1, n_2 are possible $\mathrm{n}_1 = [-1, 0, 0, 0, 1]$ and $\mathrm{n}_2 = [0, -1, 0, 1, 0]$. We are not interested in reflectance in the normal direction $R_{normal} = R_{0,0}$ with $n_1 = n_2 = 0$. For the remaining diffraction into the higher orders $R_{scat} = R_{-1,0} + R_{0,-1} + R_{0,1} + R_{1,0}$ with the good alignment of the differential detector, one could still expect to get a zero signal.

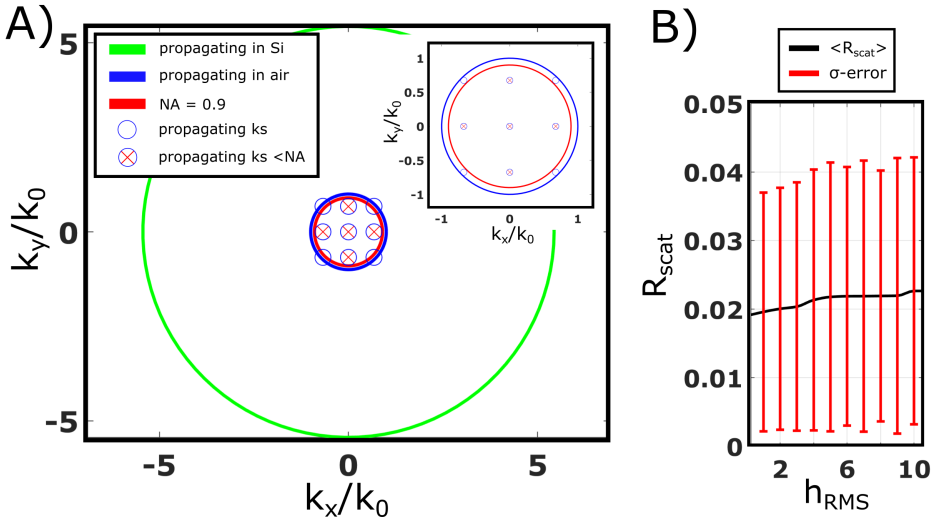


Figure 8.7: A) The Fourier space showing the propagating wave vectors for the air/silicon interface at $\lambda = 405$ nm and periodic cell of $\Lambda = 600$ nm. B) The average reflectance coefficient for the non-normal direction within the pupil.

Thus, in our worst-case scenario estimate, small roughness scattering is coming fully to one half of detector, we will rely on $R_{scat} = R_{-1,0}$. For the rough surface of $L_c = [50, 50]$ (x and y direction), the parametric sweep of the h_{RMS} from 0.25 to 10 nm, $NA = 0.9$, and $\lambda = 405$ nm as the average result of 400 initializations is shown in Figure 8.7 B). When estimating the peak-to-peak amplitude emerging from the surface roughness, we will rely on the reflectance coefficients of Figure 8.7B) combined with average roughness height from the AFM.

With the growing height of the roughness, the reflectance into the upper half space orthogonal to the surface R_0 decreases. At the same time, the reflectance into the higher angles R_{sca} increases. It is fair to assume that the scattering into the higher angles will define the amplitude of the differential signal. For our optical setup, the power at the detector P_{det} can be estimated as:

$$P_{det} = P_f R_{scat} T_{opt}, \tag{8.4}$$

where P_f is the power incident on the surface and T_{opt} is the transmission through beamsplitter and other optics; $T_{opt} \approx 0.28$ in our case. Further, taking into account the properties of the detector, the output amplitude is equal to:

$$V_{out} = P_{det} S G \tag{8.5}$$

where S is the responsivity of the detector and G is the gain of the differential detector circuit.

8.5. Accessing the noise level of the detector

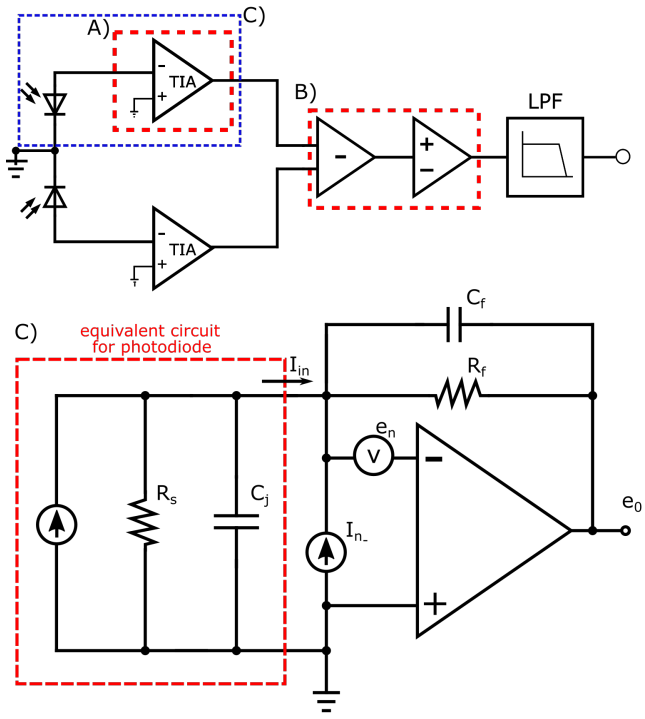


Figure 8.8: A) Sketch of the differential detector circuit. The current from each photodiode is converted to voltage by the transimpedance amplifier with a gain $G_{TIA} \approx 32K$. B) The voltages of each TIA are subtracted and amplified with a combined gain of $G_{OPA} \approx 513$. The bandwidth of the circuit is $BW = 50\ kHz$. C) Sketch for modeling of the noise of a single photodiode and TIA circuit. $R_s = 300\ M\Omega$, $C_j = 30\ pF$, $C_f = 33\ pF$, $R_f = 100\ K\Omega$.

The circuit of the differential detection (Figure 8.8) is optimized for small-signal amplification. The split detector $(ODD - 3W - 2)$ has a common ground that minimizes the pickup of the environmental noise. The photocurrent measured from each diode is converted to a voltage with a trans-impedance amplifier (TIA) with a gain $G_{TIA} \approx 32K$ (Figure 8.8 A). It is advantageous to amplify each photodiode signal separately by having two TIAs instead of subtracting the photodiodes currents, amplifying and converting to voltage. The latter is more susceptible to pickup noise. Further, the subtraction of the two signals and adding additional gain is accomplished by two operational amplifiers. The total gain of this stage is $G_{OPA} \approx 513$ Figure 8.8 B). Accordingly, the complete gain of the circuit is $G \approx 16 \cdot 10^6$. The bandwidth of the circuit is set to $BW = 50$ kHz. Since we perform particle detection in the time domain, it is convenient to express the noise in terms of either peak-to-peak (pp) or RMS values. The figure of merit of V_{rms} should be multiplied by 6.6 to get to the estimate of the peak-to-peak value. The parameters of the SD photodiode are summarized in Table 8.2.

Table 8.2: Parameters of the split detector(s): ODD-3W-2 Bi-Cell Silicon Photodiode

Active area	$2 \times 3.1\ mm^2$
Dark Current	0.9 nA
Responsivity	0.36 (633nm) and 0.1 (405nm)
Noise Equivalent Power	$2.5 \times 10^{-14}\ W/\sqrt{Hz}$
Response time	≈ 190 ns

We start by computing the shot I_s and Johnson noise I_j of the photodiode:

$$I_s = \sqrt{2q(I_p + I_d)BW}, \tag{8.6}$$

where q is the electron charge, I_p the photocurrent, I_d the dark photocurrent, and BW the bandwidth.

$$I_j = \sqrt{\frac{4k_B TBW}{R_s}}, \tag{8.7}$$

where k_B is the Boltzmann constant, T the temperature in Kelvin (estimated for temperature $25°$), R_s is the Shunt resistor of photodiode.

The total noise at the photodiode is given by:

$$I_{tp} = \sqrt{I_s^2 + I_j^2}, \tag{8.8}$$

and the associated voltage noise at the output of TIA can be written as:

$$E_{np} = I_{tp} G_{TIA}. \tag{8.9}$$

It is convention to compute the RMS noise of the field-effect transistor (FET) TIA (Figure 8.2 C) using a piece-wise approach [8]. The voltage noise is computed for the low, medium and high frequency region, and the coefficients for the voltage noise density $K_1 = 7$, $K_2 = 3.3$ and $K_3 = 2.1\ nV/\sqrt{Hz}$ are taken from the data sheet for the TIA (ADA4625-1). In the first region, from $f_1 = 0.01$ Hz to $f_c = 100$ Hz

$$E_{n1} = K_1\left[1 + \frac{R_f}{R_s}\right]\sqrt{\ln\left(\frac{f_c}{f_1}\right)}, \tag{8.10}$$

in the second region, from f_c to $f_a = 1$ kHz

$$E_{n2} = K_2 K_3 \sqrt{\frac{f_c^3}{3} - \frac{f_a^3}{3}}, \tag{8.11}$$

and the third region from f_a to $f_3 = 50$ kHz

$$E_{n3} = K_2\left(1 + \frac{C_j}{C_f}\sqrt{\left(\frac{\pi}{2}\right)f_3 - f_a}\right). \tag{8.12}$$

The output voltage component due to the current noise, with the corresponding coefficient for the current noise density $K_4 = 4.5\ fA/\sqrt{Hz}$ is equal to:

$$E_{ni} = K_4\left[1 + \frac{R_f}{R_s}\right]\sqrt{f_3 - f_1}. \tag{8.13}$$

Next, the contribution of the resistor noise of the TIA circuit is given by:

$$E_{nR} = \sqrt{4k_B T R_f BW}, \tag{8.14}$$

The total noise at the output of TIA is given by:

$$E_{ntotal} = \sqrt{E_{np}^2 + E_{n1}^2 + E_{n2}^2 + E_{n3}^2 + E_{ni}^2 + E_{nR}^2}. \tag{8.15}$$

Finally, after the differentiation of two diodes and the remaining gain of the circuit the RMS noise can be estimated for the complete circuit

$$E_{nout} = G_{OPA}\sqrt{2(E_{ntotal})^2}. \tag{8.16}$$

In Table 8.3 we summarize the different noise contributions.

Table 8.3: The noise budget of the detector circuit in μV_{RMS}

Noise Contribution	E_{nR}	E_{n3}	E_{np}	E_{n1}	E_{n2}	E_{ni}
μV_{RMS}	11.37	1.75	0.133	0.02	0.002	2.25e-4

The analytical estimate for the total output noise with no incident light at the split-detector ($I_p = 0$) is $E_{nout}^{ab} = 8.34\ mV_{RMS}$ which roughly agrees with the experimental values $E_{out}^{eb} = 7.8\ mV_{RMS}$, measured with a RMS voltmeter R&S URE3. Further, the estimate of the noise with the detector unblocked and no laser light is $E_{nout}^{e} = 14\ mV_{RMS}$. The increased level is due to ambient noise that is not considered in the analytical model.

8.6. Noise originating from vibrations

The scanning piezo stage used in the setup (P-629.2CD by Physik Instrumente) is mounted on a optical table, and not directly attached to the vertical breadboard containing the optical CFS elements. In closed-loop operation, a built-in sensor is used to monitor the position of the stage in real time. The error signal is sent to the controller to provide accurate nanopositioning. External vibrations can affect the piezo position in both horizontal and vertical directions. Based on the steady-state observation of

the feedback signal, one can identify the spectral characteristics of the vibrations of the setup (Figure 8.9).

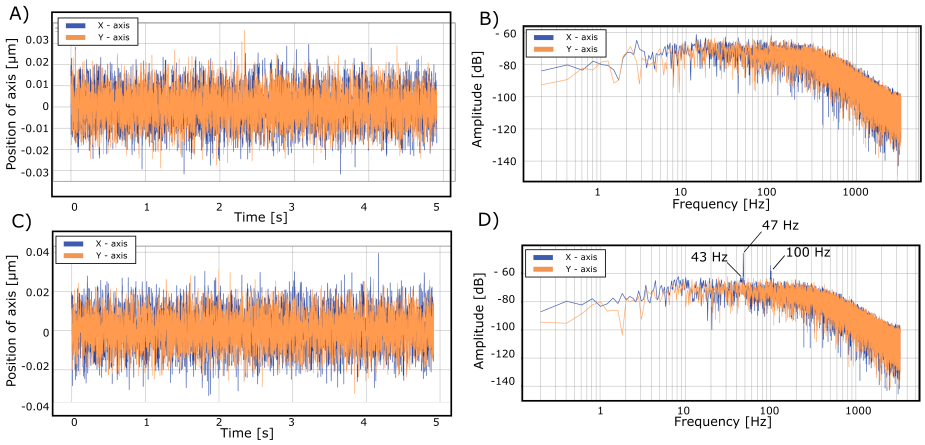


Figure 8.9: The time-domain (left) and corresponding frequency domain (right) mechanical noise of the piezo stage. A) and B) the current state of the piezo noise. C) and D) an example of how the controller of the piezo stage compensates for the multiple sources of the vibration: 43, 47 and 100 Hz.

From the time-domain signal of the x and y - axis position (Figure 8.9 A) there is up to 25 nm displacement from the target zero position. The residual noise Figure 8.9 A) and B), is caused by the electronic itself due to the high-frequency components that the piezo actuator is not able to convert into motion. The effect of this noise is a horizontal mismatch between the peaks of the signal in consecutive scans that is estimated to be 16 nm (right image of Figure 8.10 A). This could be considered as the piezo-induced limit for the particle positioning accuracy. As a comparison, in the x and y - axis position (Figure 8.9 C) there is up to 40 nm displacement from the target zero position. The cause is understood from the corresponding frequency domain representation (Figure 8.9 D) where there are spectral contributions of 43 Hz, 47 Hz, and 100 Hz. After investigation, the 43 Hz was attributed to vibrations induced by the lock-in amplifier that was situated on the same optical table. Further, one of the air pressured legs from the optical table was incorrectly placed such that there was a mechanical contact between one of the legs and the table. This allowed building vibrations to pass around to the table. After removing the lock-in amplifier from the optical table and readjusting the table leg, the 43 and 47 Hz peaks disappeared (Figure 8.9 B).

The influence of the investigated vibration to the measured data is summarized in Figure 8.10.

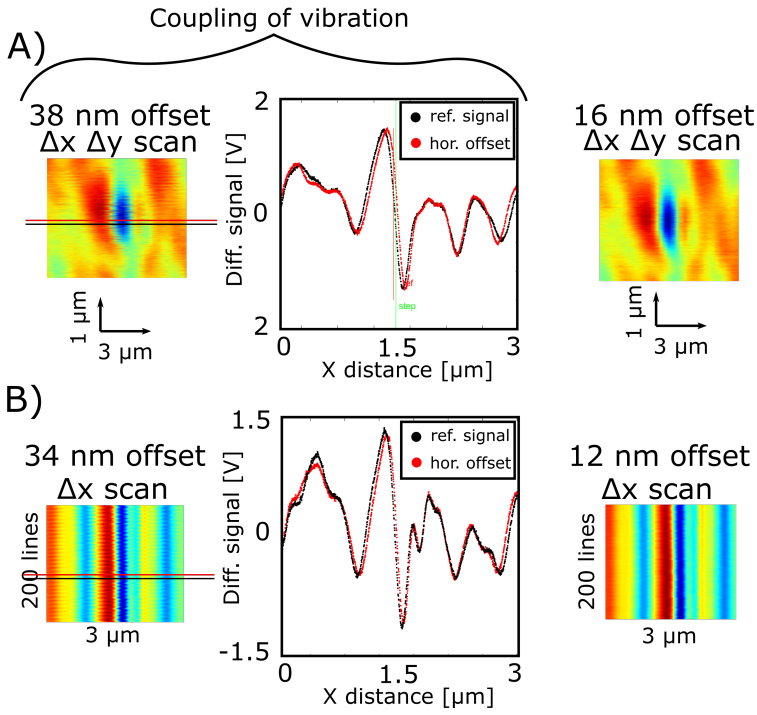


Figure 8.10: Vibrational sources couple to the measurement in the form of an horizontal offset for the particle signal (left). The reference and maximum offset signals are shown in black and red color (middle), and the remaining mechanical noise of the stage (right). A) The raster scanning of the particle in Δx and Δy in and B) repeated scanning only in Δx. The scanned particle is a 60 nm PSL sphere.

8

The 2D, Δx and Δy, (Figure 8.10 A) and 1D repeated scans of Δx, (Figure 8.10 B) for an isolated particle (in this case a PSL sphere of diamater of 60 nm) are compared between the situation before after removing the above mentioned vibration sources. Following the data processing workflow, the sampled data points are arranged line per line, according to the raster scan pattern to provide the scan map. With the initial situation, the controller of the piezo will over- or undershoot the target position thus introducing a mismatch in X and Y between scan lines. With both 43 and 47 Hz sources that corrupt the measurement process, there is an apparent ripple on the intensity of the differential signal (left part of Figure 8.10). By slicing the scan map in two places (black and red lines in Figure 8.10 left), the maximum offset between the zero-crossings of particle signal can be estimated as 38 nm for 2D and 34 nm for 1D case. This indicates the minor influence of the vibration in y-direction, with the dominant error in x-direction.

Apart from the horizontal displacement of the piezo table due to vibrations, the latter could also affect the z-position of the sample with respect to the focusing objective. The effect of a defocus on the simulated peak of LR profile is demonstrated

in Figure 8.11

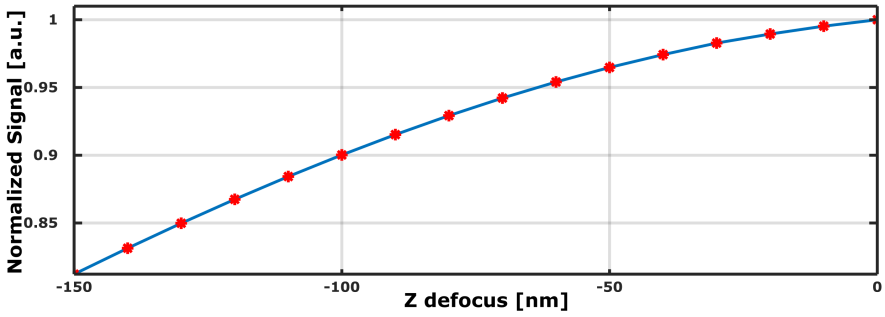


Figure 8.11: Normalized simulated defocus signal. The particle offset position $X = -200$nm combined with a z-defocus.

We can give an estimate that on average the V_{pp} drops by $c_z = 0.13\%$ per one nanometer of defocus. Further, the possible z-displacement can be computed as $\Delta z = \sigma(V_{pp})/(c_z \overline{V}_{pp})$. For the observed 60 nm PSL on silicon sample signal, amplitude fluctuation renders the defocus of $\Delta z' = 25.3$ nm which has been reduced to $\Delta z = 16.15$ nm when the remaining fluctuation due to vibration sources were removed.

8.7. Limit of detection

In order to establish the smallest LSE diameter particle that can be detected by means of the developed CFS system, we measured two types of samples. The first type are PSL nanospheres that have been spin-coated on 1-inch wafers with a mean separation of ≥ 1 μm between the isolated particles, and the second type are negative-resist square prisms or pillars made with the EBL. A summary of the fabrication and types of samples is presented in Appendix C.

For each sample, we performed at least four runs (different measurement days) accessing more than three areas on the sample (in case of PSL samples), and analyzed the signal in the time domain with the aid of the search approach, outlined in reference [9], resulting in individual amplitude labeling of hundreds of particles. Also, we study areas without particles to provide the background level estimates for the two types of samples. Figure 8.12 summarizes the overall analysis. For the detected particles sizes, the blue diamonds indicate the average value of the V_{pp} with a standard deviation shown in gray error bar. Error emerges mostly from the non-uniformity of particle sizes and vibrations in the setup. In case of the negative-resist particles the sigma is smaller due to limited amount of fabricated structures.

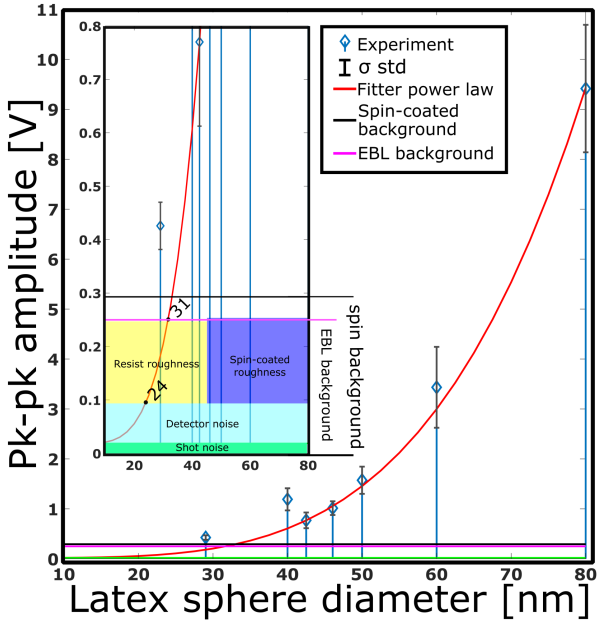


Figure 8.12: The peak-to-peak amplitude of the signal (blue diamonds with error bar of σ), fitted power-law (red curve) as function of the latex sphere diameter. Inset shows experimental background level of the spin-coated particles samples (black line) and samples with particles fabricated with the EBL (magenta line). The background level is further split into the detector noise in light cyan, shot noise in green, and worst-case scenario roughness signal from the EBL-made and spin-coated sample in yellow and purple respectively.

8

The solid red line in the plot is the fourth-power law curve with the best fit based on the error sum of squares criterion. The correction is done by normalizing to unity and multiplying with the maximum value from the experimentally acquired peak-to-peak amplitudes and adding the offset of the shot noise level U_s.

$$U' = \max U_{exp} \frac{U}{|U|} + U_s \tag{8.17}$$

Furthermore, we include the experimental background levels of the surface signal with no particles for spin-coated and EBL fabricated structures on silicon substrates, represented in the figure by black and magenta horizontal lines, respectively. We color the amplitude region of the complete detector noise with cyan, shot noise part in green. On top, the estimated pp signals from the surface roughness of 0.25 nm (in yellow) and of 0.7 nm (in purple) attributed to the developed surface of the EBL sample and spin-residue surface of PSL sample respectively. The crossing between the fitted power curve and the background measured for the EBL sample indicates the smallest detected particle that comes close to experimental one $d_{predEBL} = 31$ nm. Assuming a perfectly flat surface, the intersection between the fitted power law curve and present

detector noise is at $d_{pred} = 24$ nm particle. Note that the fact that we obtain a larger pk to pk amplitude than the one expected by the power law suggests that the latter may not be the only contribution to the detected signal, since the combined signals coming from the light scattered by the particle and the one coming from the reflection by the surface may constructively interfere, leading to a stronger signal. In addition, other high order scattering effects and resonances may also occur. This leads to the conclusion that with our present system, even smaller particles could be detected.

8.8. Analysis of pellicle sample

Samples of polysilicon-based EUV pellicle with the thickness of the free-standing film about $400 - 500$ nm thick were analyzed. The pellicle has 24×24 mm (outer dimensions) and 10×10 mm (inner dimensions) and the total thickness of the sample is about 700 μm. The sample is contaminated with unknown material. The Figure 8.13 shows a digital microscope image of the pellicle's inner part scan, where red boxes address specific regions analyzed with the CFS prototype, where the sword-shaped piece of the pellicle film is used as a reference. The image in Fig. 8.13 is obtained with a digital microscope (VHX KEYENCE, model VHX-6000), with magnification factor of 300, and the scale is marked on the figure.

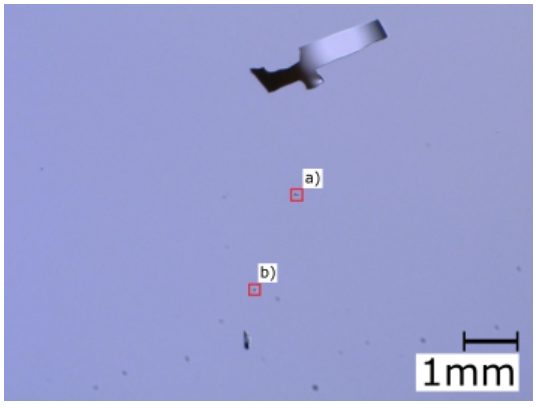


Figure 8.13: Image of the reference region of the pellicle's inner part, and the two reference a) and b) positions where the scans were performed as indicated with red boxes.

The measured time-signals in the scan area of total 338×15 μm in the reference sub-area (marked as "a" in Figure 8.13) and a total of 25×500 μm in the reference sub-area "b" are stored together and analyzed to produce the histogram of particle counts (Figure 8.14 left). The x-axis represents the time span of the particle's signal that is centered about a position of 8.4 ms pulse. In order to estimate the particle size distribution, we use the calibration curve, see Appendix C. Based on the time span of particles detected on the pellicle structure we conclude that the size of the particles is in the range of $60 - 80$ nm in diameter.

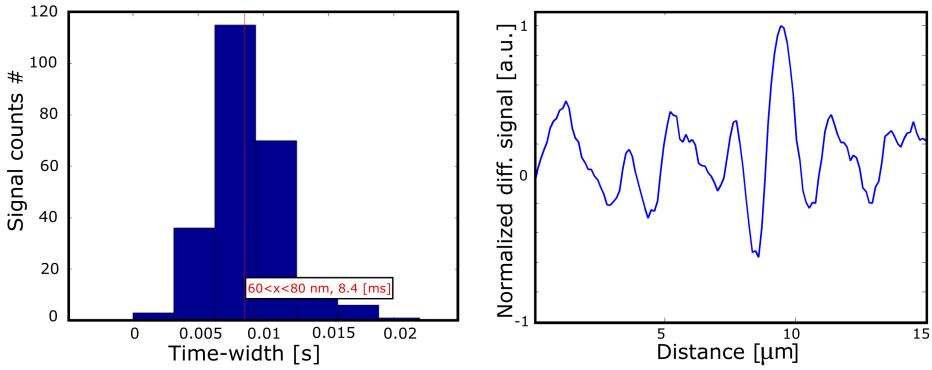


Figure 8.14: The size distribution of the particle scatterers based on pulse-width of the detected particles from the pellicle sample. Time width of the measured pulse in seconds (x-axis) - left. The typical signal as detected from the reference region a). The SNR is 10.21dB and the pulse width is 9.2 ms - right.

The one-line scan profile over the single detected particle at the reference area "a" suggests that sample is rough (Figure 8.14 right). Compared to another sample we have (silicon wafer sample covered with 60 and 80 PSL spheres), the SNR achieved for the pellicle sample is lower. Nonetheless, all the collected signals for histogram above were thresholded to have at least 10dB SNR.

8.9. Conclusions

We conclude that the detection limits of our system are dictated by multiple factors. Firstly, the presence of inherent mechanical noise of the piezo stage (Section 8.6). With the current model of the piezo stage, the uncertainty for particle localization in our setup is $\Delta x \approx 16$ nm. Further, the peak-to-peak fluctuation of the signal indicates the vibration-induced uncertainty of the focal position that is estimated to be $\Delta z \approx 16$ nm. Accordingly, the design of the setup could be improved to reduce vibrations that couple to the system. For systems requiring faster 2D scan, beam steering combined with 1D piezo scan could be an interesting option.

Next, the investigated electronic noise of the detector is bigger than the one with the Poisson distribution statistics. Based on the analysis of the differential detector circuit (Section 8.5), we identify the total noise, including all the noise components using the root sum of the squares. The shunt resistance R_s, junction capacitance C_j of a photodiode, feedback resistance R_f and capacitance C_f of the TIA are the primary parameters used in the noise analysis. In our case, the TIA's feedback resistor noise is the dominant noise source. The theoretical estimate of the resulting noise matches the experimentally measured value for the $E^e_{nout} = 14\ mV_{RMS}$.

Finally, the small scattering cross-section from isolated nanoparticles contributes to the low SNR of the detector signal as the particle diameter decreases. In this work, we demonstrate detection of nanoparticles with m $d \approx \lambda/14$ LSE diameter, where the main noise sources were due to the combined effect of detector noise and silicon

surface roughness $h_{RMS} = 0.25$ nm. We verified the reliability of the CFS bright field scanner by comparing the number of detected scatterers inside the wafer reference region with that of a benchmark SEM measurement. The detection of $LSE \approx 29$ nm particle is achieved with $SNR \approx 4$ dB. This result complies with the requirements for surface inspection in the semiconductor industry (SNR needs to be greater than 3 dB) [10]. The demonstrated detection is promising for initial inspection of a wafer in a fabrication environment, testing a reticle blank, or during the production flow (the pellicle membrane) where non-destructive surface scanning is required.

References

[1] D Kolenov, IE Zadeh, RC Horsten, and SF Pereira. Direct detection of polystyrene equivalent nanoparticles with a diameter of 21 nm (λ/19) using coherent fourier scatterometry. *Optics Express*, 29(11):16487–16505, 2021.

[2] Sarika Phadke, Jerome B Stanley, Judith D Sorge, and Dunbar P Birnie III. Clustering effects in solution-based nanoparticle/template hybrid coatings. *Journal of the Society for Information Display*, 15(12):1089–1093, 2007.

[3] Simon Schiwek, Lars-Oliver Heim, Robert W Stark, and Christian Dietz. Manipulation of polystyrene nanoparticles on a silicon wafer in the peak force tapping mode in water: ph-dependent friction and adhesion force. *Journal of Applied Physics*, 117(10):104303, 2015.

[4] Jung-Hwan Kim, Seunghyun Moon, Ji-Woong Kim, Donggun Lee, Byong Chon Park, Dal-Hyun Kim, Yoojin Jeong, Sean Hand, Jason Osborne, Peter De Wolf, et al. Advanced measurement and diagnosis of the effect on the underlayer roughness for industrial standard metrology. *Scientific reports*, 9(1):1–8, 2019.

[5] Kheelraj Pandey and Pulak M Pandey. Chemically assisted polishing of monocrystalline silicon wafer si (100) by ddmaf. *Procedia engineering*, 184:178–184, 2017.

[6] Sven Burger, Lin Zschiedrich, Frank Schmidt, Roderick Köhle, Thomas Henkel, Bernd Küchler, and Christoph Nölscher. 3D simulations of electromagnetic fields in nanostructures using the time-harmonic finite-element method. In Harald Bosse, Bernd Bodermann, and Richard M. Silver, editors, *Modeling Aspects in Optical Metrology*, volume 6617, pages 255 – 265. International Society for Optics and Photonics, SPIE, 2007.

[7] Alexei A Maradudin. *Light scattering and nanoscale surface roughness*. Springer Science & Business Media, 2010.

[8] Burr Brown. Noise analysis of fet transimpedance amplifiers. *Texas Instrum. Dallas, TX, USA, Tech. Rep.[Online]. Available: http://www. ti. com/lit/an/sboa060/sboa060. pdf*, 1994.

[9] D. Kolenov and S. F. Pereira. Machine learning techniques applied for the detection of nanoparticles on surfaces using coherent fourier scatterometry. *Opt. Express*, 28(13):19163–19186, 06 2020.

[10] Baltzinger Jean-Luc and Delahaye Bruno. Contamination monitoring and analysis in semiconductor manufacturing. *Semicond. Technol*, pages 57–78, 1999.

9

Conclusion and Outlook

9.1. Conclusion

In the beginning of the thesis we gave an overview of scattering in the context of optical metrology, described the theory that is necessary to explain the interaction of the light with the particle on the surface and provided a comparison of CFS and other surface inspection techniques and tools. Further, in the main contribution of the thesis, we have investigated different limiting aspects of the particle detector based on CFS:

- noise in the system wherein we experimentally investigate the capability for noise suppression with a heterodyne detection system for CFS (Section 3.2);
- accurate focus positioning wherein we propose the novel method of the focusing error S-curve generation (Section 5.2);
- particle signal retrieval wherein we develop and test data processing workflow for the CFS particle detector (Section 6.5);
- particle signal classification wherein we examine the ability of the trained CNN model to refuse unknown inputs (Section 7.3.3);
- remaining noise in the system wherein we experimentally verify the smallest particle that can be detected with our system. (Section 8.7)

The CFS scheme allows the use of low power lasers (as compared to darkfield techniques) and is suitable for detection of low optical contrast subwavelength nanoparticles. In Chapter 4, we provided a discussion on applying CFS in the context of detecting biological particle-like specimens and experimental validation of bio-mimicking samples.

By improving the particle scanner tools, such as one presented in this thesis, we believe we are contributing to the improvement of the process of IC's manufacturing towards microchip size reduction, and with that a better technology for our lives.

9.2. Outlook 1. Detection below the background level

One important aspect that has been shown in this thesis is the conclusion that limit on the smallest particle size that can be detected is given by the roughness of the substrate surface. This limit comes from the fact that as the particle gets smaller, the amplitude level of the differential background signal is comparable to the signal from the particle. We propose here three approaches to retrieve the location of the particles that produce comparable scattering cross-section as the background roughness.

Firstly, unless the correlation length of the roughness reaches the size scale of the particle (transverse direction), it is envisioned that the frequency content of the particle and roughness can be separated. By performing Fourier transformation, applying bandpass or cascade filters that are set to the expected particle frequency, and inverse Fourier transform, the useful signal can be recovered. Secondly, the template particle signal (synthesized) can be convolved with the global scan. In this case, the entire "image" is toned down, except for the region of interest. One will get a strong response where the correct particle location is. Thirdly, relying on the feature space representation, rather than in pixel (spatial) space, the background and particle could be separated.

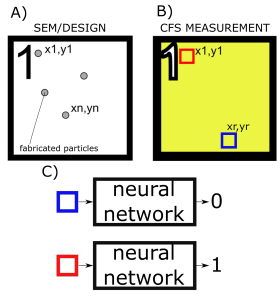


Figure 9.1: The schematic of the region on the wafer with known particle sizes. A) Imaged with high-resolution instrument and B) corresponding scattered map generated with CFS. C) The pre-trained neural network should generalize and return the label of zero for background-roughness input (blue square), and return label of one for background plus particle region (red square).

The concept here is to monitor the layer in the deep neural network (NN) where these particle images + background are far away from the background example. The penultimate layer of the NN in this regard can become the right candidate. Alternatively, the activation patterns at the convolution layers can be monitored. The main difference between NN and straightforward approach of using the convolutions is the availability of training data, trainable randomized kernels, as well as the presence of non-linear activations. With all three approaches being viable candidates, we concentrate on describing the idea of the proof-of-concept experiment with the aid of NN.

"We don't see a particle, but we know that it is there" is the summary of training the network based on the data of low a SNR or a signal-to-background ratio (SBR). Additionally, to the target design, SEM analysis can be used to confirm that the fab-

rication was successful Figure 9.1A). Furthermore, the analysis of the same region of interest is performed with CFS as shown in Figure 9.1B). Due to the fact the position of the particle is known in CFS data, the cuts can be performed over the positions from (x_1, y_1) to (x_n, y_n). All the other positions are considered to be purely the roughness of the sample, such as (x_r, y_r). The network is trained to classify between the data from (x_n, y_n) and the data from (x_r, y_r) positions. At a test time, the network should succeed to provide the highly-accurate results for the true positions of the tiny particles that were never showed to the network at the training time. N.B. aiming at the particle detection with high-density on the surface, similar training of the network should be performed on the single-pixel level, rather than on the cut-out boxes, thus imposing higher load on the creation of the training dataset.

9.3. Outlook 2. Nanoparticle shell-layer quantification

In the model derived in [Roy, S., et al. "Radially polarized light for detection and nanolocalization of dielectric particles on a planar substrate." Physical review letters 114.10 (2015): 103903.] the expressions for the scattered amplitude of the dipole embedded in the surface and measured at the far-field detector for s and p polarization was presented.

$$\begin{aligned} A_s^{dip} &= \mathcal{F}(\boldsymbol{E})(k_x, k_y, 0) \cdot \hat{\boldsymbol{s}}(\boldsymbol{k}) = -\frac{k^2}{2i\epsilon_0 k_z} \boldsymbol{\mu} \cdot \hat{\boldsymbol{s}}(\boldsymbol{k}) + 0 \\ &= -\frac{k^2}{2i\epsilon_0 k_z} \left[\frac{1}{\sqrt{k_x^2 + k_y^2}} (\mu_x k_y - \mu_y k_x + 0) \right] \\ &= \frac{1}{2i\epsilon_0} \frac{k/k_z}{\sqrt{k_x^2 + k_y^2}} \left[k k_x \mu_y - k k_y \mu_x \right] \end{aligned} \tag{9.1}$$

$$\begin{aligned} A_p^{dip} &= \mathcal{F}(\boldsymbol{E})(k_x, k_y, 0) \cdot \hat{\boldsymbol{p}}(\boldsymbol{k}) = -\frac{k^2}{2i\epsilon_0 k_z} \boldsymbol{\mu} \cdot \hat{\boldsymbol{p}}(\boldsymbol{k}) + 0 \\ &= -\frac{k^2}{2i\epsilon_0 k_z} \left[\frac{1}{\sqrt{k_x^2 + k_y^2} k} (\pm \mu_x k_x k_z \pm \mu_y k_y k_z - \mu_z (k_x^2 + k_y^2)) \right] \\ &= \frac{1}{2i\epsilon_0} \frac{k/k_z}{\sqrt{k_x^2 + k_y^2}} \left[\mp k_x k_z \mu_x \mp k_y k_z \mu_y + (k_x^2 + k_y^2) \mu_z \right] \end{aligned} \tag{9.2}$$

We expanded the electric field vector of the plane wave with wave vector $\boldsymbol{k}$ on the orthonormal basis $\hat{\boldsymbol{s}}(\boldsymbol{k})$, $\hat{\boldsymbol{p}}(\boldsymbol{k})$, where we have used Eq. 2.34 for the Fourier transform of dipole field and the orthogonal basis s and p defined in Eq. 2.10 that is perpendicular to the wave vector can be re-written using definition of k Eq. 2.2

$$\hat{\boldsymbol{s}}(\boldsymbol{k}) = \frac{1}{\sqrt{k_x^2 + k_y^2}} \begin{pmatrix} k_y \\ -k_x \\ 0 \end{pmatrix}, \quad \hat{\boldsymbol{p}}^{\pm}(\boldsymbol{k}) = \frac{1}{\sqrt{k_x^2 + k_y^2 k}} \begin{pmatrix} \pm k_x k_z \\ \pm k_y k_z \\ -k_x^2 - k_y^2 \end{pmatrix} \tag{9.3}$$

It is known that the field at the focus of a radially polarized pupil distribution has a large longitudinal component. If the dipole moment is oriented along z, $\hat{\boldsymbol{\mu}} = \mu_x \hat{\boldsymbol{x}} + \mu_y \hat{\boldsymbol{y}} - \mu_z \hat{\boldsymbol{z}}$, with $\mu_x = \mu_y = 0$, where the detector is located above the surface at the negative z position, then the electric fields of the plane waves radiated by the dipole has only p-component:

$$A_p^{dip} = -\frac{1}{2i\epsilon_0} \frac{k\sqrt{k_x^2 + k_y^2}}{k_z} \mu_z \tag{9.4}$$

The simpler than in Eq. 2.52 version of the dipole moment excitation by the incident field at the focal plane z_0, can be written without the dipole's self-coupling.

$$\boldsymbol{\mu} = \alpha \boldsymbol{E}^{inc}(\boldsymbol{r_0}) \tag{9.5}$$

Under assumption that the focus field is barely displaced $x \ll \lambda$, incident focal field is represented with $\boldsymbol{E}^{inc}(\boldsymbol{r_0}) \approx E^{inc}(0)$. Following the shift properties of the Fourier transform, if we shift the nano-particle in the focal plane along the x axis by $r = X$, then there will be a corresponding phase shift of $ik_x X$ in the scattered far-field.

$$A_s^{det} \approx 0 \tag{9.6}$$

$$A_p^{det} \approx r_p A_p^{inc} - (1 + r_p) \frac{\alpha}{2i\epsilon_0} \left[\frac{k\sqrt{k_x^2 + k_y^2}}{k_z} \right] E^{inc}(0) e^{ik_x X} \tag{9.7}$$

The goal of classification with CFS is to relate the signal recorded at the split detector with the diameter of the particle (known material and in scanning). In principle, the bigger the $E^{inc}(0)$ the bigger the signal is due to the particle (second term in Eq. 9.7). Assuming the ideal detector and perfect flatness sample, if the excitation field is increased, the signal of the particle can always overcome the limit emerging from the fundamental physical phenomenon of photon noise. If one would keep the light level constant while reducing the particle size, the detection will eventually be limited by the Poisson noise.

For applications in biology, the label-free disease sensing requires identification and quantification of various bioparticles such as DNA, RNA, proteins, viruses, exosomes, and bacteria. In established techniques of surface plasmon resonance or noble metal nanoparticle-based sensor, a typically gold particle is immobilized on a glass surface and functionalized with receptor strands to capture analyte (DNA). From the CFS perspective, to detect these strands, the change in the size of the shell that they cause of the core-shell particle has to be estimated.

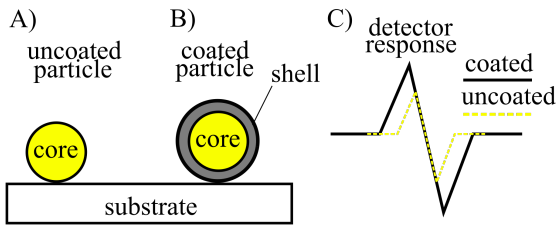


Figure 9.2: A) The schematic of the uncoated core particle and B) same core but with the shell layer on top. C) The expected signals at the detector of CFS.

The workflow is to calibrate the signal for the changes in the shell layer of the known material. In Figure 9.2, the substrate hosts the fixed diameter known material core particle A) and the same core particle with a shell layer of controlled thickness B), where, at the detector, the characteristic change of the signal is observed C).

The multipolar polarizability of spherical geometry which contains dipole for $l = 1$, $l = 2$, and octapole $l = 3$, and so on, defined for the coated nanosphere is given by

$$\alpha_l = \frac{l[(\epsilon_c - \epsilon_m)(l\epsilon + (l+1)\epsilon_c)r_c^{2l+1} + (\epsilon - \epsilon_c)(l\epsilon_m + (l+1)\epsilon_c)r^{2l+1}]}{(l\epsilon_c + (l+1)\epsilon_m)(l\epsilon + (l+1)\epsilon_c)r_c^{2l+1} + l(l+1)(\epsilon - \epsilon_c)(\epsilon_c - \epsilon_m)r^{2l+1}} r_c^{2l+1} \tag{9.8}$$

where, ϵ and ϵ_c are the complex dielectric constants of the core and the coating layer and ϵ_m is the dielectric constant of surrounding medium. Further r is the radius of the core and r_c the radius of the core plus coating layer. For the sphere coated with different material, the dipolar polarizability $l = 1$ becomes:

$$\alpha_d = 4\pi\epsilon_m r_c^3 \left[\frac{(\epsilon_c - \epsilon_m)(\epsilon + 2\epsilon_c)r_c^3 + (\epsilon - \epsilon_c)(\epsilon_m + 2\epsilon_c)r^3}{(\epsilon_c + 2\epsilon_m)(\epsilon + 2\epsilon_c)r_c^3 + 2(\epsilon - \epsilon_c)(\epsilon_c - \epsilon_m)r^3} \right]. \tag{9.9}$$

Using estimation theory, the limit of the technique can be found from the lower bound on the variance of the value of r_c when the scattered intensity is measured, assuming Poisson noise. N.B. The proposed CFS method applied to detection of bioparticles is intended to operate outside the resonance regime in such a way that sudden changes in the signal properties are absent and will not affect the calibration.

A

Theoretical background

A.1. Maxwell equations, boundary conditions, time-harmonic fields and Green's function

In this thesis, we study the interaction of light with subwavelength nanoparticles using electromagnetic theory. One approach is to quantify this interaction analytically. Another method is to solve the governing equations numerically as the geometry of the problem increases in complexity. In both cases, Maxwell's equations must be solved [1, 2]

$$\nabla\times\mathcal{H} = \mathcal{J}+\partial\mathcal{D}/\partial t, \quad \text{(A.1a)} \qquad \nabla\times\mathcal{E} = -\partial\mathcal{B}/\partial t, \quad \text{(A.1b)} \qquad \nabla\cdot\mathcal{B} = 0, \quad \text{(A.1c)} \qquad \nabla\cdot\mathcal{D} = \rho. \quad \text{(A.1d)}$$

where $\mathcal{H}$ is the magnetic field, $\mathcal{J}$ is the density of free current, $\mathcal{D}$ is the electric displacement, t is the time variable, $\mathcal{E}$ is the electrical field, $\mathcal{B}$ is the density of magnetic flux and ρ is the density of free electrical charge.

The following constitutive relations are presumed to hold,

$$\mathcal{D} = \epsilon_0\mathcal{E} + \mathcal{P}_e = \epsilon_0(\boldsymbol{I} + \chi_e)\mathcal{E} = \epsilon_0\epsilon\mathcal{E}, \quad \text{(A.2)}$$

$$\mathcal{H} = \frac{\mathcal{B}}{\mu_0} - \mathcal{P}_m = \frac{\mathcal{B}}{\mu_0} - \chi_m\mathcal{H}, \quad \text{(A.3)}$$

$$\mathcal{B} = \mu_0(\boldsymbol{I} + \chi_m)\mathcal{H} = \mu_0\mu\mathcal{H}. \quad \text{(A.4)}$$

where ϵ_0 is the electric permittivity of vacuum, $\mathcal{P}_e$ is the electric polarization, χ_ϵ the electric susceptibility and ϵ the electric permittivity of the medium. Therefore, in Eq. A.2 the electrical displacement is replaced by medium-dependent electrical permittivity times of the electric field as light interacts with a linear media. For linear media, the constitutive relationship explains also a magnetic induction. With μ_0 magnetic vacuum permeability, $\mathcal{P}_m$ magnetic polarization, the magnetic susceptibility of χ_m and

the medium's magnetic permeability of μ (Eq. A.3) and Eq. A.4). Constituent relations account for inhomogeneous media if ϵ and μ are position functions.

Next, Ohm's law describes a linear relationship between the free current in a medium and the em field. Current can be split into a source current density $\mathcal{J}_s$ and induced conductive current density $\mathcal{J}_c$. The latter is written in terms of the electrical conductivity of the medium σ.

$$\mathcal{J} = \mathcal{J}_s + \mathcal{J}_c = \mathcal{J}_s + \sigma\mathcal{E}. \tag{A.5}$$

In the case where two media share an interface, the first medium with ϵ_1 and μ_1 and the second with ϵ_2 and μ_2, the tangential and normal fields must, as a consequence of the Maxwell's equations, fulfill the boundary conditions at the interface. Remember, for example, Eq. A.1c, with the Theorem of Gauss

$$\int_V \boldsymbol{\nabla} \cdot \mathcal{B} dV = \oint_{\partial V} \mathcal{B} \cdot \hat{\boldsymbol{n}} dS, \tag{A.6}$$

where the surface that encloses a volume V is ∂V and $\hat{\boldsymbol{n}}$ is the normal unit vector to an interface pointing to medium two. The following boundary condition is identified,

$$\hat{\boldsymbol{n}} \cdot (\mathcal{B}_2 - \mathcal{B}_1) = 0, \tag{A.7}$$

Equation A.1d (Poisson's) gives, in a similar manner,

$$\hat{\boldsymbol{n}} \cdot (\mathcal{D}_2 - \mathcal{D}_1) = \rho_s. \tag{A.8}$$

where the surface charge density on the interface is ρ_s The Stoke's theorem,

$$\int_S \boldsymbol{\nabla} \times \mathcal{E} \cdot d\vec{S} = \oint_{\partial s} \mathcal{E} \cdot d\vec{l}, \tag{A.9}$$

The curve enclosing a surface S is ∂S, when applied to Eq. A.1b (Faraday's Law), produces,

$$\hat{\boldsymbol{n}} \times (\mathcal{E}_2 - \mathcal{E}_1) = 0, \tag{A.10}$$

And similarly, when applied to Eq. A.1a (Ampere's law), gives

$$\hat{\boldsymbol{n}} \times (\mathcal{H}_2 - \mathcal{H}_1) = \mathcal{J}_f. \tag{A.11}$$

Here, $\mathcal{J}_f$ defines the current on the interface between medium one and two.

Further, time-harmonic sources with a fixed angular frequency ω are essential in optics. We now consider EM sources and fields which, by implication, contain time-dependence in the form of $e^{-i\omega t}$ and by convention, define the physical quantity of the electrical and magnetic field with the real part of the complex time-dependent quantity.

$$\mathcal{E}(\boldsymbol{r}, t) = Re[\boldsymbol{E}(\boldsymbol{r})e^{-i\omega t}], \tag{A.12}$$

$$\mathcal{H}(\boldsymbol{r}, t) = Re[\boldsymbol{H}(\boldsymbol{r})e^{-i\omega t}]. \tag{A.13}$$

A

As a result, Maxwell's equations combined with constitutive relations (Eq. A.1a - Eq. A.4) are re-written as

$$\nabla \cdot \epsilon\epsilon_0 \boldsymbol{E} = \rho, \tag{A.14}$$
$$\nabla \times \boldsymbol{E} = i\omega\mu\mu_0 \boldsymbol{H}, \tag{A.15}$$
$$\nabla \cdot \mu\mu_0 \boldsymbol{H} = 0, \tag{A.16}$$
$$\nabla \times \boldsymbol{H} = \boldsymbol{\mathcal{J}}_s + \sigma \boldsymbol{E} - i\omega\epsilon\epsilon_0 \boldsymbol{E}. \tag{A.17}$$

Similarly, let $\boldsymbol{\mathcal{J}}(\boldsymbol{r}, t)$ and $\boldsymbol{J}(\boldsymbol{r})$ be the real and complex current densities.

$$\boldsymbol{\mathcal{J}}(\boldsymbol{r}, t) = Re[\boldsymbol{J}(\boldsymbol{r}) e^{-i\omega t}]. \tag{A.18}$$

Thus the alternative form for the Eq. A.17 is

$$\nabla \times \boldsymbol{H} = -i\omega\epsilon\epsilon_0 \boldsymbol{E} + \boldsymbol{J}. \tag{A.19}$$

These are complex Maxwell equations for time-harmonic fields in vacuum. If we apply $\nabla\times$ to Eq. A.15 and Eq. A.17, and multiplying corresponding equations with μ^{-1} and ϵ^{-1}) respectively one gets

$$\nabla \times \mu^{-1} \nabla \times \boldsymbol{E} - k_0^2 \epsilon \boldsymbol{E} = i\omega\mu_0 \boldsymbol{\mathcal{J}}_s, \tag{A.20}$$

$$\nabla \times \epsilon^{-1} \nabla \times \boldsymbol{H} - k_0^2 \mu \boldsymbol{H} = \nabla \times \epsilon^{-1} \boldsymbol{J}. \tag{A.21}$$

with the vacuum wave number being $k_0 = w/c$ and the complex dielectric $[\epsilon + i\sigma/w\epsilon_0] \rightarrow \epsilon$. To describe the electromagnetic field interacting with matter one needs to solve the equations above. These are second-order differential equations called **vector Helmholtz equation** for the electric and magnetic fields.

Lastly, we need to introduce the solution of the wave equation written in a complex form for a potential. Its impulse response is known as the Green's function of the wave equation. Consider time-harmonic vector potential $\boldsymbol{A}$ and the scalar potential ϕ in an infinite and homogeneous space characterized by the constants ϵ and μ:

$$\boldsymbol{E}(\boldsymbol{r}) = i\omega \boldsymbol{A}(\boldsymbol{r}) - \nabla\phi(\boldsymbol{r}), \tag{A.22}$$
$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{\mu_0\mu} \nabla \times \boldsymbol{A}(\boldsymbol{r}), \tag{A.23}$$
$$\nabla \cdot \boldsymbol{A}(\boldsymbol{r}) = i\omega\mu_0\mu\epsilon_0\epsilon\phi(\boldsymbol{r}). \tag{A.24}$$

Where the Eq. A.24 is the Lorentz gauge, a procedure to cope with redundant degrees of freedom in field variables. We can plug Eq. A.22 and Eq. A.23 into Maxwell equation A.19, multiply both sides by $\mu_0\mu$ and obtain

$$\nabla \times \nabla \times \boldsymbol{A}(\boldsymbol{r}) = \mu_0\mu \boldsymbol{J}(\boldsymbol{r}) - i\omega\mu_0\mu\epsilon_0\epsilon[i\omega \boldsymbol{A}(\boldsymbol{r}) - \nabla\phi(\boldsymbol{r})] \tag{A.25}$$

With the aid of identity $\nabla \times \nabla\times = -\nabla^2 + \nabla\nabla\cdot$, $k = \omega\sqrt{\epsilon\epsilon_0\mu\mu_0}$, and Eq. A.24 the Eq. A.25 is re-written as

$$[\nabla^2 + k^2]\boldsymbol{A}(\boldsymbol{r}) = -\mu_0\mu \boldsymbol{J}(\boldsymbol{r}). \tag{A.26}$$

which is the inhomogeneous Helmholtz equation and holds independently for each A_i of $\boldsymbol{A}$. Skipping the derivation, the equation for the scalar potential ϕ is given by

$$[\nabla^2 + k^2]\phi(\boldsymbol{r}) = -\rho(\boldsymbol{r})/\epsilon_0\epsilon. \tag{A.27}$$

This typical form of scalar Helmholtz equation (Eq. A.27) can be set to have a single point source term $-\delta(\boldsymbol{r}-\boldsymbol{r'})$ (RHS) and the scalar potential is replaced with the scalar Green's function $G_0(\boldsymbol{r},\boldsymbol{r'})$ (LHS). The idea is $\boldsymbol{r}$ denotes the location of the field point, i.e. the point at which the fields are to be evaluated, whereas the coordinate $\boldsymbol{r'}$ designates the location of the point source. In free-space, the solution to such equation is given by

$$G_0(\boldsymbol{r},\boldsymbol{r'}) = \frac{e^{\pm ik|\boldsymbol{r}-\boldsymbol{r'}|}}{4\pi|\boldsymbol{r}-\boldsymbol{r'}|}. \tag{A.28}$$

The solution with the plus sign denotes a spherical wave that propagates out of the origin whereas the solution with the minus sign is a wave that converges towards the origin. Since we are going to re-use the **free-space Green's function**, the simpler notation for the large distance is $G(r) = -\frac{e^{ikr}}{4\pi r}$ with $r = |\boldsymbol{r}|$.

A.2. Angular spectrum representation, far-field and Fourier transform by a lens

Here we outline the important relations from nano- and Fourier optics available in references [3, 4]. For the homogeneous media such as air or silicon substrate, the powerful technique called angular spectrum representation can be used to represent the optical fields. The series expansion of an arbitrary field gives a superposition of plane waves with variable amplitudes and propagation directions. Provided the electric field $\boldsymbol{E}(\boldsymbol{r})$ at some point $\boldsymbol{r} = (x, y, z)$ in space and propagation axis z, we can calculate the 2D Fourier transform at known source plane $z = const$, further the field can be propagated and reconstructed through an inverse spatial Fourier transform in the destination plane.

The 2D Fourier transform is written as $\mathcal{F}(.)$, for the general case of three spatial variables the notation of $\hat{}$ is used. With, $\boldsymbol{k} = k_x\hat{x}+k_y\hat{y}+k_z\hat{z}$, $k = wn/c = w\sqrt{\epsilon\mu}/c$, and wave vector's z-component $k_z \equiv \sqrt{(k^2 - k_x^2 - k_y^2)}$, Cartesian transverse coordinates (x, y) and the corresponding spatial frequencies (k_x, k_y) the 2D forward and inverse Fourier transforms are represented by the following two equations.

$$\hat{\boldsymbol{E}}(k_x, k_y; z) = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} \boldsymbol{E}(x, y, z)e^{-i[k_x x+k_y y]}dxdy, \tag{A.29}$$

$$\boldsymbol{E}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\boldsymbol{E}}(k_x, k_y; z)e^{i[k_x x+k_y y]}dk_x dk_y. \tag{A.30}$$

N.B. When performed on a uniform spatial grid, the forward and inverse Fourier trans-

A

forms are performed with FFT's.

We should assume that the medium is homogeneous, isotropic, linear and source-free in the transverse plane. Then a time-harmonic, optical field of angular frequency ω has to fulfill the vector Helmholtz equation (similar to Eq. A.26).

$$[\nabla^2 + k^2]\boldsymbol{E}(\boldsymbol{r}) = 0. \tag{A.31}$$

Assuming the time-dependence according to Eq. A.12, the solution of the Eq. A.30 can be plugged back to a Helmholtz equation A.31, changing the operation in square brackets [.] and integration, becoming

$$\iint_{-\infty}^{\infty} [\nabla^2 + k^2][\hat{\boldsymbol{E}}(k_x, k_y; z)e^{i[k_x x + k_y y]}]dk_x dk_y = 0, \tag{A.32}$$

Further,

$$\iint_{-\infty}^{\infty} \left[(-k_x^2 - k_y^2 + k^2)\hat{\boldsymbol{E}}(k_x, k_y; z) + \frac{\partial^2 \hat{\boldsymbol{E}}(k_x, k_y; z)}{\partial z^2}\right] e^{i(k_x x + k_y y)}dk_x dk_y = 0. \tag{A.33}$$

The equation above holds for all x and y thus the term under the square brackets is required to be zero. The function $\hat{\boldsymbol{E}}(k_x, k_y; z)$ satisfies the differential equation

$$\frac{\partial^2 \hat{\boldsymbol{E}}(k_x, k_y; z)}{\partial z^2} + \left(k^2 - k_x^2 - k_y^2\right)\hat{\boldsymbol{E}}(k_x, k_y; z) = 0, \tag{A.34}$$

The general solution to this differential equation, with $A(k_x, k_y)$ and $B(k_x, k_y)$ being arbitrary functions is given by

$$\hat{\boldsymbol{E}}(k_x, k_y; z) = A(k_x, k_y)e^{ik_z z} + B(k_x, k_y)e^{-ik_z z} = \hat{\boldsymbol{E}}(k_x, k_y; 0)e^{\pm ik_z z}. \tag{A.35}$$

The plus sign in the exponential term is defined for the propagation into the upper half-space $z > 0$ and the minus sign for the propagation into the lower half-space $z < 0$. The Eq. A.35 (**propagator**) is used to propagate the field spectrum in the spatial frequency domain. Inserting the Eq. A.35 in Eq. A.30 gives

$$\boldsymbol{E}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\boldsymbol{E}}(k_x, k_y; 0)e^{i[k_x x + k_y y \pm k_z z]}dk_x dk_y. \tag{A.36}$$

The equation above gives **angular spectrum representation** (ASR) also called angular spectrum or plane wave decomposition. By inspecting the Eq. A.36 and Eq. A.30 we notice that ASR is the inverse Fourier transform when z was set to zero and the propagation $e^{\pm ik_z}$ was included in the integral to represent the backward or forward moving waves.

It is handy to use ASR to demonstrate how the propagation can be done to the far-field regime where $\boldsymbol{r} = \boldsymbol{r_\infty}$. With unit vector $\boldsymbol{v} = (v_x, v_y, v_z) = \left(x/r, y/r, z/r\right)$, where

$r = \sqrt{x^2 + y^2 + z^2}$

$$\boldsymbol{E}_\infty(v_x, v_y, v_z) = \lim_{kr\to\infty} \boldsymbol{E}(x, y, z) = \iint\limits_{(k_x^2+k_y^2)\le k^2} \hat{\boldsymbol{E}}(k_x, k_y; 0) e^{ikr\left[\frac{k_x}{k}v_x + \frac{k_y}{k}v_y + \frac{k_z}{k}v_z\right]} . dk_x dk_y \tag{A.37}$$

Here the integration range is reduced to the $(k_x^2+k_y^2) \le k^2$ because the evanescent waves don't reach the far field zone. If for the source plane one sets $k_x \to kv_x, k_y \to kv_y$ then unit vector can be written as

$$\boldsymbol{v} = \left(k_x/k, k_y/k, k_z/k\right). \tag{A.38}$$

Since the double integral of Eq. A.37 is given in the form

$$F(k) = \iint_D f(x, y) e^{ikg(x,y)} dx dy, \tag{A.39}$$

with real functions of $f(x, y)$ and $g(x, y)$ of the real variables x, y, D closed domain which is simply connected. The asymptotic approximation to $\boldsymbol{E}_\infty(v_x, v_y, v_z)$ is given by

$$F^{(1)}(k) \sim \frac{2\pi i \sigma_c}{k\sqrt{|\Delta|}} f(x_1, y_1) e^{ikg(x_1,y_1)}. \tag{A.40}$$

indices 1 for the stationary points, determinant $\Delta \neq 0$, with σ_c taking values $\sigma_c = \pm 1$ or $\sigma_c = -i$. The $\sigma_c = -1$ for the far-field approximation [5]. Accordingly, the asymptotic solution to Eq. A.37 is given by

$$\boldsymbol{E}_\infty(v_x, v_y, v_z) = -2\pi i k v_z \hat{\boldsymbol{E}}(kv_x, kv_y; 0) \frac{e^{ikr}}{r}, \tag{A.41}$$

With the definition of the unit vector as Eq. A.38 and the solution from Eq. A.41, the Fourier representation is written in terms of the far field approximation as

$$\hat{\boldsymbol{E}}(k_x, k_y; 0) = \frac{ire^{-ikr}}{2\pi k_z} \boldsymbol{E}_\infty(k_x, k_y). \tag{A.42}$$

This relation substituted back to the ASR Eq. A.36 gives

$$\boldsymbol{E}(x, y, z) = \frac{ire^{-ikr}}{2\pi} \iint\limits_{(k_x^2+k_y^2)\le k^2} \hat{\boldsymbol{E}}_\infty(k_x, k_y) e^{i[k_x x + k_y y \pm k_z z]} \frac{1}{k_z} dk_x dk_y. \tag{A.43}$$

Having a close look at Eq. A.43 for the propagating within NA $(k_x^2 + k_y^2) \le k^2$ waves, assuming $k_z \approx k$ to remove factor $1/k_z$, the field on the LHS and RHS form a Fourier transform pair given $z = 0$. Finally, let's consider a lens which has z as optical axis, and finite extension in transverse xy plane. Under the assumption of the thin lens and paraxial approximation, the lens is characterized by a transmittance function that introduces a parabolic phase delay to the wavefront. The goal is to get a focal plane

representation of the input located d_i before the lens. Input is seen as certain transmittance function $t_A(x,y)$ modified by the plane wave of amplitude A_i. By combining the transfer function representation of the Fresnel diffraction, i.e. spatial frequency representation of the light from the source plane incident on the lens, and free space propagation by the distance $z = f$, the complex amplitude in focal plane $U_{f(k_x,k_y)}$ of the lens is given by

$$U_{f(k_x,k_y)} = \frac{e^{i\frac{k}{2f}\left(1-\frac{d_i}{f}\right)\left(k_x^2+k_y^2\right)}}{i\lambda f} \iint_{-\infty}^{\infty} A_i t_A(x,y) exp\left[-i\frac{2\pi}{\lambda f}(k_x x + k_y y)\right] dxdy, \quad \text{(A.44)}$$

With the input to the lens placed exactly at the distance $d_i = f$, the formalism becomes

$$U_{f(k_x,k_y)} = \frac{1}{i\lambda f} \iint_{-\infty}^{\infty} A_i t_A(x,y)\, exp\left[-i\frac{2\pi}{\lambda f}(k_x x + k_y y)\right] dxdy. \quad \text{(A.45)}$$

This formula is valid to describe the input-output relation for the lens of the CFS setup because paraxial approximation is satisfied. Namely, in CFS to make use of the full NA, we provide uniform illumination over the back focal plane of microscope objective by making the incident beam diameter much larger than the aperture. Inspecting the Eq. A.45 we see the forward Fourier transform with a multiplicative factor.

References

[1] H. D. Young, R. A. Freedman, T. Sandin, and A. L. Ford, *University physics*, Vol. 9 (Addison-Wesley New York, 1996).

[2] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).

[3] J. W. Goodman, *Introduction to Fourier optics* (Roberts and Company Publishers, 2005) pp. 100–110.

[4] L. Novotny and B. Hecht, *Principles of Nano-Optics* (Cambridge University Press, 2006).

[5] L. Mandel and E. Wolf, *Optical coherence and quantum optics* (Cambridge university press, 1995) pp. 100–145.

B

Schematic of the legacy split detector

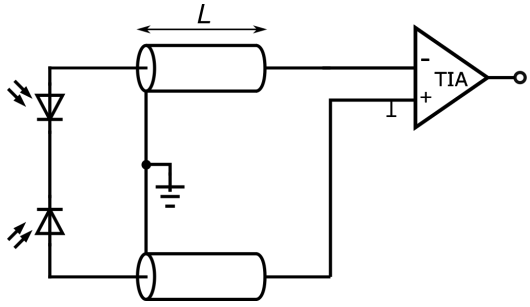


Figure B.1: The initial schematic design of the differential detector circuit. Two photodiodes are connected in reverse and are not grounded. Two coaxial cables of $L \approx 10$ cm transmit the photocurrent from each photodiode, connecting to the transimpedance amplifier (TIA).

The detector circuit (Figure B.1) is suitable for the detection of $d_{psl} \approx 100$ nm. As the scattering from the particles $d_{psl} > 100$ nm renders high SNR, the initial design would still allow for pickup of electric hum and other noise sources. The reason is the absence of the ground for the photodiodes connected in reverse as well as the presence of the coaxial cables leading to the transimpedance amplifier (TIA).

In the new design (shown in Chapter 8), the two TIAs were moved closer to the photodiodes, providing better shielding for hum noise is provided. The drawback of having a single TIA is a high impedance circuit which is generally more sensitive to noise. This is because a small current induced on a high impedance circuit (I times Z) results in a higher noise voltage.

C

Fabrication of experimental samples

C.1. Spin-coated PSL particles

In this dissertation, two types of samples were used: spin-coated PSL nanoparticles and negative resist (EBL) deposited on silicon wafers (further called as type 1 and type 2, respectively). Samples were prepared in a clean room class ISO 6 and high quality 1 inch wafers from Ultrasil were used. The general procedure for the sample preparation is outlined below:

- Clean UV/Ozone apparatus with IPA wipe and switch on for 15 minutes
- Prepare solution
- Place solution in ultrasonic bath
- Clean 1-inch Si wafer in UV Ozone for 5 minute
- Spin 0.5 ml solution on wafer @ 6100 RPM
- Place wafer in box

Table C.1: Recipes for spin-coating PSL particles that have been used for both glass and silicon substrates.

Sample # and date	Recipe. Solution of Particles.
1 22/03/2019	100 nm 3 droplets 100 nm PSL dispersion (Thermo scientific, Nanospheres, 3100A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) dilute 80 µl in 5 ml IPA (Sigma-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
2 22/03/2019	80 nm 3 droplets 80 nm PSL dispersion (Thermo scientific, Nanospheres, 3080A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) 100 µl in 5 ml IPA (Sigma-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
3 22/03/2019	60 nm 3 droplets (Thermo scientific, Nanospheres, 3060A) in 0.5 ml demi water(from Merck Simplicity UV water purification system) 60 µl in 5 ml IPA (sima-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
4 22/03/2019	50 nm 3 droplets (Thermo scientific, Nanospheres, 3050A) in 0.5 ml demi water(from Mecrk Simplicity UV water purification system) 50 µl in 5 ml IPA (sima-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
5 22/03/2019	40 nm 3 droplets (Thermo scientific, Nanospheres, 3040A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) 40 µl in 5 ml IPA (sima-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
6 13/11/2019	Lower density 50 Adapted recipe from above. Steps above are followed by an additional dilution step 500 microliter of the second dispersion were added to 5 ml IPA.
7 13/11/2019	Lower density 40 As in 6th sample
8 10/12/2019	Mixed 80 and 60 nm * 2 droplets 80 nm PSL dispersion (Thermo scientific, Nanospheres, 3080A) and 1 droplet 60 nm PSL dispersion (Thermo scientific, Nanospheres, 3060A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) * Dilute 70 µl in 5 ml IPA (Sigma-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
9 10/12/2019	Mixed 60 and 50 nm * 2 droplets 60 nm PSL dispersion (Thermo scientific, Nanospheres, 3060A) and 1 droplet 50 nm PSL dispersion (Thermo scientific, Nanospheres, 3050A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) * Dilute 70 µl in 5 ml IPA (Sigma-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking
10 10/12/2019	Mixed 50 and 40 nm * 2 droplets 50 nm PSL dispersion (Thermo scientific, Nanospheres, 3050A) and 1 droplet 40 nm PSL dispersion (Thermo scientific, Nanospheres, 3040A) in 0.5 ml demi water (from Merck Simplicity UV water purification system) * Dilute 70 µl in 5 ml IPA (Sigma-Aldrich, 2-Propanol, anhydrous, catalogusnummer 278475-1L) under vigorous shaking

C.2. Protocol for PSL sample analysis

The time span of the influence of the particle in the split detector signal depends on the particle size and has been calibrated for isolated particles of several sizes. The determination of the calibration curve is an iterative process where every new sample goes through two steps. Firstly, samples are validated to contain the target particle sizes with the highest density and, secondly, fitting and extrapolation is used to check whether the new samples are consistent with the previous measurements. The following validation procedure for the wafers from the same batch was used to make sure that after deposition the sample indeed contains the target-size particles:

- Clean the substrate firstly with pure water, and secondly with acetone, and finally with ethanol. If the supplier is trusted cleaning is not necessary. Scan the blank wafer.
- Deposit the spin liquid on the sample, with no particles in the solution, and repeat the scan.
- Make a spin-coating of the third wafer sample and make a final scan.

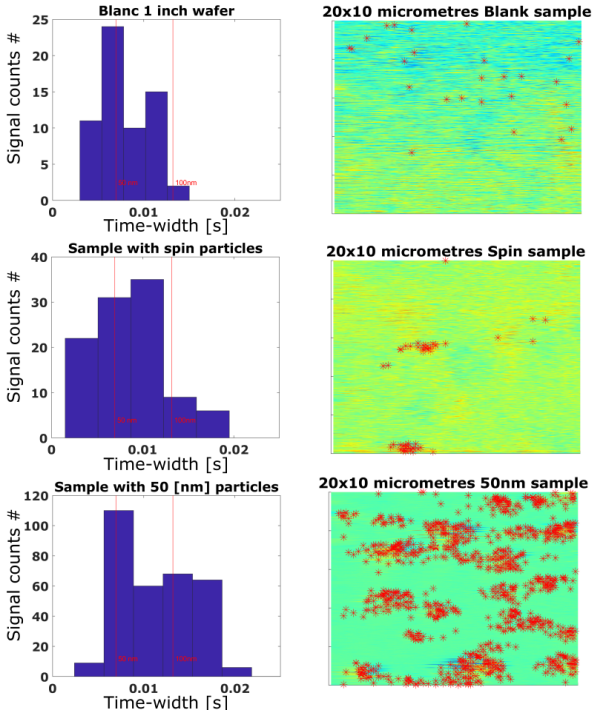


Figure C.1: Comparison between blank, blank plus spin coated, and blank plus 50 nm PSL particles of sample number 4 of the Table C.1.

We analyze reference surfaces of clean 1-inch wafer sample that comes from the manufacturer (blank), wafer sample with solvent (blank + spin), and sample with particles e.g of 50 nm (spin-coated spheres.). The following detection maps are acquired for the sample number 4 of the Table C.1.

Results shown in Figure C.1 demonstrate that wafers coming from the manufacturer are not clean on the level we do the detection. However, this additional contamination is quite sparse. A spin sample also adds the particles in the range we want to detect. Finally, the sample with nominal 50 nm particles includes a high density of particles with the peak that corresponds to 50 nm based on the calibration curve. However, the problem of the final sample (with 50 nm nanoparticles) was that there were too many particles and for this reason we detected a lot of clusters. For the next iteration of samples number 5 and 6 of the Table C.1 the density of particles was reduced so that the clustering could be minimized. The Figure C.2 on the right side shows the raw scans, and on the left side, the corresponding histograms.

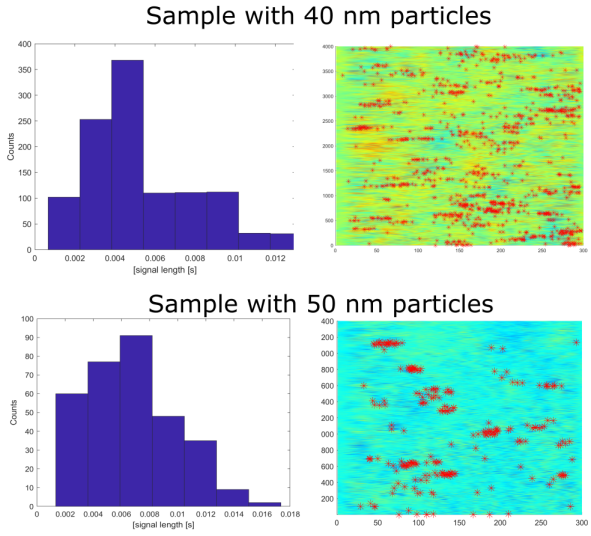


Figure C.2: Histograms (left) and raw scan (right) of the samples with 40 nm (top) and 50 nm (bottom). The scan area is $20 \times 20\ \mu m^2$. Sample number 5 and 6 of the Table C.1

For the sample with nominal 50 nm particles, we can discriminate about $9 - 12\ \frac{\#particles}{area}$, whereas for the sample of nominal 40 nm, $26 - 28\ \frac{\#particles}{area}$. This density was considered satisfactory and was used for the deposition of particles on the glass samples with and without an enhancing layer (Section 4.5). Note that although the deposition of 30 nm particles was planned, after some iterations with TNO, we decided not to do that because it has been observed that smaller particles (< 30 nm) tend to form clusters and thus were not suitable for calibration purposes of our scatterometer.

In the following example of calibration curve the particles have refractive index of $n = 1.58$ (PSL) and experimental curve acquired for $40, 50, 60, 80$ and 100 nm diameter spheres on top of silicon wafer with cubic interpolation towards 30 nm particle (Figure

C.3).

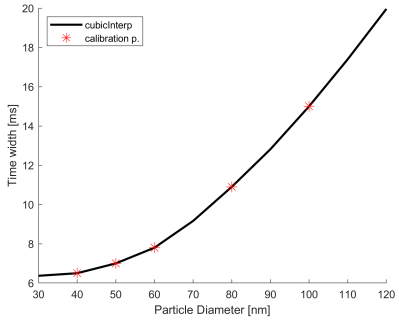


Figure C.3: The calibration curve for the CFS prototype. Time width of the measured pulse as function of PSL particle diameter.

Every experimental point is the mean from the histogram of the one-size sample and the fit is defined based on Error Sum of Squares (SSE) metric.

C.3. Negative resist particles

We continue by describing the fabrication of the type 2 (resist-samples).

- The wafer was cleaved to smaller pieces (1×1 or 2×2 cm).
- Chips were spun with the HSQ XR-1541 diluted in MIBK resist at different speed to control the thickness.
- E-Beam exposure: 100 kV Raith ebeam lithography system, beam current of 112 pA (estimated spot size of 1.9 nm). The dose is 7250 $\mu C/cm^2$ for the $50 \times 50 \times 25$ nm sample and $9500 \mu C/cm^2$ for $25 \times 25 \times 25$ nm sample.
- Development: MF322 (1 min) , MF322: H_2O ~ 1:5 (30 sec), H_2O (1 min).

It is crucial to make a dilution of the e-beam resist and perform the spinning at different speed, to optimize for the heights of the future pillars. Table C.2 summarizes the developed recipes and shows the results of the thickness measured with an ellipsometer (Woollam M5000).

Table C.2: HSQ XR-1541 was diluted with MIBK. And spun with different speeds

Recipe #	Dilution HSQ:MIBK (by volume)	Spinning speed (RPM)	Measured thickness [nm]
1	1:2	4000	53
2	1:2	6000	47
3	1:3	6000	25

Given that the refractive indices n are 1.57 (PSL) and 1.45 (resist) at room temperature with a probe wavelength of 405 nm, we estimate the LSEs in Table C.3 for

the square prism and cylinder resist particles using the formulas for their volumes and deducing the equivalent sphere diameter:

$$\frac{1}{6}\pi d^3 = wlh_{resist} = \frac{\pi}{4}d'^2 h_{resist}, \tag{C.1}$$

$$d_{sphere-equiv} = \left(\frac{6}{\pi}wlh_{resist}\right)^{\frac{1}{3}} = \left(\frac{3}{2}d'^2 h_{resist}\right)^{\frac{1}{3}}, \tag{C.2}$$

where w and l are the width and the length of the resulting rectangular prism ($w = l$ for our fabricated structure), d' and h_{resist} are the diameter of cylindrical particle and the height of the resist. Finally, we compute the Latex Sphere scattering Equivalent for CFS setup:

$$LSE = d_{sphere-equiv} * \left(\frac{(m_{resist}^2 - 1)(m_{psl}^2 + 2)}{(m_{resist}^2 + 2)(m_{psl}^2 - 1)}\right)^{1/3}. \tag{C.3}$$

Table C.3: Resit particle types for which the LSE is estimated

Volume of resist particle [nm^3]	Shape of particle	Sphere diam. equiv. [nm]	LSE [nm]
50x50x25	Square Prism	49.23	46
50x25	Cylinder	45.42	42.5
25x25x25	Cube	31.02	29

We gratefully acknowledge the TNO for help provided by Christiaan Hollemans christiaan.hollemans@tno.nl, Anton de Jong anton.dejong@tno.nl, Jacques Donck jacques.vanderdonck@tno.nl for making the spinning of particles. Further, Martijn Brouwer martijn.brouwer@tno.nl and Jan Cornelis jan_cornelis.wolff@tno.nl for making the thin layer deposition. Next, Kodai Hatakeyama kodai.hatakeyama@tno.nl for the AFM measurements. We also acknowledge Iman Zadeh I.EsmaeilZadeh@tudelft.nl for the fabrication of negative resist samples.

Acknowledgements

This research project would not have been possible without the support of many people, remarkable individuals, whom I would like to acknowledge. I want to express gratitude to my supervisor and promotor, S.F. Pereira, who was abundantly helpful and offered patience, invaluable assistance and support while guiding me to the successful completion of this project. I extend my appreciation to H.P. Urbach, second promotor of this project, for insightful comments and hard questions. Deepest gratitude to the members of the supervisory committee, W. Coene, S. Bäumer and M. Tas, without their knowledge and assistance this study would not have been successful. I am pleased to say thank you to R.C. Horsten, T. Zuidwijk, and C.T.H. Heerkens, who facilitated, gave me access to the laboratory and research facilities and allowed me to learn from their grasp of the practical aspects of science. I also recognize the support of I.E. Zadeh for hands-on cleanroom work in the final year of the project. Great thanks are due to an exceptional HR: Y. van Aalst, C. J. Christie, and FEM support, M. Hammerschmidt and S. Burger, for ensuring my survival and continued sanity. There are dozens of people who provided many experiences whom I wish to thank.

The following is a non-exhaustive collection of the memories and observations that I've collected during my PhD years at TU Delft. I'm thankful to everyone who was involved and I hope you will forgive me if I have missed you. Starting with the senior members; Working together, Silvania and Paul are an excellent combination of experimentalist and theoretician. Having regular debates during progress meetings was an essential experience for me to learn from two worlds. Given their involvement in teaching, supervision, grant writing, international cooperation, and leadership, it is remarkable that Silvania is so welcoming and friendly, and Paul maintains his good sense of humour. They made it possible to enter the domain of Optics and Photonics and made this journey comfortable for me.

From the early days in the group, the "people across the corridor" impressed me with their working attitude. I'm speaking about Roland and Thim. They start early, never forget to have a good laugh in the morning, and get things done before five o'clock. I would not believe it if someone told me that one of them could not solve an engineering task, or miss an opportunity to make a good dad joke.

The right atmosphere in the group is everyone's job; however, it was always Yvonne who kept it under control. I would also credit her for being very active and helpful. To Wim, Stefan and Marnix, I wish to thank you for making our meetings productive, relaxed and positive. It was always a pleasure to visit an ASML with an invitation from Wim and the great help from Ilse.

I'm thankful to Nandini for giving me her keys and being cheerful; Omar for being happy-go-lucky and telling quality stories; Aurele for his enthusiasm and eagerness to introduce the modernism to the group; Bob for his willingness to share his knowledge of geometrical optics; Gao for his skill in pitching the science; Jeff for his "But, but, but..."; Peter for singing; last, but not least, to Joseph for being the smartest person

in the group.

I'm grateful to the people from TNO and Kavli - Christiaan, Martijn, Anton and Jacques for their quality fabrication and intelligent discussions; Marc, Charles, Hozanna and Eugene for the safety meetings, friendly introductions and supplying the good stuff.

Thanks to all the researchers. "First, you learn to crawl, and then you learn to walk, then you learn to run." As a new PhD in the group, you meet the legacy of "prehistoric" students; you are influenced by fellow old school students and your generation, and at the end of the project you should pass it to the new wave. The legacy of Alberto, Lei, Sarathi, Nitish, Arthur Siewert v.d. Nes, and Arthur Wachters could not be overestimated. Your well-documented research and software made it possible for me to kick-start my own efforts. I could only fantasize about being as useful to the new generation as the old-school crew was for me. The key people who helped me and provided great fun and love were Daniel and Priya. I got a lot of insight and inspiration from Daniel: he cared about me a lot, and I hope, by the time I submit my thesis, you'll have your defense. With Priya, it was always a great pleasure to share the office and laugh about some stupid stuff. Thanks to Matthias for cakes, yachting and courage to seize the project money. I miss the surprised, high-pitched grumble of Fellipe, I miss the quality German humor of Andreas and honesty of Marco.

It was an immense pleasure to share the office with Po-Ju, Po-Sheng, Xiujie and Min. Thanks to Zhe for holding composure and being cool; to Yifeng and Ying for being somewhat extraverted Chinese; to Sander for his philosophy and depth; to Paolo and Xukang for "high-loss", being humble but splendid; Doris for leaking Chinese codes and passion for fashion; Jin for his skills in training cats. I thank Zheng for his profound knowledge of scattering, Iman for his enormous fabrication skill, Lauryna for taking life easy.

For the new generation - Sven, Alex, Thomas og and Thomas junior I think you are very well prepared for the job. I'm happy to graduate as the one and only unbeaten champion of the office in basketball and ping pong (what's written in the thesis is true). You should finally bring Optica to the Nature publication and let Paul keep his promise. I'm thankful to Richard, Justine and Davy for letting me help and supervise and I wish you all the best.

Thanks to Ruben, Arturo, Esther, Zhongwen, Anna, Kafei, Hamed, Luca, Nishant and Florian. Further, I acknowledge, High Tech Systems and Materials Research Program with Project n. 14660, financed by the Netherlands Organisation for Scientific Research (NWO), Applied and Technical Sciences division (TTW) for funding this research, ASML and TNO for the contributions in money and in kind.

Last, but not least, I would like to thank the most important women in my life; my mother Liubov and my wife, Kateryna. Thank you for your endless love and support. They have cherished with me every great moment and supported me whenever I needed it. For me and Kateryna it has been the happiest time living together in Delft, but there is more to come. Thanks to Valery for keeping my mother busy. I'm thankful to Dmytro and Oksana for raising such a nice daughter and giving us life lessons. Thanks to Anna, Viktor, Alena, Denis, Andrey and Vova for the good times visiting us. I'm nourishing the memory of meeting my extended family of Sasha, Natasha, Ruslan, Kateryna, my lovely nieces Sofia and Veronica.

The source of my life energy also resides with my friends. I have great friends, Andrii and Gennadii, unique in many ways. I hope one day we will reunite together with our own families somewhere in Europe. Thanks to Alexander for letting me know about the Netherlands and keeping good relations. Thanks to Vardan for helping us integrate and for her extraordinary taste for life.

Curriculum Vitæ

Dmytro KOLENOV

14-10-1993 Born in Kharkiv, Ukraine.

Education

2000–2010	High School Diploma Lyceum 161 "Impuls" - Kharkiv, Ukraine
2010–2015	Bachelor and Master of Science in Applied Physics V. N. Karazin Kharkiv National University - Kharkiv, Ukraine
2015–2017	Master of Science in Space Physics UiT The Arctic Unversity of Norway - Tromsø, Norway
2017	PhD student in CFS at the optics research group Delft University of Technology - Delft, Netherlands

Awards

2020	Spotlight in Optics: Information Acquisition, Processing, and Display (OSA)
2015	Graduated with honours: V. N. Karazin Kharkiv National University
2015	Norwegian Quota Scheme Scholarship

List of Publications

6. **D. Kolenov, I. E. Zadeh, R. C. Horsten and S. F. Pereira**, *Direct detection of polystyrene equivalent nanoparticles with a diameter of 21 nm ($\sim \lambda/19$) using coherent Fourier scatterometry*, Opt. Express **29(11)**, 16487-16505 (2021).

5. **D. Kolenov, D. Davidse, J. Le Cam, and S. F. Pereira**, *Convolutional neural network applied for nanoparticle classification using coherent scatterometry data*, Appl. Opt. **59**, 8426-8433 (2020).

4. **D. Kolenov and S. F. Pereira**, *Machine learning techniques applied for the detection of nanoparticles on surfaces using coherent Fourier scatterometry*, Opt. Express **28**, 19163-19186 (2020).

3. **D. Kolenov, H. P. Urbach, and S. F. Pereira**, *Effect of polarization in evanescent wave amplification for the enhancement of scattering of nanoparticles on surfaces*, OSA Continuum **3**, 742-758 (2020).

2. **D. Kolenov, P. Meng, and S.F. Pereira**, *A highly sensitive laser focus positioning method with sub-micrometre accuracy using coherent Fourier scatterometry*, Measurement Science and Technology **31** (6) 064007 (2020).

1. **D. Kolenov, R. C. Horsten, S. F. Pereira**, *Heterodyne detection system for nanoparticle detection using coherent Fourier scatterometry*, Proc. SPIE **11056**, Optical Measurement Systems for Industrial Inspection XI, 110561A (21 June 2019).