

Università degli Studi di Modena e Reggio Emilia

Dipartimento di Scienze Fisiche, Informatiche e Matematiche

PhD School in Physics and Nanosciences

XXXVII Cycle

**Computational Ghost Imaging
in Transmission Electron Microscopy**

Advisor:

Prof. Stefano Frabboni

Co-advisor:

Dr. Vincenzo Grillo

Candidate:

Dr. Lorenzo Viani

PhD. School Coordinator:

Prof. Marco Affronte

Academic Year 2023/2024

Contents

Abstract	V
Italian Abstract	VI
Introduction	1
 I Theoretical Background	 3
1 The Transmission Electron Microscope	3
1.1 Electrons in a TEM	3
1.1.1 Wave functions and wave packets	4
1.1.2 Spatial and Temporal Coherence	5
1.2 The Resolution Limit	7
1.3 Magnetic Lenses	8
1.3.1 The Paraxial Equation	11
1.3.2 Aberrations	12
1.4 Layout of a TEM Column	17
1.4.1 Illumination System	18
1.4.2 Imaging System	19
1.5 The Sample Holder	21
1.6 The Aperture Holder	22
1.7 STEM HAADF Imaging and the HAADF detector	23
1.8 Pixelated Detectors	25
1.9 Off-axis Electron Holography	28
1.10 Ptychography	30
 2 Ghost Imaging	 33
2.1 Origin of the Technique: Quantum Ghost Imaging	33
2.2 Classical Ghost Imaging	34
2.3 Computational Ghost Imaging	35
2.4 Applications in Optics and Single-Pixel Imaging	36
2.5 Motivation to make Ghost Imaging in Electron Microscopy	38
 3 Structured Electron Beams	 41
3.1 Beam Shaping Methods	41
3.1.1 Electron wave functions and weak phase objects	42
3.1.2 Thin-film phase plates	43
3.1.3 Electrostatic phase plates	44
3.1.4 Magnetic phase plates	46
3.1.5 Optical modulation	47
3.2 Our Devices and Optimal Beam Shaping for Ghost Imaging	48
3.2.1 Electron Vortex Beams and the QC phase plate	51
 4 Electron Computational Ghost Imaging	 54
4.1 Hybrid Scanning-ECGI Method	56

5	Single Pixel Imaging Algorithms	58
5.1	Non-Iterative methods	58
5.2	Iterative Methods	60
5.3	Ptychography-inspired variation of Alternating Projection	61
6	The Multislice Method	63
II	Materials and Experimental Methods	65
7	Simulations	65
7.1	Expected Far Fields of the Phase Plates	65
7.1.1	Double Sorter 1	66
7.1.2	Quadrichopsticks	69
7.2	Multislice	74
7.3	Simulated Experiments	77
7.3.1	Effect of the number of different probes	77
7.3.2	Effect of the scanning on the Traditional-GI and AP algorithms	78
7.3.3	Simulation of probe reconstruction with ptychography-inspired variation of AP	83
7.3.4	Simulation of atomic resolution imaging	86
8	Experimental Setups	88
8.1	Microscope Setup - Double Sorter I	88
8.2	Microscope Setup - Quadrichopsticks	89
9	Characterization of the Phase Plates	90
9.1	Double Sorter I	90
9.2	Quadrichopsticks	92
10	Electronics and Software for Measurements	100
10.1	Electronics	100
10.2	Python	103
10.2.1	HTTP Server	103
10.2.2	ECGI measurement	105
11	Samples	109
11.1	Microscale binary masks	109
11.2	Au nanoparticles and twisted MoS ₂	109
III	Experimental Results	112
12	Double Sorter I Setup	112
12.1	Acquisition of the Probes	112
12.2	Signal Acquisition	112
12.3	Microscale Imaging	114

13 Quadrichopstick Setup	116
13.1 Acquisition of the Probes and Optimization	116
13.2 Signal Acquisition	117
13.3 Low Convergence - Nanoscale Imaging	118
13.4 High Convergence - Atomic Resolution Imaging	127
13.5 Some considerations about the experimental protocol	131
14 Conclusion	133
Acknowledgements	134
Bibliography	136

Abstract

The field of Transmission Electron Microscopy (TEM) counts numerous imaging techniques that can measure and collect information about a sample by quantifying the amplitude and/or phase effects it produces on a transmitted electron wave. However, this analytical capability is limited by the presence of aberrations that can only be partially compensated. As proven by Scherzer, circularly symmetric electron optics are invariably affected by unavoidable spherical aberration. Compensating for its effect requires the use of complex and expensive aberration correctors that extend the limit imposed by aberrations.

In an effort to extend the information sampling beyond the aberration limit and also to mitigate the need for those components, we draw inspiration from a well-established technique in the field of light optics, Computational Ghost Imaging (CGI). CGI is an amplitude imaging method that uses a sequence of structured illuminations and the corresponding transmitted intensities measured on a single-pixel detector to recover an image of the sample. When the technique is adapted to TEM, aberrations are no longer detrimental as long as their effect on the structured illumination is known. That allows for overcoming the aberration-limited resolution even on uncorrected microscopes.

The chapters in this thesis present the efforts to transfer the technique to TEM, going from the proof-of-concept stage to a fully developed experimental setup of Electron Computational Ghost Imaging (ECGI). The project required finding an optimal method to perform electron beam shaping and a single-pixel detector to measure the transmitted intensity. For the beam shaping I used an electrostatic phase plate placed in the microscope's probe-forming system, while the single-pixel detector of choice was the High Angle Annular Dark Field (HAADF) detector because of the properties of the electrons scattered at high angles.

After a theoretical introduction, the thesis documents the work made to characterize two different models of electrostatic phase plates, the simulations done to verify the adequacy of the HAADF signal for the realization of the technique, and all the steps needed to implement a functioning experimental setup. These efforts materialized in the measurement of experimental images of different samples both at the micro- and nanoscale. In particular, the microscale samples were binary amplitude masks, and the experimental setup at the microscale was designed at the proof-of-concept development stage using the first model of phase plate. After the validation of the results from the first experimental setup, the second phase plate model was designed in pursuit of atomic resolution amplitude imaging. The experimental setup was updated and optimized for nanoscale imaging, with the goal of measuring Au nanoparticles in earlier attempts and a bidimensional material for the final result. We were able to experimentally demonstrate ECGI reconstruction with resolution down to 1nm. The results at atomic resolution are still difficult to interpret. A better experimental protocol and analysis method will be required to improve those results. Finally, the thesis explores possible future developments in that direction.

Italian Abstract

Il campo della Microscopia Elettronica a Trasmissione (TEM) conta numerose tecniche per produrre immagini che possono misurare e raccogliere informazioni su un campione quantificando gli effetti di fase e/o ampiezza che questo produce su un'onda elettronica trasmessa. Tuttavia, questa capacità analitica è limitata dalla presenza di aberrazioni che possono essere compensate solo parzialmente. Come dimostrato da Scherzer, le ottiche per elettroni con simmetria circolare sono invariabilmente affette da aberrazione sferica. Compensare il suo effetto richiede l'utilizzo di correttori di aberrazioni complessi e costosi, che estendono il limite dato dalle aberrazioni.

In uno sforzo per estendere il campionamento di informazione oltre il limite delle aberrazioni, ed anche per mitigare la necessità di quei componenti, traiamo ispirazione da una tecnica consolidata nel campo dell'ottica della luce, il Ghost Imaging Computazionale (CGI). Il CGI è un metodo per produrre immagini di ampiezza che utilizza una serie di illuminazioni strutturate, e le corrispondenti intensità trasmesse misurate su un detector a singolo pixel per ricostruire un'immagine del campione. Quando la tecnica è adattata per la TEM, le aberrazioni non sono più dannose, purché il loro effetto sull'illuminazione strutturata sia noto. Ciò consente di superare la risoluzione limitata dalle aberrazioni anche su microscopi non corretti.

I capitoli di questa tesi presentano gli sforzi fatti per trasferire la tecnica in TEM, passando dallo stadio di prova di concetto fino a un apparato sperimentale completo per il Ghost Imaging Computazionale a Elettroni (ECGI). Il progetto ha richiesto di trovare un metodo ottimale per realizzare fasci elettronici strutturati e un tipo di detector a singolo pixel per misurare le intensità trasmesse. Per generare un fascio strutturato io ho utilizzato un piatto di fase elettrostatico inserito nel sistema di formazione del fascio del microscopio, mentre il detector a singolo pixel scelto è stato il detector ad anello per Dark Field ad alto angolo (HAADF), per via delle proprietà degli elettroni diffusi ad alto angolo.

Dopo un'introduzione teorica, la tesi documenta il lavoro svolto per caratterizzare due diversi modelli di piatti di fase elettrostatici, così come le simulazioni fatte per verificare l'adeguatezza del segnale HAADF per la realizzazione della tecnica, e tutti i passaggi necessari per implementare un apparato sperimentale funzionante. Questi sforzi si sono concretizzati nella misura di immagini di campioni sia alla micro- che alla nanoscala. In particolare, i campioni alla nanoscala erano maschere d'ampiezza binarie, e l'apparato sperimentale alla nanoscala è stato progettato allo stadio di sviluppo di prova di concetto usando il primo modello di piatto di fase. Dopo la validazione dei risultati del primo apparato sperimentale, il secondo modello di piatto di fase è stato progettato con lo scopo di produrre immagini di ampiezza a risoluzione atomica. L'apparato sperimentale è stato aggiornato e ottimizzato per lavorare alla nanoscala, con l'obiettivo di misurare nanoparticelle di oro nei primi tentativi e un materiale bidimensionale per il risultato finale. Si è potuta dimostrare la realizzazione di una ricostruzione ECGI sperimentale con risoluzione di 1nm. I risultati a risoluzione atomica rimangono di difficile interpretazione. Saranno necessari un protocollo sperimentale e un metodo di analisi migliori per migliorare quei risultati. Infine, la tesi esplora possibili sviluppi futuri in questa direzione.

Introduction

Electron Microscopy is a branch of Condensed Matter Physics that explores the numerous ways in which it is possible to acquire information about a sample by probing it with an electron beam. A Transmission Electron Microscope is an instrument that uses one or multiple kinds of signals that are emitted as the beam interacts with the sample to image it at extremely high resolution, past the physical limits of ordinary light microscopes. Since its invention in the first part of the twentieth century, the Transmission Electron Microscope has seen the development of numerous techniques that produce images of the amplitude and/or the phase of samples, with applications in many different areas, from material science to biology and medicine.

In the early days of Electron Microscopy, Scherzer demonstrated that all electron optical systems that use circularly symmetric magnetic lenses are invariably subject to spherical aberration. Very often, that and other aberrations pose a limit to the analytic capability of electron microscopes, and compensating for their effect requires the use of complex and expensive aberration correctors. In this thesis project I explored the possibility of adapting a technique from the field of light optics, called Computational Ghost Imaging, to find an alternative way to overcome the limitations posed by aberrations. Ghost Imaging is an amplitude imaging technique, that combines a set of structured illuminations and the corresponding transmitted intensities measured on a single-pixel detector to image the sample. The power of Computational Ghost Imaging relies in the fact that the aberrations can be incorporated in the process that generates the structured illuminations needed by the method, paving the road to overcoming the aberration-limited resolution even on uncorrected microscopes. In the three Parts of this thesis, I show the work done to turn this idea into a functioning experimental apparatus.

Part I provides the theoretical background necessary to contextualize the research work showed later. It begins in Section 1, with a description of the working principles of a Transmission Electron Microscope and two standard Electron Microscopy techniques that were most relevant for my research. Section 2 presents Ghost Imaging, from its origin as a quantum-based method to its applications in the field of light optics. Section 3 contains a summary of the different methods that have been developed to generate structured electron beams and concludes with a more detailed explanation of the one chosen for this project, the use of electrostatic phase plates. Section 4 introduces the concept of the experimental method to be developed in this thesis, Electron Computational Ghost Imaging (ECGI). Section 5 illustrates the different image reconstruction algorithms that have been used to process the experimental data and produce the results. The first Part concludes with Section 6, that quickly describes the Multislice method, a technique used to simulate the propagation of electron beams in a material, that was necessary to determine the range of sample thicknesses for which the ECGI method would apply.

Part II presents the simulation work and the steps that were taken to go from the conceptual stage to the actual realization of a functioning experimental apparatus. Section 7 contains the main body of simulation work done for this thesis, from the simulation of the electrostatic phase plates to those of an entire experiment. Section

8 briefly discusses the setup of the Transmission Electron Microscopes used for the experiments. Section 9 reports the results of the experimental characterization of the real phase plates used to generate structured electron beams. Section 10 shows the electronics and software components of the experimental setup that were needed in addition to the TEM. And last, Section 11 talks about the samples that were chosen to be used during the experiments.

Part III is the most important one, as it presents the experimental results obtained at the different stages of development of the experimental apparatus. Specifically, Section 12 presents the results obtained with the earliest version of the experimental setup, while Section 13 shows the most recent results and the estimation of the highest resolution demonstrably accessible with the technique. The Part concludes with some final remarks concerning the current limitations of the experimental setup and indications on possible strategies that should be adopted to improve it beyond the level reached with my efforts on this project.

Part I

Theoretical Background

This part will provide a theoretical description of the theory behind the techniques and instruments that were used for my PhD work. Particularly, the focus will be on the main instrument, the Transmission Electron Microscope, and the different kinds of Ghost Imaging developed over the years.

1 The Transmission Electron Microscope

The Transmission Electron Microscope (TEM) is an instrument that extracts electrons from a source and manipulates them through a series of electron-optical elements to prepare a controlled probe state on a sample. Then, it uses the electrons transmitted by the sample, or some other product of the electron-sample interaction, to measure high spatial resolution images of the sample, or to map certain properties.

The first Transmission Electron Microscope was built by a German physicist, Ernst Ruska, in 1933. It was just a prototype, a device that could provide an optical magnification by manipulating electron beams with magnetic fields. Now, more than ninety years after that first experimental demonstration, Transmission Electron Microscopes still use that same principle, perfected by the work of hundreds of scientists over time. The invention of the TEM led to the development of Electron Microscopy, an entire branch of Condensed Matter Physics. Electron Microscopy itself branches into several fields, from the demonstrations of open hypotheses in theoretical physics to material sciences studies and biophysics. In the context of this PhD work, the TEM is the main instrument used to set up the experimental technique of interest, so it is necessary to go over its basic working principles.

1.1 Electrons in a TEM

In its most ordinary meaning, the word "microscope" refers to an instrument that is used to observe microscopic samples using light. This is what is called a Visible Light Microscope (VLM), which uses photons in the visible part of the spectrum. In Transmission Electron Microscopy (TEM), instead, the particles that are probing the sample to study it are the electrons. Electrons are charged massive fundamental particles, and this provides advantages and disadvantages in comparison to photons, which are massless and neutral. Nonetheless, according to de Broglie's law, electrons exhibit the same wave-particle duality as photons. As a consequence, they can be visualized both as particles and waves, with a wavelength given by de Broglie's relation:

$$\lambda = \frac{h}{p} \tag{1.1}$$

where p is the electron's linear momentum. For a classical particle, the momentum would be given by the well-known formula $p = m_e v_e$, as the product of the electron's mass and velocity. However, in a TEM the electrons are accelerated by very large acceleration voltages, of the order of hundreds of kV . Then, the probes can reach very short wavelengths and the diffraction-limited resolution can be increased way beyond

the limit of an ordinary microscope. Such accelerating voltages are so large that the electrons reach relativistic speeds, making it necessary to introduce a relativistic treatment of the calculation. The momentum must be derived by looking at the relativistic kinetic energy acquired by the electron after it has been accelerated.

Particles traveling in a medium have a mean free path dependent on their energy, and the mean free path in a dense medium for electrons at standard TEM energies is typically very small, of the order of a few tens of nanometers. So, the inside of a TEM is pumped to high- or ultra-high vacuum, and the electrons propagate in vacuum until they meet the sample. So it is possible to assume vacuum propagation when calculating the momentum. Also, it is assumed that the electrons are accelerated by the accelerating voltage and reach the sample at the same energy. This constraint is usually satisfied by electron microscopes, where the electron-optical elements found along the beam path are not supposed to alter the beam's energy spread (or at most to narrow it in the case of a monochromator). Under these assumptions, the electron's kinetic energy is given by:

$$E_K = E - E_0 = \sqrt{p^2 c^2 + m_{0,e}^2 c^4} - m_{0,e} c^2 \quad (1.2)$$

where E_0 is the electron rest energy and $m_{0,e}$ is the electron rest mass. If we solve this equation with respect to the momentum p and substitute in Eq. 1.1 we obtain the relativistic electron wavelength:

$$\lambda = \frac{hc}{\sqrt{E_K^2 + 2E_K E_0}} \quad (1.3)$$

It is possible to estimate the error made using the classical formula by comparing the wavelengths calculated in both ways. At an accelerating voltage of 100 kV, for instance, the relative error is already:

$$\frac{\lambda_{cl} - \lambda_{rel}}{\lambda_{rel}} = \frac{3.88 pm - 3.70 pm}{3.70 pm} \simeq 5\% \quad (1.4)$$

so it is quite significant. Using the relativistically corrected formula at 300 kV, the operating voltage used for our experiments at high resolution, the resulting wavelength is $\lambda = 1.96 pm$.

1.1.1 Wave functions and wave packets

The dual nature of the electron requires different mathematical tools to describe its behavior, and some phenomena are best described when treating electrons as waves. That is the case for several aspects of TEM, where the wave nature of the electrons becomes relevant to make sense of the experimental observations. To this end, electron waves are described through *wave functions*, typically identified with the Greek letter $\Psi(\vec{r}, t)$. A wave function describes an electron wave, and it depends on both the spatial coordinates $\vec{r} = (x, y, z)$ and the temporal coordinate t . It is a complex probability amplitude function, and it can not be directly measured [1]. The wave function is important because its square modulus is proportional to the probability to find the electron within a certain volume of space at a given time:

$$P(\vec{r}, t) = |\Psi(\vec{r}, t)|^2 = I(\vec{r}, t) \quad (1.5)$$

the square modulus of the wave function is also equal to the intensity of the wave if measured at the corresponding coordinates. In electron microscopy, the wave intensity is the basic property measured by the detectors. The wave function should always be normalized to unity, because the probability of finding an electron somewhere in the universe must always be 1 if the electron exists.

Within the wave optics framework, electron waves are represented by a complex wavefunction with an amplitude and a phase [1]:

$$\Psi = \Psi_0 \exp[i\varphi] = \Psi_0 \exp [i(\vec{k} \cdot \vec{r} - \omega t)] \quad (1.6)$$

Particularly, Equation 1.6 describes a plane wave, which is an unrealistic description because the electron's momentum would be perfectly known, in violation of Heisenberg's uncertainty principle, and its probability density can not be normalized to one as required earlier. Still, it is a convenient approximation in some cases. Electron waves that propagate in free space must obey the Schrödinger equation:

$$i\hbar \frac{\partial \Psi(\vec{r}, t)}{\partial t} = \frac{\hbar^2}{2m} \vec{\nabla}^2 \Psi(\vec{r}, t) \quad (1.7)$$

that is a *linear* and *homogeneous* equation governing the evolution of the wave function. Equation 1.7 is restricted to the case of electrons in vacuum, but it can be used in more complicated settings by adding other terms to it, to represent other individual particles or complex systems like materials. Thanks to the linearity and homogeneity of the Schrödinger equation in vacuum, one valid solution for it is a linear superposition of plane waves. This mathematical construct is called a *wave packet*. A wave packet can be written as:

$$\Psi(\vec{r}, t) = \int \Psi(\vec{k}) \exp(i 2\pi \vec{k} \cdot \vec{r}) \exp(-i\pi \frac{\hbar \vec{k}^2}{m} t) d^3 \vec{k} \quad (1.8)$$

so it is a superposition of plane waves, each with its own amplitude and momentum. A wave packet can realistically represent an electron, as it satisfies the probability normalization condition stated before, and is not perfectly localized in terms of position and momentum as required by Heisenberg's uncertainty principle.

1.1.2 Spatial and Temporal Coherence

One further step for the description of electrons inside a TEM is the concept of *coherence* [2][3]. When electrons are described as wave packets, it is necessary to consider that they might be subject to the typical phenomena observed in wave-optics, like interference. Then, an important property is the coherence of electrons, that expresses to which degree the electrons of the beam can interfere with themselves or other electrons. Coherence is divided into *spatial coherence* and *temporal coherence*, that refer to two different possible modes of interference, and are associated to two different values of *coherence length*.

Spatial coherence, also called transverse coherence, is the maximum distance between two points, in a plane perpendicular to the electron beam, where interference can occur [3]. The spatial coherence is affected by the size of the electron source and by the maximum illumination angle with which electrons reach the observation plane

compared to a reference optical axis. It is quantified by the spatial coherence length l_s , that represents the width of the wave packet, and it can be estimated to be:

$$l_s = \frac{\lambda}{2\beta} \quad (1.9)$$

where β is the electron beam's illumination angle. Figure 1.1 shows the decoherence effect arising as electron beams propagate at slightly different angles with respect to the optical axis. At a distance $P1$ from the optical axis, the wavefronts of beams A and B or A and C are half a wavelength apart, causing maximum destructive interference. Inside the region of diameter l_s , centered on $P0$, the wavefront can be assumed to be well defined. The dependence on the source size becomes clear when thinking of improving coherence by decreasing the illumination angle at the observation plane. Imagining to use a diaphragm with a circular aperture to limit the illumination angle, only some of the electrons emitted by the source can go through the aperture. At fixed number of emitted electrons, the smaller the source is, the larger the number of electrons that will be emitted within the illumination angle.

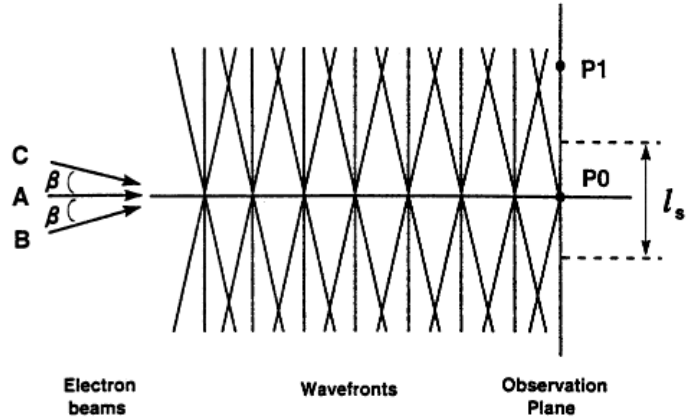


Figure 1.1: Wavefronts of an electron beam, with B and C propagating at a nonzero angle β . By shifting transversely to the optical axis, the wavefronts grow progressively further and further apart, producing decoherence effects. Adapted from [3].

Temporal coherence, also called longitudinal coherence, is associated with the length of the wave packet along the propagation direction [3]. The temporal coherence length l_t depends on the energy spread of the beam $e\Delta V$ (the uncertainty on the energy of the emitted electrons), and is expressed as:

$$l_t = \frac{2V}{\Delta V} \lambda = \frac{\lambda^2}{\Delta \lambda} \quad (1.10)$$

where λ is the wavelength and V is the accelerating voltage of the source. The temporal coherence is relevant when there is a need to observe interference between beams that are split coherently and then overlapped after traveling on different paths. In that case, interference fringes can be observed if the difference of the optical paths for the two beams is smaller than the longitudinal coherence length. Because the temporal coherence length depends of the energy spread of the source, it can be quite relevant to have stable sources with a small energy spread, to maximize the longitudinal coherence length.

1.2 The Resolution Limit

An important metric for any imaging technique is the spatial resolution that it can achieve. The spatial resolution, from here onward simply referred to as resolution, is the smallest separation between two different features in the sample. In general, the final resolution of any imaging method can be affected by multiple factors, but it is intrinsically lower-bound because of diffraction. Thus, the maximum theoretically available resolution for an imaging method is the diffraction-limited resolution, the smallest separation that can be resolved between two different features in the sample in the absence of aberrations. Any features closer than this distance cannot be distinguished because of diffraction.

In many respects, electron waves can be described in the same way as electromagnetic waves. Then, the definition of the diffraction limited resolution can be borrowed directly from light optics. In the absence of aberrations, when a point source is imaged by a lens, it will be spread into a disk. This happens because the lens is finite in size and cannot collect all the light emitted by the point source to make a perfect image. So, because of diffraction, a point is turned into a rippling function called Airy disk.

The diffraction-limited resolution is identified by the Rayleigh Criterion, which is based on the Airy disk. Rayleigh's Criterion asserts that the minimum separation to image two distinct features in the sample is identified as the distance where the first minimum of one's Airy disk overlaps with the maximum of the other's. An example is shown in figure 1.2. Numerically, the diffraction-limited resolution is:

$$R_{th} = 0.61 \frac{\lambda}{NA} = 0.61 \frac{\lambda}{n \sin \beta} \quad (1.11)$$

where λ is the wavelength and NA is the numerical aperture of the lens, product of the refractive index n of the medium between source and lens and sine of the lens' outermost collection semi-angle β . For electron microscopes, the refraction index of the surrounding medium is that of vacuum, so it is 1. The lens's collection semi-angles are typically fractions of a degree, so the sine can be approximated by the angle β . At 300 kV and 7 mrad collection angle, Rayleigh's Criterion gives a theoretical resolution of:

$$R_{th} = 0.61 \frac{1.96 \text{ pm}}{7 \text{ mrad}} \simeq 1.4 \text{ \AA} \quad (1.12)$$

However, this value is better than what is experimentally obtained in a simple TEM by a factor of 2-4, because of aberrations. Aberrations are deviations of the path of the electron beam from the path of ideal propagation. For waves, this translates to distortions in the phase of the wave, that ultimately lead to distortions in the image. Aberrations will be discussed in more detail in Subsection 1.3, while talking about the properties of magnetic lenses.

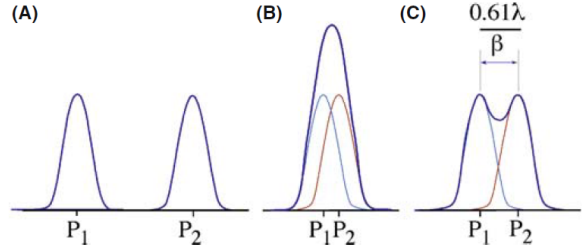


Figure 1.2: (A) two separate Airy disks, images of two separate points P_1 and P_2 of the sample. (B) when the disks' centers are closer than their radii, they appear as a single feature. (C) condition where Rayleigh's Criterion is met. Adapted from [2].

1.3 Magnetic Lenses

Magnetic lenses are one of the key components of a TEM. These objects apply the actual magnification/demagnification effects on the electron beam, using magnetic fields. In a normal TEM there are several different lenses, each one designed for a specific purpose. So, before moving to a description of the structure of a TEM, it is useful to understand how magnetic electron lenses are made and the physical principles behind their operation.

Electron lenses have been first produced only about a century ago by Ruska and Knoll [4][5], and their design process is both conceptually very simple and practically quite challenging due to the high precision needed to machine the components and assemble them. Figure 1.3 shows a simple representation of the main parts of a magnetic lens. Electron lenses are built by surrounding a pair of cylindrical pieces of soft iron, called the *pole pieces*, with a set of copper coils. The pole pieces can be made from two separate pieces of iron, or they can be a single element, but they are always separated by a certain gap along the optical axis. The electrons need a path to pass through the lens, so magnetic lenses have a cylindrical hole at the center, called *bore*. When an electric current is run into the copper coils, it generates a magnetic field inside the lens' bore. Because of the cylindrical symmetry of the lens structure, the field also has an axial cylindrical symmetry. This kind of lens is also called round lens for this reason.

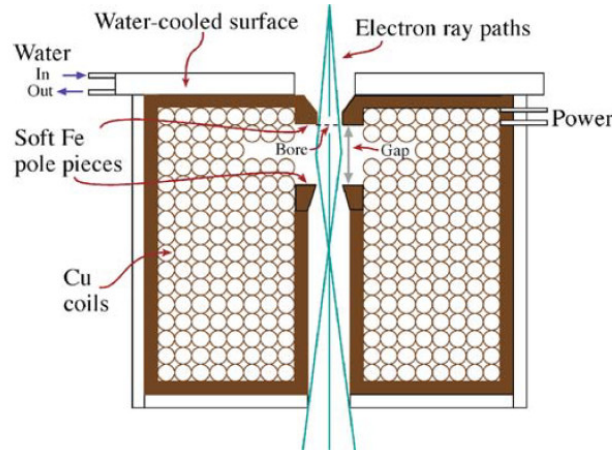


Figure 1.3: Schematic of a sagittal section of an round magnetic electron lens. The beam travels through a cylindrical hole in the central part of the lens, where it feels the magnetic field produced by the surrounding coils and concentrated between the pole pieces. Adapted from [2].

The magnetic field has two components, one transverse to the optical axis and one parallel to it, and both vary along the optical axis. In a well designed lens, the magnetic field generated by it is concentrated inside the bore, between the pole pieces, in the region traversed by the electrons. The focusing effect of a magnetic lens was demonstrated in 1926 by Busch [6][7][8]. He showed how the focal of a magnetic lens can be calculated as:

$$\frac{1}{f} = \frac{\eta^2}{4V^{*2}} \int_{z_i}^{z_f} B^2(z) dz \quad (1.13)$$

where $\eta = \frac{e}{\sqrt{2em_0}}$ and $V^* = [2V(1 + \frac{e}{2m_e c^2}V)]$ is the relativistically corrected acceleration potential. Equation 1.13 tells us that the focal of a magnetic lens changes depending on the strength of the magnetic field that it is producing. The field can be regulated by tuning the current running through the coils, so basically the focusing effect of a magnetic lens is regulated by selecting an appropriate value of the current in the coils. The gap between the pole pieces and the size of the bore can also determine the behavior of the field through the lens, and the various TEM lenses have different gaps and bore sizes to tune their range of possible focal lengths for their specific function in the microscope.

The effect of the magnetic field can be better visualized if one considers what happens to a single electron coming into the lens. First of all, a convention: usually the optical axis of the microscope, that runs parallel to the column, is assigned to the z axis in a right-hand Cartesian frame of reference, where x and y are the transverse components parallel to the various optical planes of the microscope. Assuming perfect cylindrical symmetry of the magnetic field, it is possible to switch to polar cylindrical coordinates without loss of generality, so we can define the radial and transverse components in the usual right-hand orientation. When an electron is moving in a region with magnetic and electric fields, it feels a Lorentz force given by equation 1.14.

$$\vec{F}_L = -e(\vec{E} + \vec{v} \times \vec{B}) \quad (1.14)$$

where $-e$ is the electron's charge, $\vec{E}$ is the electric field, $\vec{v}$ is the electron's velocity and $\vec{B}$ is the magnetic induction field (we can identify it with the magnetic field $\vec{H} = \vec{B}/\mu_0$ since the electron is in vacuum). The field of the lens and a radially projected electron trajectory are shown in Figure 1.4, that also introduces the fact that the electron trajectory is rotated while crossing the lens in panel (b).

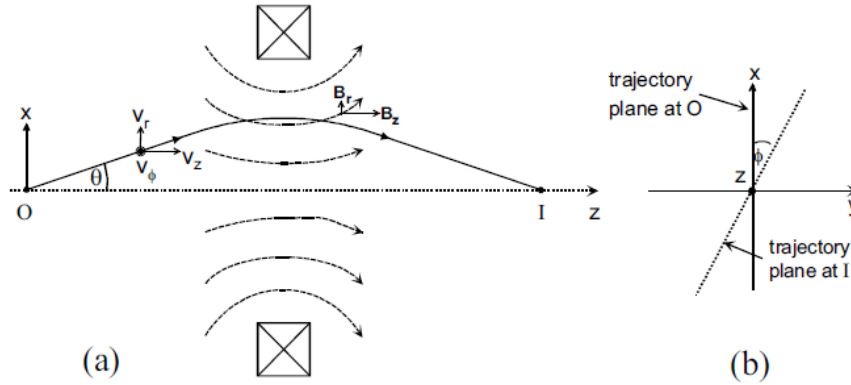


Figure 1.4: (a) schematic electron lens and magnetic field, with the radial projection of an electron trajectory. (b) rotation effect of a magnetic lens, the velocity of the incident electrons at the object plane O is rotated while propagating to the image plane I . Adapted from [9].

Figure 1.5 provides a more detailed description of the effect of a magnetic lens. It shows the radial, transverse and z components of the magnetic field, the Lorentz force and the velocity of the electrons as they change along the optical axis. B_z is always

positive, but B_r changes sign through the lens. If an electron enters the lens with an initial velocity $\vec{v} = (v_r, 0, v_z)$ as in Figure 1.4 (a), it will feel a transverse Lorentz force that pushes it to move in a spiraling motion around the optic axis, because the Lorentz force is dominated by $F_\phi = -ev_z B_r$.

As the electron approaches the center of the lens, the radial component of the field reduces and the z component grows. So, the electron will acquire a transverse velocity, that then in turn interacts with B_z . Close to the center of the lens, the Lorentz force is then dominated by $F_r = -ev_\phi B_z$, a radial force that is pushing the electron *towards* the optical axis.

After the central part of the lens, the z component of the field starts decreasing, and the radial component increases to a positive value. This essentially keeps the focusing effect towards the optical axis but this time the transverse component of the velocity is decreased. As the electron exits the lens and both the field components go to zero, the electron velocity is restored to its initial components. However, in the process the electron has come closer to the optical axis and its velocity has been rotated to a plane at an angle from the entrance one. This rotation effect is well known by electron microscopists, and it is accounted for by manufacturers while designing the instruments.

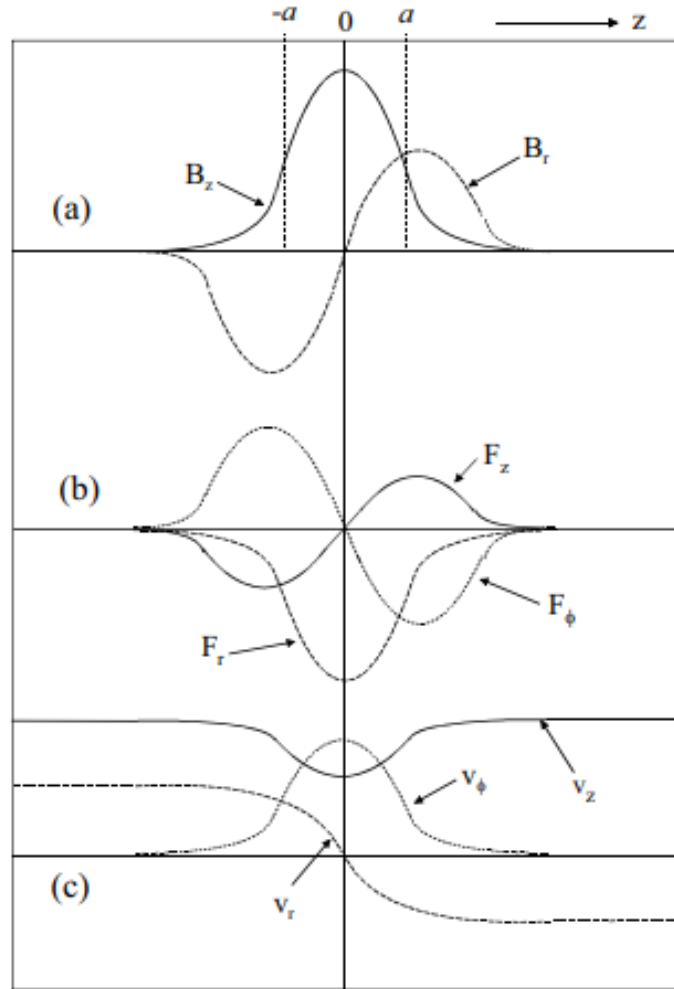


Figure 1.5: (a) radial and z components of $\vec{B}$. (b) components of the Lorentz force on the electron in (c), whose initial velocity is assumed in the x - z plane (no transverse component). (c) velocity of the electron. Its components may change, but the modulus of the velocity stays the same, because $\vec{B}$ cannot do any work on the electron. Adapted from [9].

Now that the focusing mechanism of a magnetic lens is clear, it is useful to make a little comparison with light optics, like the glass lenses of a visible light microscope. Essentially, a magnetic lens can be considered to be analogous to a *converging lens*. Converging lenses have two main functions:

- they can collect a part of the rays emanating from a point of an object and project them onto another point in an image (it actually is a disk, as said in Subsection

1.2)

- they can focus incident parallel rays to a point in their back-focal plane and vice-versa, turn a point source in the front-focal plane into a parallel beam.

Figure 1.6 shows both situations:

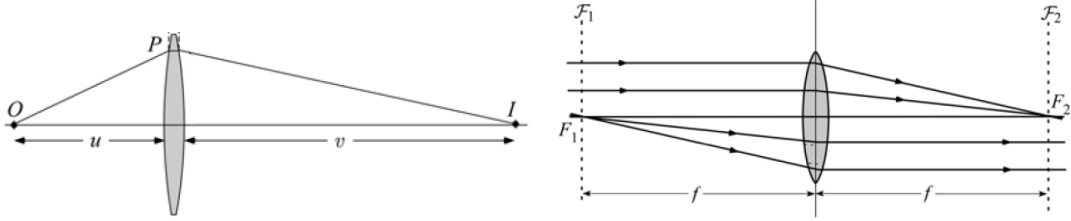


Figure 1.6: Left: example of a converging lens focusing a point from the object O into a point in the image I . Right: converging lens focusing a parallel beam into a point in the back-focal plane $\mathcal{F}_2$ or vice versa, spreading a point in $\mathcal{F}_1$ to a parallel beam. Adapted from [10].

So, magnetic lenses are similar to glass light lenses. However, glass lenses exploit standard light refraction at the interface between two materials, and the focal of the lens cannot be changed because it is determined by the properties of the interface. Magnetic electron lenses, as seen before, have a tunable focal length. Another point in common between the two families of lenses is the presence of optical aberrations. All real optical systems are subject to aberrations, deviations of the electron beams paths from their ideal paths in the optical system. Round electron lenses are at a natural disadvantage in this case, as demonstrated by Scherzer [11]. Scherzer proved how a system of rotationally symmetric lenses must have a positive spherical aberration (definition later in this Section), so this aberration is inherently present in all electron microscopes because all electron lenses are designed with cylindrical symmetry. In addition to this, it is quite difficult to control the magnetic field in the lens' bore precisely and make it perfectly symmetrical, so this always introduces some ulterior aberrations in the lenses.

1.3.1 The Paraxial Equation

In the previous Section it was said that the aberrations are deviations of the electron beam paths from their ideal paths in the optical system. The ideal paths of electron beams can be derived from the equations of motion set by Equation 1.14. The derivation presented here will follow the treatment in [12]. The calculation will assume a particle description of the electron, and the presence of the sole magnetic induction field along the path of the electron beams. By rewriting the Lorentz force and the electron velocities through time derivatives of the electron position, Equation 1.14 becomes:

$$m \frac{d^2 \vec{r}}{dt^2} = -e \frac{d\vec{r}}{dt} \times \vec{B} \quad (1.15)$$

Equation 1.15 is a vectorial differential equation, so it has three components, one for each Cartesian coordinate. In electron microscopy, it is customary to turn the time

dependence into a position dependence along the coordinate associated with the optical axis, usually z . This makes the x and y coordinates dependent on the z coordinate, as the electron move through the lenses and along the optical axis. The position of the electrons in a given plane z can also be expressed using a set of polar cylindrical coordinates, introduced through a complex notation of the kind:

$$\omega(z) = x(z) + iy(z) = u(z) \exp[i\vartheta(z)] \quad (1.16)$$

where u and ϑ are the radial distance from the axis and the angle in the xy plane. This change of coordinates, and a series of approximations, leads to the possibility of rewriting the components of Equation 1.15 as:

$$\begin{aligned} \frac{d\vartheta}{dz} &= \sqrt{\frac{e}{8mV}} B(z) \\ \frac{d^2u}{dz^2} + \frac{e}{8mV} B^2(z)u(z) &= 0 \end{aligned} \quad (1.17)$$

This is known as the Paraxial equation, and it is the equation that describes the ideal trajectories of electrons inside the magnetic fields of electron lenses. In a realistic setting, it is valid for electrons that enter the lenses close to the optical axis and at small angles with it. Electrons that do not meet those conditions are subject to larger aberrations.

1.3.2 Aberrations

In Subsection 1.2 it was mentioned that the resolution of a microscope is lower-bound by diffraction, but the aberrations can further reduce it. After discussing the basic properties of electron lenses, it is worthwhile to have a quick overview of the main aberrations as well. An optical aberration is defined as the deviation of the electron beam path from the path of ideal propagation given by the Paraxial equation. Aberrations can be introduced in an optical system in many ways, and some are always present because of the physics behind electron lenses, while others can be introduced by defects in the fabrication of the components or contamination of the optics over time. As a rule of thumb, the more elements are added to an optical system, the more aberrations will be introduced as well.

In general, the effect of aberrations is described by the presence of higher order terms in the paraxial equation 1.16. In this case, the second order linear differential equation in u becomes a nonlinear differential equation in the form [12]:

$$\omega'' + \frac{eB^2(z)}{8mV}\omega = P(\omega, \bar{\omega}, \omega', \bar{\omega}', z) \quad (1.18)$$

where ω'' , ω' , $\bar{\omega}$ are the second derivative, first derivative, and complex conjugate of ω . The nonlinear term $P(\dots)$ comprises all the nonlinear terms that depend on the electron distance from the optical axis, and the slope of its trajectory as it enters the lens' field. $P(\dots)$ is generally assumed to be a small perturbation to the paraxial equation. Aberrations are not all equal, and it is possible to devise a classification based on their origin mechanism. The rest of the Subsection will provide a synthetic description of the main aberrations relevant for electron microscopy.

First, Chromatic aberrations: they are caused by the fact that a real electron beam is not perfectly monochromatic, but it always has an energy spread. The electrons emitted by an electron source do not emerge all in the same conditions, at fixed energy and from the same spot. The emission of electrons from a material is a stochastic process, so an electron source emits particles at slightly different energies. This is the source of the temporal coherence effect, as seen in Subsection 1.1, and it also has a consequence in terms of aberrations. The physics behind the chromatic aberration can be understood by looking at Equation 1.13. It tells us that the focal of a magnetic lens depends on the electron energy, so electrons with different energies (like light with different colors) will be focused at different points along the optical axis. This in turn means that no matter which plane is chosen to look at an image, a fraction of the beam will always be out of focus, and a point like source will be imaged as a disk. The greater the beam monochromaticity is, the less relevant the chromatic aberrations will be in turn.

Then, there are Geometric aberrations: the name comes from the fact that they can be easily identified in geometric optics, for monochromatic beams, as deviation of the beam path from that of ideal imaging. Here the term *ideal imaging*, also called *Gauss imaging*, refers to the path obtained in paraxial condition. Geometric aberrations have been classified in different ways, but in general it is possible to distinguish them in two groups: *axial* aberrations, for which the spatial deviation of the aberrated beams from the gaussian position depends only on the angle that the beam forms with the optical axis, and *off-axial* aberrations, that depend both on the angle of incidence and the transverse distance of the source point from the optical axis.

By defining the angle ϑ between the ray and the optical axis, and the distance r of the point source from the optical axis, it is possible to quantify the aberrations in terms of their order, that is the sum of the exponents of r and ϑ in the expression of the aberration. The most common geometric aberrations are the Five Seidel aberrations [16], that all share a third-order dependency on ϑ and r when summing their exponents. They are:

- third-order Spherical Aberration
- Off-axial Astigmatism
- Off-axial Coma
- Field Curvature
- Distortion

The last two are normally not a problem in high-resolution electron microscopy because they are dominated by r , which is small at large magnification, so it is more important to focus on the first three that mostly depend on the angle ϑ .

Spherical Aberration is the most important aberration in the microscope. It is caused by the fact that the lens' focal is not uniform across the entire lens, but it becomes gradually shorter for electrons passing through the outer zones [2] [8]. The name *spherical* aberration originates from optics, where it was first characterized as an effect that would prevent perfect focusing of light rays focused by a spherical mirror or lens.

The effect of spherical aberration is to increase the curvature of an incident spherical wave front as it crosses the lens. Figure 1.7 shows this effect. Electrons that enter the lens at a larger angle experience a stronger focal than the paraxial ones, and the result is that a beam with angular aperture ϑ is imaged in the Gaussian plane as a central peak on the optical axis, surrounded by a halo of decreasing intensity. In the presence of spherical aberration, the tightest focus is found in the plane of least confusion, and not in the Gaussian plane of focus determined by the paraxial equation. The image is always broadened to a disk, even for an ideal point source.

Figure 1.7 clearly shows that a point source P that is emitting a spherical wave front of aperture $\vartheta = \beta$ is imaged to a broad disk in the Gaussian plane, with a radius $R_G = C_S \beta^3$. In the plane of least confusion, the disk radius is $R_{min} = 0.25 C_S \beta^3$. C_S is the spherical aberration coefficient, and its typical value for the objective lens, the strongest of the microscope, are between 0.5-2 mm. There are different ways to cope with spherical aberration, but it is undoubtedly the most significant one as it is caused by the device itself and must always be accounted for.

Astigmatism is another very common aberration, and it is found when an optical element causes the beam to be focused more along certain transverse directions than others. The easiest example is the two-fold off-axial astigmatism, where the beam deviation from the ideal propagation depends on $r^2 \vartheta$, so it goes linearly with the angle. There is also 2-fold axial astigmatism, that only depends on the angle ϑ . and is not a Seidel aberration, because it is of lower order. Figure 1.7 shows the effect of the two-fold astigmatism on a beam, where a conical beam with uniform opening angle is focused less in the yellow (Meridional) plane than in

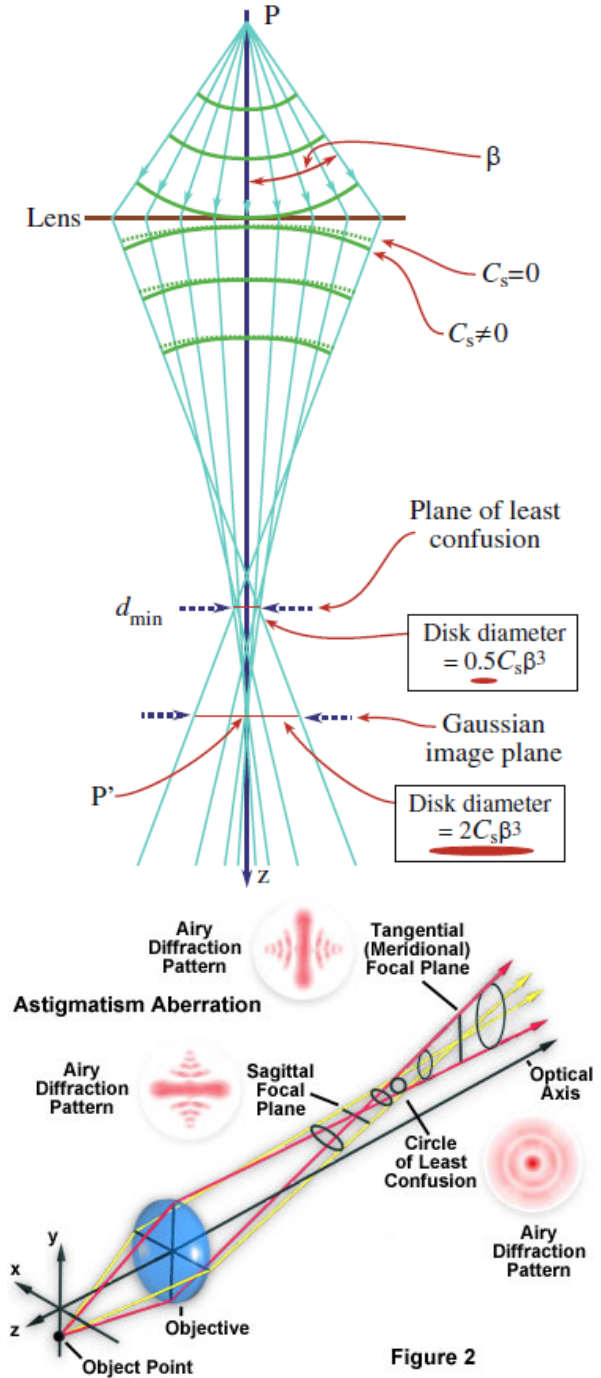


Figure 1.7: Top: diagram of the effect of spherical aberration: a spherical wavefront emitted in a cone of aperture β by a point source P is focused at different points along the optical axis. P' is the position for the Gaussian image plane, but the smallest spot is found in the plane of least confusion. Adapted from [2].

Bottom: effect of the two-fold astigmatism, the beam is focused differently depending on the polar angle of the trajectory of its rays, creating a disk of least confusion. Adapted from [13].

the red (Sagittal) one, creating once again a disk of least confusion between the two different focal points. The radius of this disk is $R_{ast} = \vartheta \Delta f$, so it is proportional to the difference between the position of the foci for the two directions. Astigmatism can be induced by inhomogeneities in the bore of the lenses, or by contamination of the aperture diaphragms, but the two-fold kind is easily corrected by using a dedicated octupole, called a *stigmator*. Astigmatism is such a common problem, and so easily solved, that there are stigmators at multiple positions along the optical axis in a TEM.

The last to be mentioned is Coma. Just like for Astigmatism, Coma is actually divided in two kinds, off-axial and axial. The beam deviation from the gaussian image point depends on $r\theta^2$ for Off-axial coma, so it is actually a Seidel aberration, while axial coma only depends on ϑ^2 . Regardless, both kinds of coma distort the image of the beam into a cone or *comet* shape, hence the name *coma*.

As a final note, it is important to say that the treatment of aberrations so far was mostly based on geometrical optics. Of course, it is also possible to discuss aberrations in terms of wave optics. In that case, the aberrations are seen as perturbations of the paraxial wave front that introduce an additional phase shift. The analogy is understood by considering that the change of beam path seen in geometrical optics aberrations is a change of optical path, so it must correspond to a phase shift in terms of waves optics. A clear description is provided in [14], where the discussion is mostly done for aberrations in image formation under parallel illumination (so, for TEM). However, the treatment of aberrations remains valid even when discussing probe formation with convergent illumination, that is the condition of STEM. In both cases, the discussion is mostly reserved to Axial aberrations, that are most relevant for high-resolution imaging.

The axial aberrations of the wave front are all accounted for through the aberration wave function $\chi(\omega)$ [12], that depends on the azimuthal (θ) and polar (ϕ) angle coordinates of the wave front as defined in Figure 1.8. For a better understanding of the meaning of the aberration function expression for axial aberrations, let us consider what happens to the emission from an object point on the optical axis. Because the object point O is a point source, it will emit a spherically symmetric wave front W . Part of this spherically symmetric wave front will be collected by the lens. In absence of aberrations, the spherical wave front entering the lens will be then imaged into another perfectly spherical wave front that converges into the image point (it is actually an image disk because of diffraction, as explained in Subsection 1.2). In presence of aberrations the ideal wave front W is distorted to an aberrated wave front W' , and

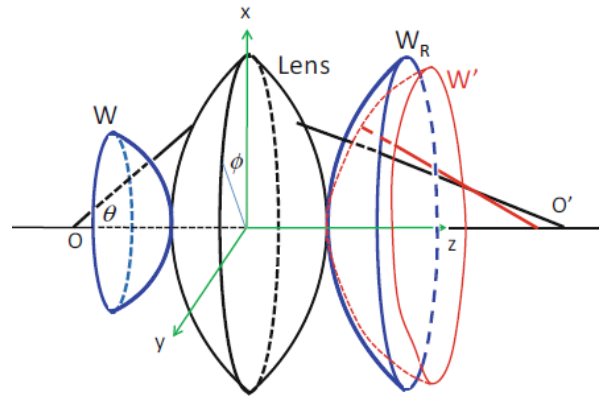


Figure 1.8: Wave front altered by axial aberrations. The unaberrated wave front W is imaged in W_R in absence of aberrations. If aberrations are present, W is imaged into W' , where the shape of the new wave front depends on the aberration function. Adapted from [12].

the amount of distortion at each position can be mapped by a 2D function of the angular part of a spherical coordinate system, the pair (θ, ϕ) . If the two angular coordinates are combined in a complex coordinate $\omega = \bar{\omega}^* = \vartheta \cos \phi + i\vartheta \sin \phi$, the aberration function can be conveniently expressed as a power series involving combinations of θ and ϕ :

$$\begin{aligned} \chi(\omega) = \text{Re} \left\{ \frac{1}{2} \bar{\omega}^2 A_1 + \frac{1}{2} \omega \bar{\omega} C_1 \right. & \quad (I \text{ order}) \\ & + \frac{1}{3} \bar{\omega}^3 A_2 + \omega^2 \bar{\omega} B_2 \quad (II \text{ order}) \\ & \left. + \frac{1}{4} \bar{\omega}^4 A_3 + \frac{1}{4} (\omega \bar{\omega})^2 C_3 + \omega^3 \bar{\omega} S_3 + \dots \right\} \quad (III \text{ order and higher}) \end{aligned} \quad (1.19)$$

Each aberration corresponds to one of the terms in the power series expansion, has its own complex coefficient and is in general a complex function that acts on the complex wave front of the electron wave. The letters used for each coefficient aid in identifying aberrations of similar kind, in the Typke-Dierksen notation [15]. For example, the letter *A* is reserved to astigmatism, *B* to coma, *C* for spherical aberration, *S* for the star aberration, and so on. Figure 1.9 shows a table with some examples of how a convergent electron probe is aberrated by each Axial aberration in the first few orders.

		Symmetry						
	round	1	2	3	4	5	6	
Order	1	Defocus 		2-fold astigmatism 				
	2		Coma 		3-fold astigmatism 			
	3	Spherical 		Star 		4-fold astigmatism 		
	4		4th-order coma 		Three-lobe 		5-fold astigmatism 	
	5	5th-order spherical 		5th-order star 		Rosette 		6-fold astigmatism 

Figure 1.9: List of the first few Axial aberrations. As per Equation 1.19, the order is the pedix of the complex coefficient, and the symmetry is determined by the number of the aberrations' symmetry planes found along the coordinate ϕ . Adapted from [16].

1.4 Layout of a TEM Column

This section provides a description of the main components of a TEM, with the focus on those that are relevant for this PhD work. A TEM is a very complex instrument, but its structure can be once again compared with that of a Visible Light Microscope to understand its basic working principles. In a VLM there is a source of light, placed at the bottom of the microscope. The light is usually directly shined on the sample, but even the simplest microscopes offer an option to regulate the brightness of the light. Once the light has crossed the sample, it is collected by a set of lenses that magnify the image of the sample and project it onto the detector, the microscopist's eye.

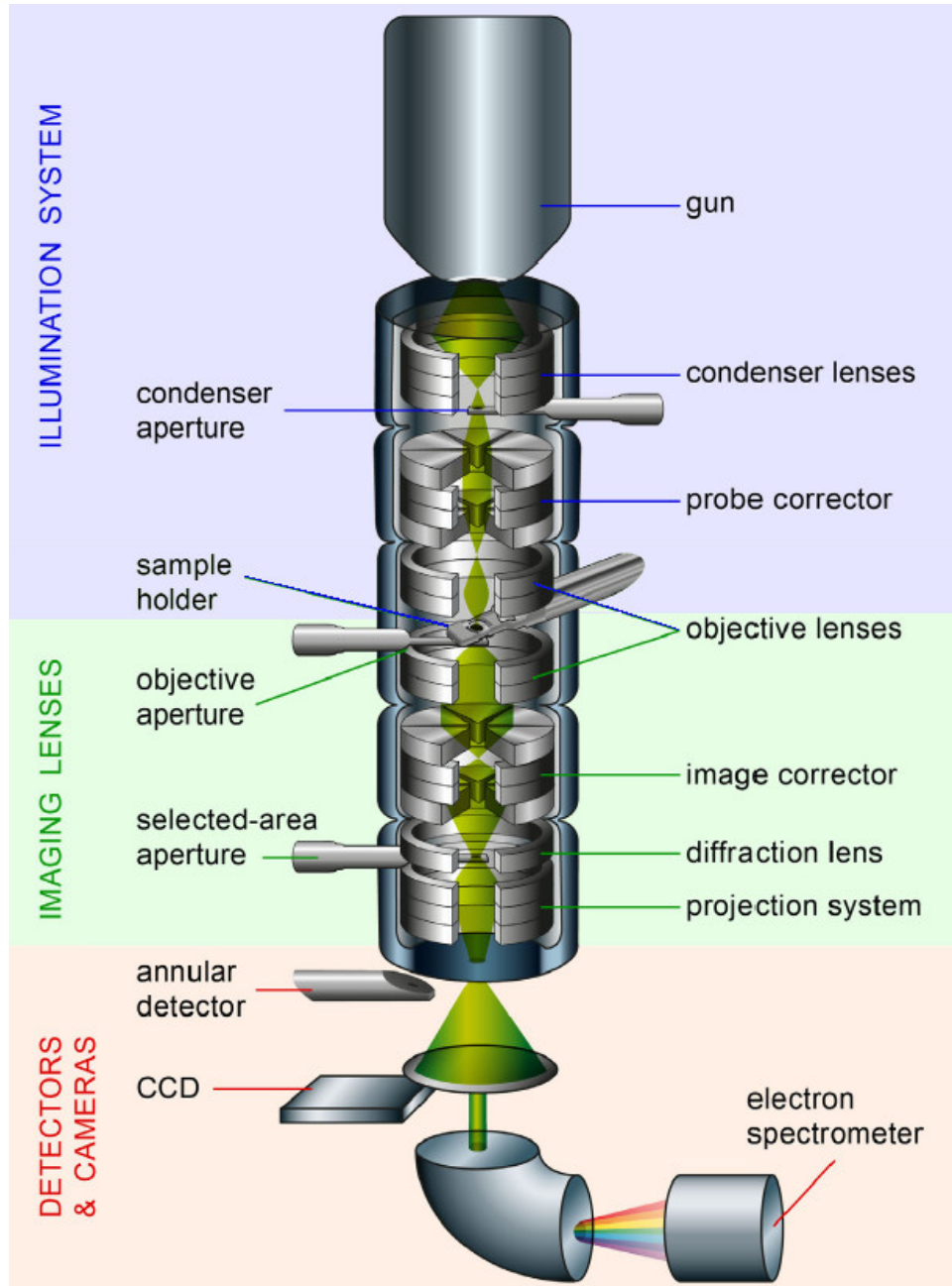


Figure 1.10: Simplified schematic of a modern double-corrected TEM column, indicating the most relevant planes and aperture positions, as well as their division between the tasks of probe formation (illumination system), and image formation and detection (imaging system). Adapted from [17].

So, the optical setup of a TEM can be roughly divided into two sections:

- an illumination system, that contains the source and the optical elements needed to prepare the probe
- an imaging system, that collects the transmitted electrons and projects them to a set of cameras and detectors, either to make imaging or spectroscopy.

Figure 1.10 shows a synthetic view of an analogous partition of a TEM column.

1.4.1 Illumination System

The illumination system consists of the elements needed to produce and shape the electron beam, preparing the probe in the desired state to illuminate the sample. The main elements are the source (electron gun), the condenser lenses and the condenser apertures.

The source is a critical part of the system, because it affects the quality of the beam that is produced and intrinsically limit the quality of the images that can be produced by the microscope. There are several figures of merit for the electron gun, with three of them being most relevant. The source size, the brightness and the energy spread. A good source should have a small size, high brightness, and small energy spread.

The source physical size determines the minimum size of the probe beam, because the lenses in the illumination system can demagnify the source, but this becomes easier if the initial source size is already small. Actually, in a TEM the electrons emitted by the real source are focused into a crossover point that is used as a virtual source for the illumination system. This virtual source is demagnified by the first condenser lens, so it is possible to partially tune the source size, and this will affect the spatial coherence as mentioned in Subsection 1.1.

The source's axial brightness along the optical axis is defined as the emission current density per unit solid angle:

$$\beta_{source} = \frac{\Delta I}{\Delta S} \frac{1}{\Delta \Omega} = \frac{j}{\pi \alpha^2} \quad (1.20)$$

it essentially determines the strength of the electron gun. Its meaning can be easily understood by imagining that the source is a flashlight used to look around in a dark room. If the light is dim and spread over a large cone, it only vaguely outlines the shapes of the objects. If the light is strong and concentrated in a smaller cone, it clearly shows the area where we point it. The axial brightness is constant along the optical axis [8].

The energy spread ΔE is the maximum difference in energy between electrons that are emitted from the gun. All the optical elements of a TEM are designed to operate at a set of nominal accelerating voltages, but the electron beam emitted by the source cannot be perfectly monochromatic, and some electrons will be emitted with a slightly larger or smaller energy than the average. So, the greater the energy spread is, the worse the optical components can function, and chromatic aberrations become more significant. Also, the beam will lose temporal coherence as the energy spread grows,

as reported in Subsection 1.1.

The next elements would be the condenser lenses, but electron lenses are treated in greater detail in Subsection 1.3 so the discussion here will be quick. The role of this combination of magnetic lenses is essentially to demagnify the source to a point, called crossover, and then either project this point onto the sample in a convergent probe, or spread the intensity in a parallel, plane-wave-like illumination condition, depending on the technique's requirements.

Then, there are the condenser apertures. Each magnetic lens has an associated aperture position, with a dedicated aperture holder where a diaphragm can be placed. A diaphragm is an opaque object, usually made of metal, that blocks the beam except for a vacuum region, typically a circular hole, called aperture. Apertures of different size are used to select only part of the electron beam, with variable radii between 1 *mm* and 10 μm . Remember that in a TEM the axial brightness is constant, and the effective source size can be tuned by the first condenser lens. Once the source size is fixed, the only way to get high currents at the specimen plane is to use large apertures for the remaining lenses in the illumination system. So, the use of apertures offers a trade off between the total current that reaches the sample and the resolution, as each aperture and lens can limit the fraction of beam that enters the next optical element along the column. Smaller apertures can select only the central part of the beam, increasing its spatial coherence, but this will also increase the diffraction-limited resolution set by the aperture as per Equation 1.11. Aperture holders are discussed more thoroughly in Subsection 1.6.

The last element that can be present in the illumination system is the probe corrector, a remarkable device that has been developed to compensate high-order aberrations in the probe by using a sequence of multipoles. It allows to greatly reduce the effect of the aberrations when imaging, so that larger apertures can be used to form the probes, improving the available resolution and the current at the same time. Our technique in principle makes use of aberrations, so we set it up on a machine that was not probe-corrected and we will not discuss this tool further.

1.4.2 Imaging System

Moving further along the scheme in Figure 1.10, we reach the imaging system. The first listed component is the sample holder, a metallic rod that can house the sample in a slot close to one of the tips. Thanks to a dedicated fast-entry system, it can be inserted and retracted from the column without breaking the high-vacuum inside. The sample position can be controlled using the sample stage, a dedicated system that can precisely move the holder tip along three Cartesian axes *x,y,z*, and one or two angles. There can be different kinds of sample holders, depending on the shape of the sample and the kind of property that has to be studied. More details on the sample holder are given in Subsection 1.5.

The sample holder is placed *inside* the objective lens, that is usually split into two parts, called pole-pieces. The objective lens is the most powerful lens of the microscope, so it has a short focal distance and a large magnification factor. Because it provides the strongest magnification, it is also the main source of aberrations for the imaging system.

So, it is usually designed to have minimum natural astigmatism. The objective lens also has a corresponding aperture position, in its back focal plane. In normal TEM imaging, the objective aperture is used to select only a part of the sample's diffraction, enhancing contrast in the final image. In doing so, it also determines the collection angle of the objective lens, so it is as important as the lens itself for normal TEM imaging, because it will determine the effect of aberrations on the image and thus the final resolution.

Next in line is the image corrector, a device that mirrors the function of the probe corrector in the illumination system. It comprises a set of multipoles that can be tuned to correct high-order aberrations. It is used when the technique of interest needs to have as little effects from the aberrations as possible when forming the images. It is the case of standard TEM, where the illumination system forms a plane-wave-like probe, with a flat intensity and phase profile, and the image is formed depending on the phase and amplitude modulation that the plane wave acquires after the interaction with the sample.

After the image corrector there is the diffraction lens, that is used in combination with the projection system to switch between the microscope's image mode and diffraction mode. In image mode, the image plane of the objective lens is conjugate to the plane of the camera, and a magnified image of the can be acquired. In diffraction mode, it is the back-focal plane of the objective that is conjugate to the detector plane, and one can see the sample's diffraction. The selected-area aperture is placed in the objective lens' image plane.

Finally, there are the detectors. The picture shows the annular detector, that once inserted along the beam path provides an annular surface to collect electrons scattered between an inner and outer detection angle, while the main beam is transmitted. Next is the fluorescent screen, that depending on the microscope model can be directly visible to the user through a lead glass window, or it can be observed remotely with a TV camera if the TEM's control panel and PC are in a different room than the instrument itself. Then there is usually a Charge Coupled Device (CCD) camera to acquire high-resolution images, and an electron spectrometer if the microscope is equipped to make Electron Energy Loss Spectroscopy (EELS).

1.5 The Sample Holder

The sample holder is the microscope component designed to house the sample and manipulate it inside the microscope. The sample holder is inserted into the microscope through the stage, a specific mechanism designed to keep the inside of the column in vacuum while exchanging samples. The stage is designed to let the sample reach the inside of the objective lens. There are two kinds of stages, designed to work with different geometries of objective lenses, that depend on the microscope's final purpose. If the microscope's primary purpose is to work at very high resolution, the objective has to have a short focal, and the pole pieces must be very close (as explained in Sub-section 1.3). In this case, a *top-entry* stage is used, where the sample is immersed into the objective lens from the top. This kind of objective-stage combination is severely limited in the number of detectors that can be placed near the specimen because of the narrow gap, and only allows small tilt angles for the sample, lest it crashes into the pole pieces. If the microscope is to be used for analytical electron microscopy the constraint on the pole piece gap can be relaxed. In this case, the upper and lower objective pole pieces are usually split, and the sample holder can be easily inserted from the side through a *side-entry* stage. The sample can be manipulated more freely, and it is possible to insert custom holders that do a variety of tasks such as tilting, rotating, heating, cooling, biasing, straining, ... or a combination of these effects.

Since all the microscopes used for this PhD work were produced by Thermo Fisher Scientific, in Figure 1.11 is presented a schematic showing the design of a double-tilt Thermo Fisher sample holder:

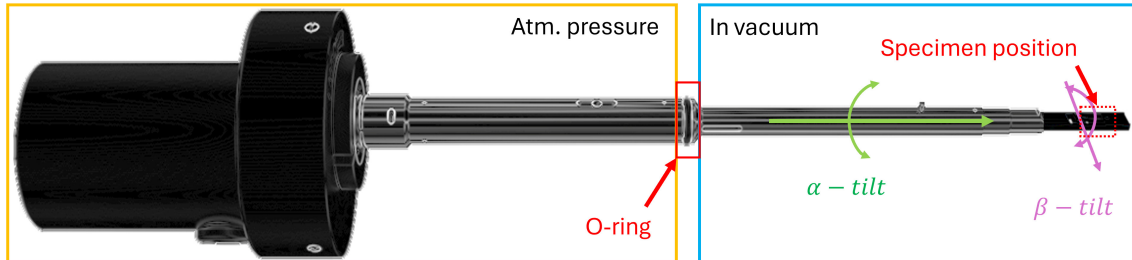


Figure 1.11: Schematic of a Thermo Fisher Scientific double tilt holder, showing the separation between the in-vacuum section and the in-air section, as well as the two possible tilt axes and corresponding angles. The α -tilt motion is displayed on the holder rod for readability, but the only part that actually moves is the tip where the sample is clamped.

In Figure 1.11, it is clearly visible how the holder is separate into two sections by the O-ring. The left side remains at atmospheric pressure and houses most of the mechanical parts that control the sample motion, while the right side reaches inside the column and physically holds the sample. The specimen position is where the holder reaches inside the objective lens' pole pieces. This kind of holder allows for controlled motion with five degrees of freedom. Three are for shifting the sample in three perpendicular directions, x, y, z , where z -shifting is along the optic axis and x, y are in a plane transverse to it. The remaining two are for tilting the sample, with a rotation around the holder axis (α -tilt) or around an axis perpendicular to the optical axis *and* the holder's axis (β -tilting). The sample is kept in position by a screw or clamp, and possibly also

a spacer metallic ring if it is too thin and remains loose when mounted.

As mentioned earlier, sample holders come in a variety of designs to test samples in different conditions, and this can mean that the holders themselves have a slightly different structure, or some additional connections for the passage of wires or tubes to flow liquids. This holds true for aperture holders as well, though custom aperture holders are more uncommon.

1.6 The Aperture Holder

Given the large number of lenses inside of a TEM, and the requirement to select only parts of the beam at different positions in the column, a TEM is bound to have at least a few aperture positions. Aperture holders are not inserted through fast-entry stages, so in order to swiftly swap apertures without breaking the vacuum they are designed to hold a few different diaphragms at the same time. Typically, modern aperture holders have four available positions with interchangeable apertures, but custom holders can have less than four. For this thesis work it is relevant to examine a custom holder with only three positions, two for normal apertures, and one to house a Micro ElectroMechanical System (MEMS) chip. It is displayed in Figure 1.12.

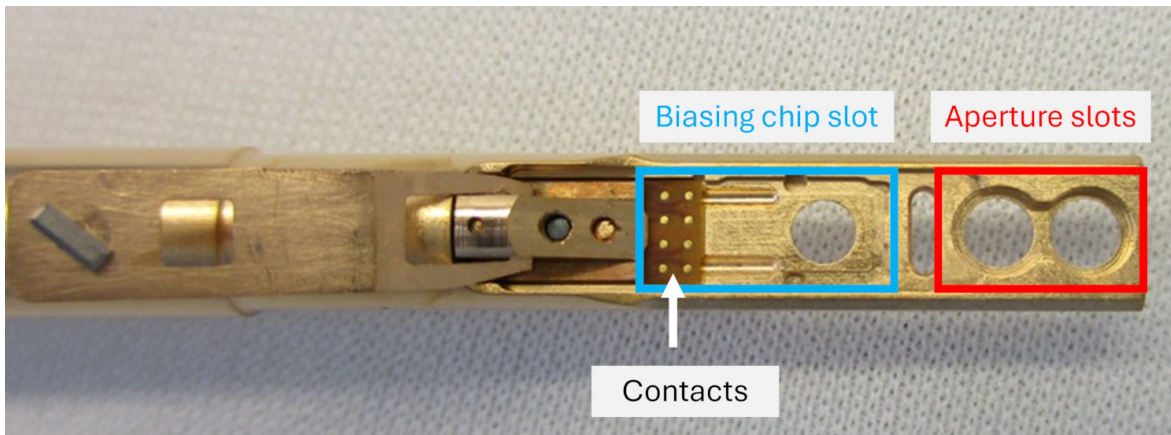


Figure 1.12: Detail of the tip of a custom biasing aperture holder. Aperture diaphragms are inserted by sliding them in from the side, while the chip is to be put with contact-side facing down. The MEMS chip clamp is on the left, and it folds over the chip to keep it in place, leaving only a tiny square window exposed. The clamp is kept in place by a tiny bolt that lock in the hole between the MEMS slot and the aperture slots.

In this kind of holder, two of the four usual aperture positions are replaced by a MEMS chip slot, with a total of eight electrical contacts. The two remaining positions are left for two apertures, so it is possible to seamlessly swap between the special aperture and the ordinary ones, always leaving the microscope fully operational.

1.7 STEM HAADF Imaging and the HAADF detector

The simplest operation modes of a Transmission Electron Microscope are the TEM mode and the Scanning-TEM (STEM) mode. In TEM imaging, the beam is prepared in parallel illumination condition, so that the sample is presented with a flat amplitude and phase profile. After the interaction with the sample, the objective aperture is used to select only a portion of the scattered electrons. As discussed in Subsection 1.4, the objective aperture is in the back-focal plane of the objective lens, so it can be used to isolate only a few or just one diffraction spot. This enhances the final image's diffraction contrast, and is particularly useful to image crystalline samples. When only the central spot is selected, the imaging mode is called TEM Bright Field (BF). The central spot contains only the unscattered electrons, and since vacuum is not scattering, the image will display vacuum regions as the brightest, because they provide the strongest signal. So, the sample will be displayed with dark features on a bright background of the field of view (hence the name Bright-Field). If instead the central spot is excluded, the imaging mode is called TEM Dark-Field (DF) as now vacuum and less scattering regions appear darker, and the only parts of the sample that appear bright are those that scatter in the selected region of the diffraction space. Figure 1.13 shows the two TEM modes in panels (A) and (B).

In STEM mode, the beam is prepared in convergent illumination, so it is focused in a tight spot on the sample. The spot is scanned in a raster pattern through a pair of scanning coils that allow to move the beam on a square grid by keeping the same incidence angle. At each scanning point the beam will produce a diffraction pattern, and similarly to TEM BF and DF, only a part of the scattered electrons is used to form an image. In the case of STEM, though, the image is formed pixel by pixel, one for each scanning point. Each pixel's value is assigned proportionally to the total measured scattered intensity on the detector. If the central spot is measured, the resulting image will be a STEM BF. For STEM DF, instead, the intensity is measured on an annulus in the diffraction pattern, centered on the central spot. The Bright-Field/Dark-Field distinction is just like that of TEM images, with STEM BF images having dark features on a bright background and vice-versa for STEM DF. Figure 1.13 shows the distinction in panels (C) and (D).

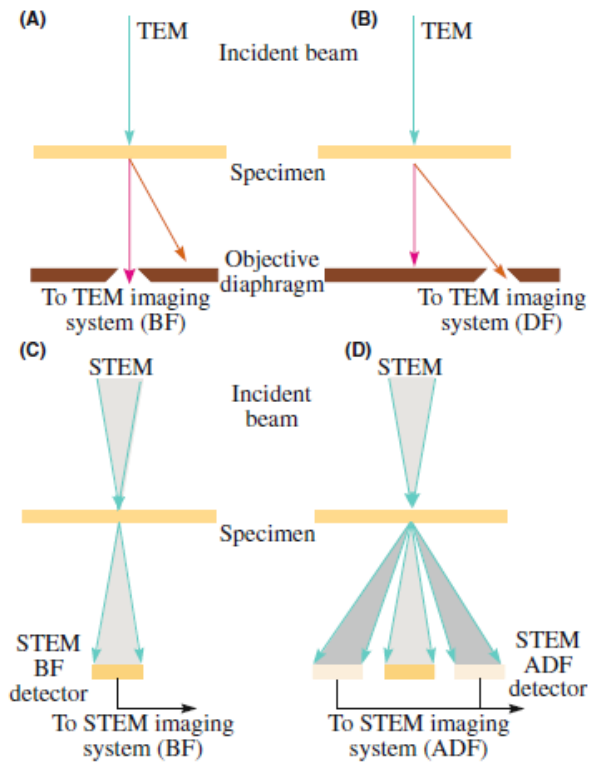


Figure 1.13: Schematic of the TEM and STEM imaging modes, with the two possible kinds of signal collection, Bright-Field and Dark-Field. (A) TEM BF, (B) TEM DF, (C) STEM BF and (D) STEM DF. Adapted from [2].

In STEM, the resolution is determined by the size of the probe, because it sets the smallest separation that can be resolved between adjacent positions. The size of the probe is controlled by the angle that the outermost part of the beam forms with the optical axis, called the *convergence angle*. Usually, it is identified by the Greek letter α . The convergence angle is not the only significant one: there is also the *collection angle* θ . Choosing a certain range of collection angles means deciding if the contribution to image formation comes mostly from elastically or inelastically scattered electrons, and this has an effect on the image's contrast. There are three ranges that separate equally many regimes for STEM imaging:

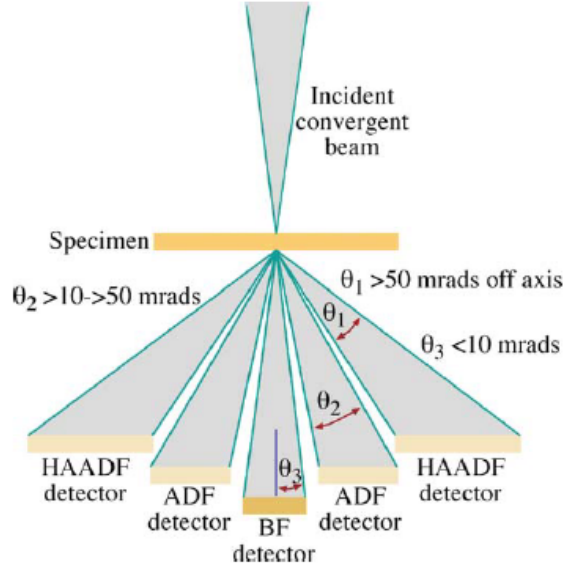


Figure 1.14: Schematic of the STEM collection angles, with the corresponding conventional angular range values associated to each imaging mode. Adapted from [2].

- $\theta < \alpha$ for Bright Field imaging
- $\alpha < \theta \lesssim \alpha + 40 \text{ mrad}$, for Annular Dark Field (ADF) imaging
- $\theta \gtrsim \alpha + 40 \text{ mrad}$, for High-Angle Annular Dark Field (HAADF) imaging

Particularly, STEM HAADF imaging uses only the strongly scattered electrons at high angle, that are mostly elastically scattered. The contrast in the HAADF signal is mainly dependent on mass-thickness contrast [2][18][8][22], and thus also on the atomic number Z of the atoms that form the sample. There is an extensive literature [19][20][21][22] showing how STEM HAADF imaging can be described as *incoherent*, and how that leads to the direct interpretability of peaks in atomic resolution HAADF images as atoms or atomic columns. It has been shown how, for thin samples (few tens of nm), the exit wavefunction from the sample can be factorized in a probe-dependent term and a sample-dependent term. Since the same probe is used to scan all the points that make up an image, any structures that appear are directly associated with the sample-dependent term.

The factorization is possible thanks to the fact that, when a crystalline sample is put in zone axis, or in a near-zone axis configuration, the impinging electrons experience a channeling effect from the electrostatic potential of the atomic columns. Particularly, in the s-state approach [23][24], this effect is attributed to the lowest-energy bound eigenstate of the two-dimensional projected potential of the atomic columns, and it remains valid even in the presence of defects or grain boundaries. In fact, in that probe-sample geometry it is only the highly-localized s-type Bloch states that contribute significantly to the intensity at the atom sites, which is where the scattering to the high-angle detector takes place. The importance of these aspects of HAADF imaging will be further discussed in Subsection 7.2. For now, however, it can be said

that after these considerations, STEM HAADF imaging can be seen as an ideal technique to look at amplitude samples, that mostly affect the amplitude of the probe's wavefunction.

The idea of selecting a certain annulus of the diffraction of a sample begs the question of how this is physically done. Going back to Figure 1.10 of Subsection 1.4, it appears clear that the sample and the annular detector needed for STEM HAADF imaging are separated by the entire block of lenses that form the imaging system. All those lenses provide the perfect setup to tune the collection angles θ of the HAADF detector to a certain range. With the microscope set up for STEM HAADF imaging, the sample's plane is conjugate to the detector's plane, where the diffraction pattern becomes visible. The HAADF detector has fixed inner and outer radii, and it exposes a projected annulus of beam-sensitive material to the impinging electrons. The material produces a current that is picked up by the detector's electronics and amplified before being read by the microscope to make a digital image.

So, the role of the imaging system in STEM is to propagate the beam at an effective Camera Length CL , the distance between the sample and the screen, that can be tuned arbitrarily between a set of discrete values. Changing the camera length changes the collection angles subtended by the detector, with typical values that range between $CL = 100 - 150 \text{ mm}$ for inner collection angles $\theta \sim 50 \text{ mrad}$. It is important to note that in this case the HAADF detector is working as a single-pixel detector: for each point in the scan it is only measuring the integrated intensity. Thus, the presence of aberrations in the imaging system has little relevance for STEM, while the aberrations of the illumination system drastically affect the resolution because they can blur the probe and limit the minimum scanning step.

1.8 Pixelated Detectors

In the context of this PhD work it is necessary to introduce another class of detectors: pixelated detectors. The term *pixelated* refers to the fact that these detectors have more than a single detecting element, combined in an array. Some detectors can be just split into four different quadrants, for instance, while others have square or rectangular sensors with thousands of pixels on each side, to make arrays with tens of thousands of pixels. In this last group we find Charge Coupled Device (CCD) and Monolithic Active Pixel (MAP) cameras, devices built with arrays of sensing elements, that can detect incident electrons. The detection mechanism can be *indirect* or *direct*, depending on the design of the sensing elements. Indirect detection is based on the use of a scintillator, a material that produces photons from the interaction with the incident electrons. The photons emitted by the scintillator are then detected by a sensor array of pixels coupled to the film by an array of optical fibers. So, the incident particles trigger the emission of photons that are then carried to the sensor, which measures them and builds up charge stored in a dedicated potential well for each pixel. The need for a scintillator to mediate the measurement process motivates the term *indirect* detection. Direct detection systems instead can directly measure the incident electrons, without the mediation of the scintillator. In both kinds of detector, each pixel's side is typically only a few μm , and detectors developed for TEM usually have a resolution up to 4kx4k pixels. CCD cameras belong to the class of indirect detection

systems, while MAP cameras are based on CMOS technology and can be either direct or indirect detection systems.

An important difference between CCDs and MAPs is the method used for the readout process, with CCD sensors that are designed to read out the charge stored in each line of the image array sequentially, while MAPs have dedicated electronics for the individual readout of each pixel. This causes difference in the performance and speed of the two kinds of detectors, with MAPs being faster thanks to the parallel readout allowed by the individual electronics of each pixel. A general takeaway is that CCD technology used to dominate the market, but thanks to major improvements in the design of MAPs, nowadays they are widely used in high-end applications, like the production of fast, high-resolution TEM cameras. In most TEM applications there is no need to make high-framerate imaging, however some fields like Ultrafast-TEM or in-situ TEM can greatly benefit from the use of ultrafast cameras to capture the evolution of the system under study.

For completeness, here are listed some general properties of a detector camera:

- *Number of pixels* - ideally, the sensor should have a large number of pixels to get a high resolution image of the sample or diffraction pattern that is being measured. Modern cameras typically have a resolution between 1024x1024 and 4kx4k pixels, with the option of capturing images only from a sub-area of the detector or to rebin it, using the full area of the sensor with larger pixels obtained by combining the signal from 2x2, 3x3, 4x4, ... pixel clusters.
- *Noise* - noise in the detector can come from different sources: the sensor readout noise, when some charge is left in the sensor after the read; thermal or dark noise, that is seen even when the sensor is not exposed to the beam, and rises from heat-generated ionization of the semiconductor layer, reduced by cooling; and finally, shot noise, that is caused by the variability of the signal itself.
- *Detection Quantum Efficiency (DQE)* - DQE is a parameter that varies between 0 and 1, and it represents the fraction of the incoming signal that is effectively detected. A detector with a DQE of 1 would be able to perfectly record all the information encoded in the signal to measure, but in practice all detectors have a DQE smaller than 1. A good detector can have a DQE around 0.5. If the detector is linear in its response, the DQE can be expressed as [2]:

$$DQE = \frac{\left(\frac{S_{out}}{N_{out}}\right)^2}{\left(\frac{S_{in}}{N_{in}}\right)^2} \quad (1.21)$$

if it is perfect, it preserves the original signal-to-noise ratio, otherwise it reduces it, losing part of the information.

- *Dynamic range* - The dynamic range represents the detector's range of recording and is the ratio between the largest and smallest signals that it can record. Ideally, a good TEM detector should have a large dynamic range, because even for the

recording of a simple diffraction pattern there is a very large difference between the intensity of the central spot and the fainter outer diffraction spots.

- *Uniformity* - a good detector should respond linearly to changes in input signal and show a fairly uniform response across many pixels.

1.9 Off-axis Electron Holography

The characterization of one fundamental component of the experimental setup required to make a quantitative measurement of the phase shift of an electron wave that interacted with it, as reported later in Section 9. There are a number of methods that can recover the phase of a sample, like Ptychography, or the Transport of Intensity Equation. Among them, one technique that is both relatively easy to set up and at the same time provides quantitative measurements of the phase is Electron Holography (EH). It was proposed for the first time in 1948-49 [25][26], and it was introduced as a method to achieve high-resolution imaging. Electron Holography takes its name from the use of electron holograms, that are interference patterns between two electron waves. After its introduction in the 1950s, EH was steadily developed, and in the 1990s it counted twenty different sub-varieties [27], some experimentally demonstrated and some only theoretically predicted. Nowadays, it is a well established technique used for imaging at medium and high resolution. It is used to study materials and their defects, but also the electrostatic and magnetostatic fields generated by a sample or in vacuum.

Of the many possible varieties of Electron Holography, the one of interest for this thesis is the Off-Axis in-focus TEM Electron Holography, mostly just referred to as Off-Axis Electron Holography [28]. As the name suggests, it is realized in TEM mode, so with a parallel illumination of the sample. The core component of the microscope needed for this technique is called *electron biprism*, and is usually just a very fine ($< 1 \mu\text{m}$) metallic wire, positioned in place of one of the Selected Area apertures in the Imaging System (see Subsection 1.4). The biprism is placed on the path of the beam and biased positively. As the electrons pass near the wire, they are attracted towards the region at higher potential. In this way, the biprism is able to split the wavefront in two separate waves and at the same time it brings them to interfere further down the column. If the illumination is coherent, in the absence of a sample, the resulting interference pattern is simply a set of fringes. The interference pattern is called hologram, and the holographic interference fringes that appear in it have a certain spacing that depends on the applied bias voltage.

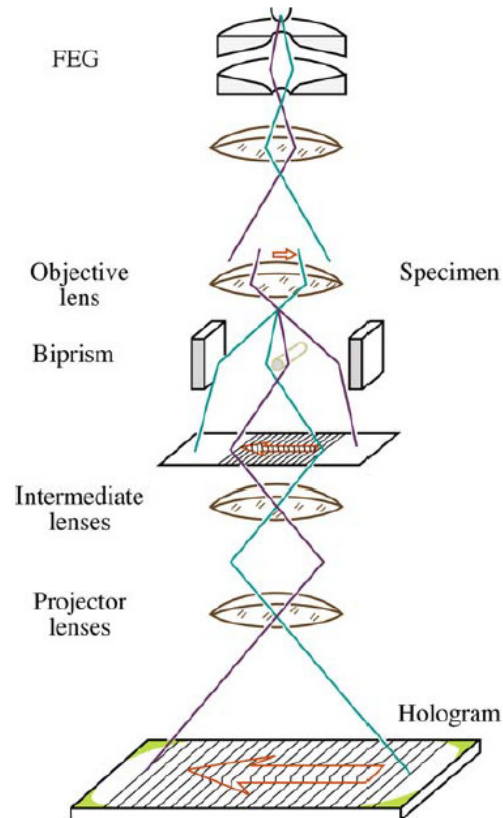


Figure 1.15: Diagram of the Off-Axis Electron Holography method. Notice how the biprism brings the electron wave on either of its sides to interfere with the other. Adapted from [2].

The 'off-axis' attribute of this holography method's name can be better understood when considering what happens when a sample is inserted under the beam. The sample should be positioned in a way that occupies the region roughly on one side of the

biprism. As the electron wave reaches the sample plane, half of it will pass through vacuum, and its phase and amplitude will remain unaltered. This is called the *reference* wave. The other half will cross the sample, being modulated by it both in phase and amplitude. As it carries information about the sample, it is now called the object wave. When the two waves are brought to interfere to form the hologram, the object and reference wave will be at an angle with respect to each other, as their propagation direction has been altered by the biprism. So, the object wave sees the reference wave coming in as if from an "off-axis" direction at the image plane. The fact that the two waves interfere at an angle with each other is the crucial step for the entire procedure. To understand why, it's best to go over the equations that describe this phenomenon.

Because of the coherent illumination, the hologram wave function in the image plane can be written as the sum of the reference and object waves:

$$\psi_{holo}(\vec{\rho}) = \psi_{ref}(\vec{\rho}) + \psi_{obj}(\vec{\rho}) \quad (1.22)$$

where $\vec{\rho} = (x, y)$ are the transverse coordinates in the image plane. The resulting intensity seen in the image will be given by its square modulus, according to Equation 1.5:

$$I_{holo}(\vec{\rho}) = |\psi_{ref}(\vec{\rho})|^2 + |\psi_{obj}(\vec{\rho})|^2 + 2|\psi_{ref}(\vec{\rho})||\psi_{obj}(\vec{\rho})| \cos(\varphi_{obj}(\vec{\rho}) + \varphi_{ref}(\vec{\rho})) \quad (1.23)$$

If the illumination is parallel, then the reference wave is just a plane wave with amplitude 1. Its phase can be written to account for the angle between itself and the object wave, and only has to effectively include a tilt term in the form of a 2D reciprocal space vector $\vec{q}_c$:

$$\psi_{ref}(\vec{\rho}) = \exp[2\pi i \vec{q}_c \cdot \vec{\rho}] \quad (1.24)$$

and substituting in Equation 1.23 gives:

$$I_{holo}(\vec{\rho}) = 1 + |\psi_{obj}(\vec{\rho})|^2 + 2|\psi_{obj}(\vec{\rho})| \cos(\varphi_{obj}(\vec{\rho}) + 2\pi i \vec{q}_c \cdot \vec{\rho}) \quad (1.25)$$

where in the last part of the formula the spatial dependencies were omitted for brevity. Now, the final step: if the intensity in the image plane is Fourier transformed, the presence of the tilted reference wave causes the phase and amplitude information of the object wave to be shifted from the central peak in reciprocal space. That is because:

$$\begin{aligned} \mathcal{F}[I_{holo}(\vec{\rho})] &= \delta(\vec{q}) + \mathcal{F}[|\psi_{obj}(\vec{\rho})|^2] + \\ &\delta(\vec{q} + \vec{q}_c) \otimes \mathcal{F}[|\psi_{obj}(\vec{\rho})| \exp(i\varphi_{obj}(\vec{\rho}))] + \\ &\delta(\vec{q} - \vec{q}_c) \otimes \mathcal{F}[|\psi_{obj}(\vec{\rho})| \exp(-i\varphi_{obj}(\vec{\rho}))] \end{aligned} \quad (1.26)$$

So, at the center of the Fourier transform there is a peak carrying information about the amplitude of the object function and the reference image, i.e. the TEM-BF image of the sample. Then, there are two peaks centered at the positions marked by the reciprocal vectors $\vec{q} = +\vec{q}_c$ and $\vec{q} = -\vec{q}_c$, called *sidebands*, that carry information about the object wave function and its complex conjugate. The extraction of the phase information about the wave function can be done by digitally selecting only one of the sidebands with an aperture, shifting it back to the $\vec{q} = 0$ position, and inverse-FT the

image. With the resulting complex image, the phase is just the arctangent of the ratio of its imaginary and real parts:

$$\varphi_{obj}(\vec{\rho}) = \arctan \left(\frac{\Im(\mathcal{F}^{-1}[SB])}{\Re(\mathcal{F}^{-1}[SB])} \right) \quad (1.27)$$

where SB is a filtered sideband. The size and smoothness of the digital aperture filtering function determines the level of detail and amount of artifacts that can appear in the reconstructed phase image. phase images produced in this way might exhibit discontinuities if the phase shift exceeds 2π . A class of algorithms, called *unwrapping* algorithms, has been developed to solve this issue and recover a smooth phase landscape with an *unwrapped* phase defined over a range that exceeds the usual $[0, 2\pi]$. Finally, if there is the possibility of a perturbation being present in the vacuum reference wave itself, it is possible to measure a reference hologram of the vacuum, by bringing the sample far away from the interference region. The resulting vacuum reference hologram can be processed in the very same way as the sample hologram, selecting the same sideband and using the same exact filtering function. Then, before the calculation of the phase, the inverse-FT of the sample hologram can be normalized to the inverse-FT of the vacuum reference hologram. That additional step removes the effect of vacuum perturbations. In that case, the equation for the phase becomes:

$$\varphi_{obj}(\vec{\rho}) = \arctan \left(\frac{\Im(\mathcal{F}^{-1}[SB]/\mathcal{F}^{-1}[VRSB])}{\Re(\mathcal{F}^{-1}[SB]/\mathcal{F}^{-1}[VRSB])} \right) \quad (1.28)$$

where $VRSB$ is the additional vacuum reference hologram's filtered sideband.

1.10 Ptychography

As it was mentioned while introducing Holography, Ptychography is another phase imaging method that is available in electron microscopy, and it is currently quite popular. Just like electron Holography, Ptychography is a kind of computational imaging method, because it does not directly produce an image of the sample, but instead recovers it later by processing data recorded during the measurement session.

Ptychography produces complex images of the object function of a sample, recovering both its amplitude and phase. The more recent algorithms for Ptychography solve simultaneously for the amplitude and phase of the probe used for the measurement as well, characterizing both the illuminating probe and the sample. The structure of the equations at the heart of the technique are similar to the ones for the imaging technique used for this thesis. Thus, Ptychography is introduced in this Section, and its application to the topic of this thesis is further described in Subsection 5.3.

The experimental procedure involves scanning some probe with limited support on the sample, and measuring the corresponding diffraction pattern in the near or far field. The limited support probe can be made either by scanning an aperture in close proximity of the sample or, as it is more standard in a TEM, by scanning a slightly defocused convergent probe on the sample. It is important that the probe is extended and not point-like, so that the illuminated areas at different points overlap partially. Ptychography exploits the overlap between adjacent regions to stitch them together, imposing constraints on the recovered object function for each position. So, the scan-

ning is a fundamental element of the technique. Figure 1.16 shows a diagram of the object function update process.

The procedure described in Figure 1.16 is the most basic version of the Ptychographical Iterative Engine (PIE) algorithm [29]. It yields a solution for the phase of the object wave, that is lost when the diffraction pattern is recorded, as per Equation 1.5. The reconstruction of the sample's object function is carried out by repeatedly alternating between real and Fourier space, imposing constraints on the sample illumination and to its diffraction. The size of the field of view for the object function at a given position is determined by the resolution of the recorded diffraction patterns and the probe size.

However, it is possible to store an arbitrarily large total reconstruction of the area of interest by stitching together parts of the object function recovered from adjacent regions. The stitching procedure requires the adjacent regions to partially overlap each other, and this also means that the object's phase and amplitude recovered at one position are constrained to be consistent with all those recovered at adjacent overlapping regions. The stitching of adjacent regions at each step is not done by brutally substituting a cutout of the object function into the larger composite image. Rather, the estimate of the object function is used to alter the composite image. So, as the calculations go on, the phase and amplitude of the object function smoothly converge to a solution that is compatible with all the recorded diffractions.

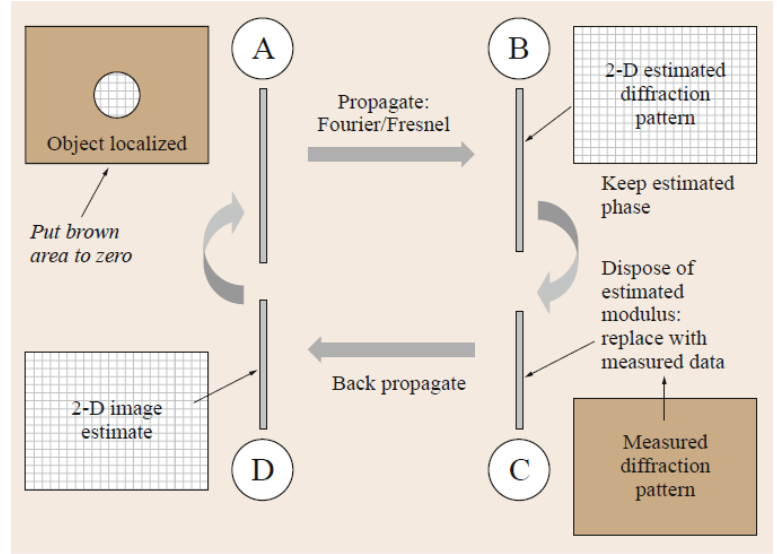


Figure 1.16: Diagram of the object function update process. (A) real space portion of the object illuminated by the probe at one of the scanning points. The area outside of the probe support is put to zero, imposing a real-space amplitude constraint. (B) the wave function is propagated to the diffraction plane of the detector, giving an estimate of the diffraction pattern. (C) the measured diffraction pattern is used to impose a Fourier space constraint on the modulus of the object function, while the estimated phase is left untouched. (D) back-propagation to the sample plane gives an estimate of the object function, that is used again in (A). Iterating the propagation between real and Fourier space makes the phase converge to a possible solution. Adapted from [28].

As the technique developed over time, the PIE algorithm branched into a family of Ptychography algorithms. In particular, Maiden introduced the *extended* PIE (ePIE) algorithm [30], an improved version of the preexisting PIE algorithm, to include the reconstruction of the probe in parallel with that of the object function. To find the link between our modified version of the AP algorithm and ePIE, it is necessary to provide a brief description of the equations behind it. The object and probe functions

are called $O(\vec{r})$ and $P(\vec{r})$, with $\vec{r}$ being the real space coordinates. The ePIE method assumes that the interaction of probe and object can be modeled by a simple product between their complex functions. The experimental measurements produce a set of J diffraction patterns, that are simply the recorded intensities of the exit wave from the sample, propagated to the detector plane:

$$I_j(\vec{u}) = |\mathcal{F}[O(\vec{r})P(\vec{r} - \vec{R}_{s(j)})]|^2 \quad (1.29)$$

where $\vec{u}$ are the Fourier-space coordinates and j is the index representing the j -th diffraction pattern, while $\vec{R}_{s(j)}$ is the shift between probe and sample introduced to record the j -th diffraction. The initial guess of the exit wave from the sample at a given scan position is then given as:

$$\psi_j(\vec{r}) = O_j(\vec{r})P(\vec{r} - \vec{R}_{s(j)}) \quad (1.30)$$

This initial guess is propagated to the Fourier space, where the amplitude of its Fourier transform is updated with the recorded square root of the detected diffraction:

$$\Psi_j(\vec{u}) = \sqrt{I_{s(j)}(\vec{u})} \frac{\mathcal{F}[\psi_j(\vec{r})]}{|\mathcal{F}[\psi_j(\vec{r})]|} \quad (1.31)$$

The updated exit function is then computed by backpropagating that function to the sample plane:

$$\psi'_j(\vec{r}) = \mathcal{F}^{-1}[\Psi_j(\vec{u})] \quad (1.32)$$

That updated exit wave is then used to compute an update of the object function:

$$O_{j+1}(\vec{r}) = O_j(\vec{r}) + \alpha \frac{P^*(\vec{r} - \vec{R}_{s(j)})}{|P(\vec{r} - \vec{R}_{s(j)})|_{max}^2} (\psi'_j(\vec{r}) - \psi_j(\vec{r})) \quad (1.33)$$

and similarly, the probe can also be updated as:

$$P_{j+1}(\vec{r}) = P_j(\vec{r}) + \beta \frac{O^*(\vec{r} + \vec{R}_{s(j)})}{|O(\vec{r} + \vec{R}_{s(j)})|_{max}^2} (\psi'_j(\vec{r}) - \psi_j(\vec{r})) \quad (1.34)$$

In Equations 5.6 and 5.7, the $*$ stands for the complex conjugate, as the object and probe functions are complex. The coefficients α and β can be interpreted as learning rates, parameters that determine the strength of the update steps. The ePIE algorithm is usually started with a flat amplitude and phase representing vacuum for the initial guess of the sample, and with a simulated or experimentally measured guess of the probe. The above mentioned method is applied at all the J positions of the raster scan to complete one iteration, and then starts back from the beginning. As the number of iterations over the data set increases, the recovered complex functions for the sample and probe converge to the correct result.

2 Ghost Imaging

This section covers the main technique of interest for this PhD work, Ghost Imaging, giving an overview of its origin and evolution as a light optics imaging method and some of its applications. The term *Ghost Imaging* refers to a transmission imaging technique first proposed in the field of light optics. Usually, when an image is formed through a transmission-based system, the probe is prepared in a specific state at the sample plane. There, it interacts with the sample and propagates through it. After that, it is collected by some detection system. In the case of the first design of ghost imaging experiments, this very simple sequence of events was broken down, and the image of the sample was produced using probes that had never directly interacted with the object to see. Thus, it was as if a *ghost* of the probe had interacted with the sample, producing a *ghost image* of it. The phenomenon behind this peculiar imaging technique was actually quantum entanglement, and the experiment exploited the correlation of the pairs, measuring the particle that had not interacted with the sample.

2.1 Origin of the Technique: Quantum Ghost Imaging

In fact, the first proposed experimental setup [31] was based on the use of quantum entanglement between photons. The original paper from 1995 [31] does not actually call the technique *ghost imaging*, that name was first set further down the line. Figure 2.1 shows the original experimental setup, with a synthetic description of the experiment:

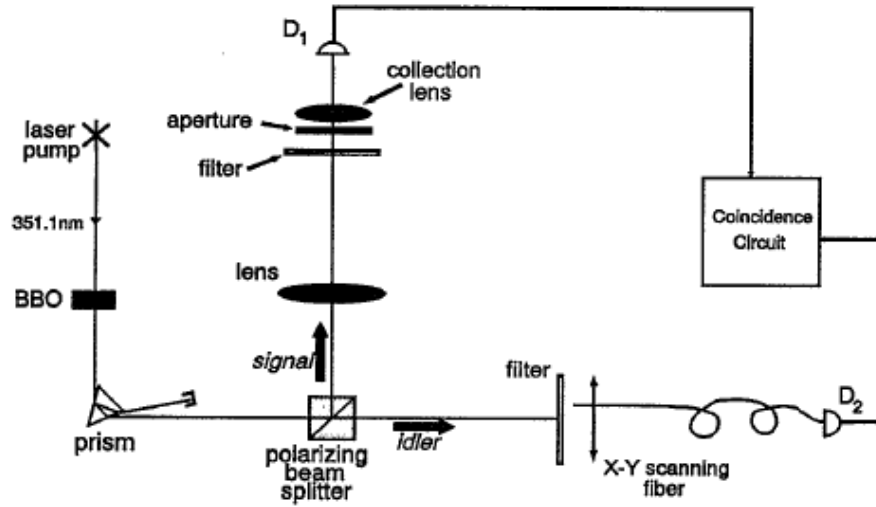


Figure 2.1: Schematic of the experimental setup of the first Ghost Imaging experiment. A laser is used to excite entangled photons in orthogonally polarized pairs called *signal* and *idler* from a Beta Barium Borate (BBO) crystal. A polarizing beam splitter sends the two photons onto two different optical lines, one with the sample aperture and one with a scanning fiber tip. Both lines end on photodiodes D_1 and D_2 , where the first acts as a bucket detector and the second as a pixelated detector, retrieving the idler photon position. Recording the coincidence counts on the signal line as a function of the fiber tip's x-y position in the idler line recovers the image of the aperture. Adapted from [31].

This experiment showed how using the spatial and temporal correlation of entangled photons would allow to reconstruct a sample by measuring particles that had never

interacted with it in the first place. There was also a magnification effect of the reconstructed image thanks to the lens placed on the signal line: the relationship between the distance of the aperture from the lens, the distance of the fiber tip's scan plane from the lens, and the lens's focal obeyed the thin lens equation [10]. However, later theoretical [32][33] and experimental [34][35] works that used non-entangled particles showed how it was possible to relax this particular requirement, maintaining a similar setup but using classical fields instead.

2.2 Classical Ghost Imaging

The setup proposed in [32][33][34][35] used pseudothermal light, that can be described in terms of classical electromagnetic fields. Figure 2.2 schematizes a possible classical (non-quantum) version of the experiment and provides a description of the principal elements.

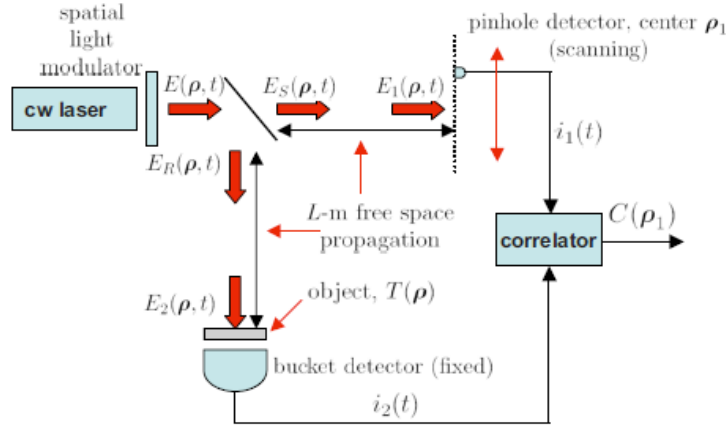


Figure 2.2: Diagram of the experimental setup for classical Ghost Imaging. A continuous-wave (cw) laser source emits a field that is modulated by a Spatial Light Modulator (SLM) into a wave $E(\rho, t)$, ρ being the transverse spatial coordinates perpendicular to the optical axis. As in the other setup, this field is then split on two lines, a *signal* line and an *reference* line using a beam splitter. The fields propagate in free space until the sample and pinhole detector planes, respectively. On the signal line, the field interacts with a sample object of transfer function $T(\rho)$. The signal line ends on a bucket detector, while the reference line ends up on a scanning pinhole detector. Correlating the current $i_1(t)$ from the pinhole detector, that is scanned in plane, with the current $i_2(t)$ from the bucket detector allows reconstructing the sample's transmission function $T(\rho)$. Adapted from [36].

In this case, the source is a laser, modulated by a Spatial Light Modulator (SLM). An SLM is a device that can impart an amplitude, phase, and/or polarization modulation to an incident light beam. SLMs are usually made with micromirror arrays or liquid crystals, and can be programmed to impart an almost-arbitrary modulation. The advantage of using an SLM instead of single photons is that the acquisition process becomes easier, and the SLM can produce very stable different field landscapes. Then, in a nutshell, what happens is that on the reference path the high-resolution pinhole detector measures the field as it is before its interaction with the sample. On the signal line, the bucket detector measures the result of the interaction. Computing

the cross-correlation between the bucket detector signal and the photodiode signal at every point reveals the sample's transmission function $T(\rho)$.

There are a couple aspects worth pointing out: first, as mentioned before, the only requirement for the field is that after the splitting, the signal and reference fields maintain a high spatial correlation. This is mandatory, because it is demanded by the correlative nature of the measurement. The second notable thing is that the resolution of this imaging method is in part limited by the scan step of the pinhole detector, that is supplying the spatial information. This issue can be better understood by considering another implementation, where the pinhole detector is substituted by a high-resolution pixelated detector that measures the entire reference field at once. In that case, a snapshot of the field can be correlated with a measured intensity on the signal line, and the transmission function can be recovered through inverse methods (a more detailed explanation is given in Section 5). The resolution of the image of the field is then necessarily dictating the final resolution of the reconstructed transmission function because of the finite pixel size.

2.3 Computational Ghost Imaging

In his 2008 article [36], Shapiro showed how it would be possible to completely remove the reference field measurement part from the classical Ghost Imaging setup. Figure 2.3 shows the proposed modified setup:

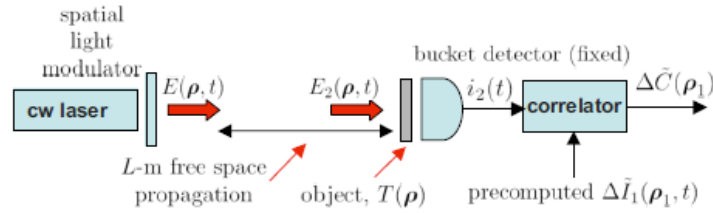


Figure 2.3: Diagram of the experimental setup for computational Ghost Imaging. A continuous-wave (cw) laser source emits a field that is modulated by a Spatial Light Modulator (SLM) into a wave $E(\rho, t)$, ρ being the transverse spatial coordinates perpendicular to the optical axis. The field propagates in free space until the sample plane. There, it interacts with a sample object of transfer function $T(\rho)$. The signal line ends on a bucket detector that measures a current $i_2(t)$. The field at the object plane is precomputed to the desired resolution, and is then used for the calculation of the transmission function under the assumption that the experimental field matches the simulated one. Adapted from [36].

Effectively, this setup completely eliminates the need for expensive high-resolution detectors, because the field at the sample level is assumed to be perfectly controlled by the SLM. This assumption is quite realistic for light optics, where Spatial Light Modulators allow very accurate encoding of arbitrary wavefronts in the probe [37]. Then, in the experiment, the SLM can be used to illuminate the sample with any structured probe, including the ones whose analytical form is known and that can be simulated precisely. As the SLM controls both the amplitude and the phase of the modulated field, the probes are known not only in the SLM plane but also in their free space propagation to the sample plane, where it is possible to calculate their

intensity. The structured probes are stored in the computer memory, and the imaging method is now based on correlating the spatially-resolved *computed* intensity images of the probes with the corresponding total transmitted intensities measured at the bucket detector [36][38]. This new take on the technique was then called *Computational Ghost Imaging*. Notice how, if the sample is perfectly transparent everywhere and only changes the phase of the probes, no information can be retrieved. The quantity measured by the detector is the total transmitted intensity, that is proportional to the square modulus of the field, $I_T \propto |E(\rho, t)|^2$, and so the phase of the field is lost. If the sample only has a phase effect, the total transmitted intensity is equal to the total incident intensity for each probe, and the signal doesn't carry any information about the sample's transmission function $T(\rho)$. So, in this form, Computational GI is inherently an Amplitude imaging technique, because it is only sensitive to the sample's amplitude. There have however been attempts to make the technique also phase-sensitive, within the quantum entanglement framework, as reported in [39], and using Transport of Intensity Equation [40], but they will not be discussed here.

2.4 Applications in Optics and Single-Pixel Imaging

Conceptually, Computational Ghost Imaging is analogous to another technique from the field of computational imaging, called Single-Pixel Imaging (SPI) [38]. In Computational Ghost Imaging, the illumination is structured by an SLM and the detection occurs directly after the wave is transmitted by the sample. In Single-Pixel Imaging, instead, the object is illuminated with a non-structured probe and the SLM is positioned after the sample, to structure the transmitted intensity before the detection. Formally, the two methodologies are equivalent, and are faced with the same challenges of accurate beam structuring and signal measurement. It is interesting to quickly survey the advancements in both fields, to see that these techniques are well established in light optics.

Computational GI and SPI have complementary schemes, where the first uses structured *illumination* and the second uses structured *detection*. In either case, the goal is to reconstruct an image of the sample comprising N pixels, with a series of M different structured beams. Keeping the Computational GI framework, a useful assumption is to consider the measurement process as linear. Then, the signal measured at the bucket detector for the m -th structured beam, S_m , is simply obtained by the total integrated intensity of the product of the m -th structured probe $P_{(x,y),m}$ and the object's transmission function $O_{(x,y)}$. Then, using M different probes, the 2D estimate of the object's transmission function can be recovered as [41]:

$$O_{(x,y),M} = \frac{1}{M} \sum_{m=1}^M S_m P_{(x,y),m} \quad (2.1)$$

Equation 2.1 represents the simplest kind of Computational GI/SPI algorithm, reconstructing the object's transmission function as a weighted sum of the structured beams, with weights given by the intensity of the corresponding signals at the bucket detector. If a structured probe is well transmitted (i.e. it mostly illuminates transparent regions of the sample) then it will be weighted by a larger signal. Vice-versa, a probe that is poorly transmitted will be weighted by a smaller coefficient. Summing all the weighted probes creates contrast in the image that reflects the sample's transmission

function. Figure 2.4 shows both the measurement process and the above-mentioned image reconstruction algorithm. A more detailed discussion of the SPI algorithms will be given in Section 5, but Equation 2.1 already presents the key elements needed to carry on the analysis here.

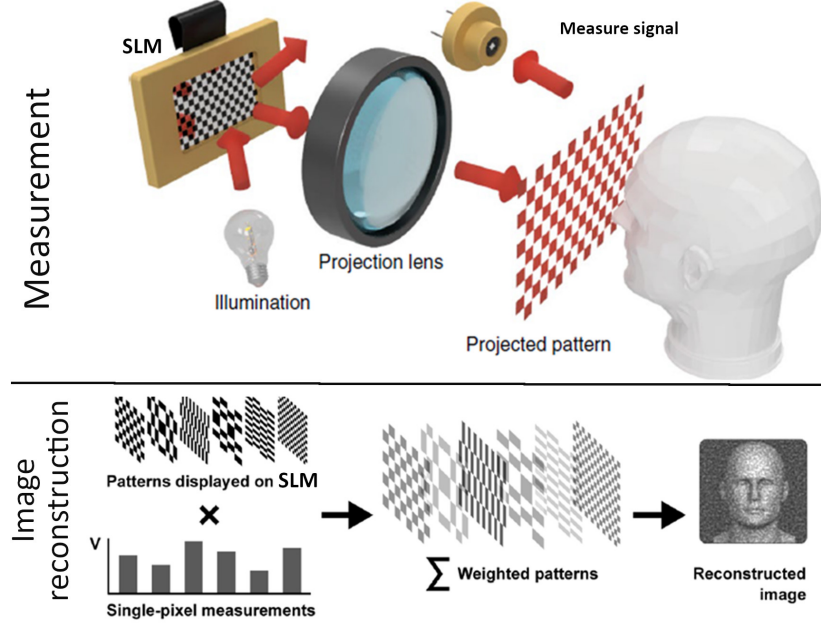


Figure 2.4: Essential schematic of a Computational GI experiment and the simplest image reconstruction algorithm. The measurement process is in reflection geometry, but the principle is equivalent to the transmission geometry previously explained in Subsection 2.3. The image reconstruction process is that of Equation 2.1. Adapted from [41].

The optimal goal is to fully reconstruct the sample’s image by determining all of its N pixel values correctly. To do that, one needs either $M = N$ different *orthogonal* probes, with a null overlap for each pair, or $M \gg N$ different non-orthogonal probes, to get a Signal-to-Noise ratio $SNR \gg 1$ [42]. Since the introduction of Computational GI and SPI, different strategies have been employed to reduce the number of different probes needed, optimize their ordering to be more dose-effective while illuminating the sample, and to improve the reconstruction algorithms. Also, there has been testing on the different kinds of beam modulation available.

Some of the first efforts towards improving Computational GI were devoted to including *Compressed Sensing* (CS) into the image reconstruction algorithms. The idea behind compressed sensing, also called compressive sampling [43][44], is that natural images like those obtained by imaging a real scene are usually sparse in the spatial frequency domain, and can be reconstructed by using only a few coefficients in the appropriate representation basis. In [45] is found an example of the use of CS for a Computational GI setup, with results yielding 3.5 times larger SNR for the mixed CS-GI approach compared to the standard Computational GI. The results were encouraging, but the reconstructed images were only a few tens of pixels on each side, and the computational cost of CS grows rapidly with the number of pixels in the target image. As time went on and machine learning started becoming more mainstream and appealing, several groups dedicated some time to integrating different kinds of Neural Networks in the

image reconstruction process [46][47][48][49] and even to make SPI videos at a good framerate [50] of 30 fps.

Another interesting case study is presented in [51], where the focus was put on optimizing the measurement by finding an ideal ordering of the Hadamard matrix basis [52] patterns that would lead to a good reconstruction of the sample with the least number of different probes. In the paper, it was found that switching the order of the Hadamard basis matrices to a specifically designed one allowed recovering a 128x128 image with a sampling ratio N_{probes}/N_{pixels} of 5% at the signal peak signal-to-noise ratio (PSNR) of 23.8 dB, reducing significantly the total number of required patterns for the reconstruction while improving the image quality.

Finally, a quick remark about the different methods to produce structured illumination. Programmable Spatial Light Modulators, where the structured illumination can be fully controlled, are mostly based on Liquid Crystal technology or Digital Micromirror Devices. However, these technologies might not be available at all scales and wavelengths, so in some cases it might be more useful to resort to speckle patterns. A deterministic way to do this is described in [53], where a slowly varying phase mask is used to produce caustic patterns in the far field of the SLM. Figure 2.5 shows the results obtained using caustic patterns, both in simulations and in the real experiment.

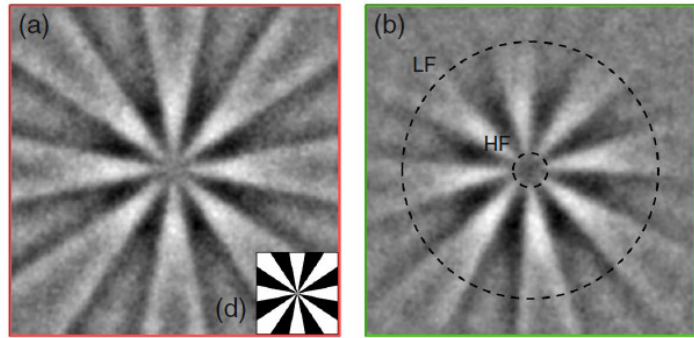


Figure 2.5: Simulated (left) and experimental (right) comparison of a Siemens star target (inset) reconstructed with 30.000 caustic patterns at a 768x768 pixels resolution. LF and HF dotted lines mark the Low- and High-Frequency limits of the reconstruction, where the contrast falls below $1/e^2$. Adapted from [53].

2.5 Motivation to make Ghost Imaging in Electron Microscopy

In principle, even the STEM mode of a TEM can be seen as a kind of Computational GI. Conceptually, making a square STEM image of $N \times N$ pixels is the equivalent of making a Computational GI measurement with N^2 different convergent probes. The illumination system of the microscope is producing a point-like probe that is scanned over a raster, and for each point of the grid the BF (or HAADF) detector is measuring a single value of transmitted intensity. In this case, each pattern can be thought as being zero everywhere except in the pixel where the probe is, where it has a value of one. The detector signal is weighting the pixel value, and to reconstruct a full image one needs exactly $N \times N$, because all of the patterns are perfectly orthogonal to each other.

Following this premise, it is possible to imagine that, with an appropriate device capable of producing structured illumination for electrons, it would be possible to draw inspi-

ration from the developments of Ghost Imaging in light optics and adapt the technique for electron microscopy. A paper from 2018 showed a realization of Ghost Imaging with electrons at the UCLA PEGASUS beamline [54], but so far there has been no development concerning a realization of the technique in a TEM, where the compact structure of the machine makes it difficult to introduce an electron beam modulator in the setup.

As it was shown in Subsection 1.3, in a TEM the aberrations degrade the image resolution in ordinary imaging modes. The most interesting and powerful aspect of Ghost Imaging is that those same aberrations that are detrimental to the standard operation of a TEM could then be incorporated into the probe formation mechanism. Thus, they would no longer be harmful for image formation, but they would simply be another parameter to account for in the generation of the structured illumination, and the structured probes could be generated by using larger convergence angles. Under these circumstances, Ghost Imaging would provide a way to overcome the aberration-limited resolution of uncorrected instruments, simultaneously eliminating the need for expensive high-resolution pixelated detectors and probe correctors [55]. Figure 2.6 shows an example of a simulation that compares the results of an aberration-limited STEM HAADF image with a CGI measurement, illustrating how the latter can access information at spatial frequencies that are precluded to the former.

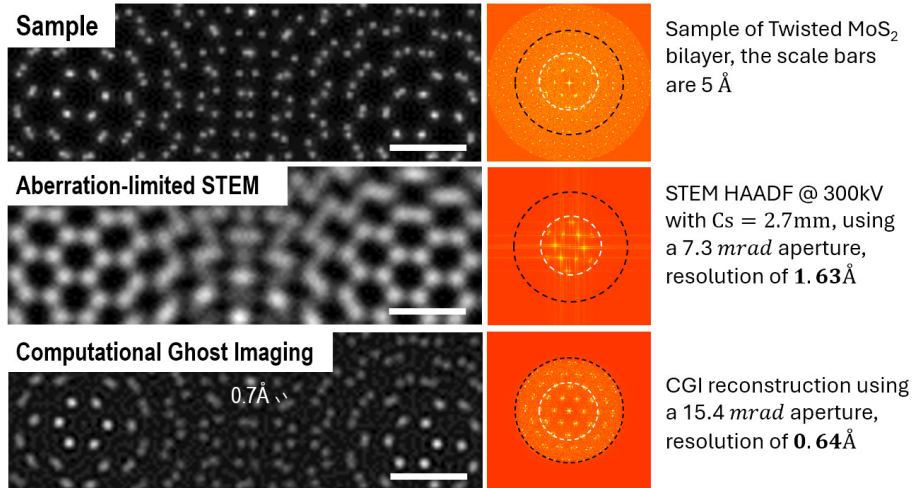


Figure 2.6: Comparison of an aberration-limited STEM HAADF image of a sample and a simulated CGI reconstruction made with structured probes at twice the convergence angle of illumination. The Fourier transform of the images show clearly the difference between the sampled spatial frequencies in both images. Adapted from [55].

Another element to bear in mind is that the framework of Ghost Imaging in optics is valid because there is a linear relationship between the signal S measured on the single-pixel detector and the overlap of the probe with the sample. This can be understood from Equation 2.1, where to reconstruct the transmission function each illumination is weighted by the signal measured by the single pixel detector. The equation only works because:

$$S \propto \int_A I_{ill}(x, y) T_{sam}(x, y) dx dy \quad (2.2)$$

where the signal from the detector is proportional to the overlap between the structured illumination I_{ill} and the sample's transmission function T_{sam} , calculated over an area

A that contains all the regions illuminated by the different probes. In a TEM, the interaction of the electron beam with the sample has several products [2], like secondary electrons, elastically backscattered electrons, X-ray and cathodoluminescence photons, inelastically scattered electrons, ... And each one of them could represent a different source for the signal needed for GI, possibly changing the properties probed with this imaging scheme. The only requirement that has to be satisfied is the linearity of the signal measurement imposed by Equation 2.2. In Section 7, it is shown how the validity of this criterion is verified for our experimental setup.

Finally, it is our opinion that after demonstrating a fully functional setup capable of high-resolution imaging, it would be possible to begin adopting strategies that make use of prior information about the sample, to favor an accurate image reconstruction process. Also, another further development for the technique would be to adopt methods of image compression, to optimize the distribution of the electron dose on the sample and minimize it at the same time. That would make the technique viable also for the imaging of beam-sensitive materials, because the use of structured probes in place of point-like convergent ones could limit the irradiation damage caused by the exposure to the beam. For instance, a pioneering paper of Van Broek [56] has compared single pixel imaging and traditional STEM-ADF in view of standard acquisition and compressed sensing.

3 Structured Electron Beams

This section covers the different methods that have been demonstrated to produce controlled electron beam shaping in TEM, including the theory about the devices used for the experimental work of the PhD.

3.1 Beam Shaping Methods

Section 2 has extensively explained how a key aspect of classical and Computational Ghost Imaging is to be able to produce structured illumination. For light optics, this task is easily solved with the help of Spatial Light Modulators, that are extremely versatile and reliable. SLM have been developed to work in transmission and reflection geometry, and can provide both amplitude and phase modulation. It is instructive to take a look at a Digital Micromirror Device (DMD), a kind of SLM, to deduce the most relevant aspects of this technology in the context of Ghost Imaging. Figure 3.1 shows an example of a DMD and the kind of structured illumination that it can produce.

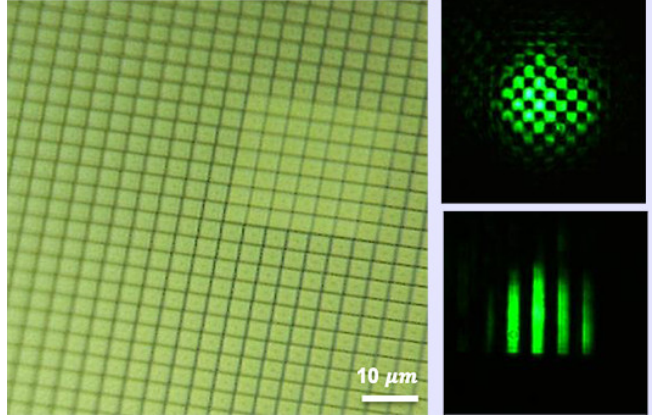


Figure 3.1: Digital Micromirror Device showing the size of the single mirror elements and two examples of structured probes produced with an illuminating green laser. Adapted from [57].

DMDs work with binary amplitude-only modulation. A DMD is essentially an array of micron-sized mirrors, each one working as a pixel that can be turned on or off individually by rotating it. This kind of device can generate the Hadamard matrices as a set of completely orthogonal structured probes. The Hadamard matrices patterns, like the two examples in Figure 3.1, are perfectly known analytically as they are arrays of binary pixels (0 or 1), and can be simulated reliably for the image reconstruction process. Taking DMDs as a reference, one can identify the main aspects that the ideal beam shaping technique to be used when adapting Computational GI to electron microscopy must have. It must be:

- Effective in producing electron wave Amplitude modulation, as GI is inherently an amplitude technique
- Programmable, so that several different amplitude masks can be generated by changing some control parameters
- Reproducible by simulation, so that a computational analog of the structured probes can be simulated

In the case of electron optics, these requirements are not easily met. So far, there has not been any advancement in the field of electron beam shaping that can be compared to the introduction of SLMs in light optics. There is no device equivalent to an SLM for electron beams that offers the same level of control over the illumination. However, the

field is not entirely uncharted, and there are several different methods that have been explored over the years to make structured electron beams with specific properties.

3.1.1 Electron wave functions and weak phase objects

Before diving into the description of all the various beam shaping methods, it is convenient to provide a simple description of the theory behind the manipulation of electron wave functions. As stated in Section 1, electrons can be described both as particles and as waves, and in the second case they can be represented through wave functions. What happens if an electron wave interacts with a sample or an electrostatic or magnetostatic field? In TEM, the samples are usually very thin, and the electrostatic/magnetostatic fields that might be present around the sample are much weaker compared to those imparted by the microscope's optics. For this reason, TEM makes use of the Weak Phase Object approximation, that simplifies the description of the effects of such perturbations on the electron wave function. Within this approximation, for thin enough sample, or in presence of weak electric or magnetic fields, the wave function after the interaction is [58]:

$$\Psi = \Psi_0 \chi(x, y) = \Psi_0 \exp[i \Delta\varphi(x, y)] \quad (3.1)$$

So, the initial wave function Ψ_0 only experiences a phase shift due to the interaction with the perturbing element, while amplitude effects are mostly neglected. The relativistically corrected phase shift can be expressed as [58]:

$$\Delta\varphi(x, y) = \frac{2\pi e}{\lambda} \frac{E_0 + E}{E 2E_0 + E} \int_{Z_i}^{Z_f} V(x, y, z) dz - \frac{e}{\hbar} \int_{Z_i}^{Z_f} A_z(x, y, z) dz \quad (3.2)$$

where the integrals represent the projected electrostatic scalar potential V and the projected z component of the vector potential A_z . In the case of a thin, purely electrostatic and homogeneous sample, it is possible to derive a Mean Inner Potential (MIP) V_{MIP} that models the average effect of the sample on an impinging wave. The projected value of the Mean Inner Potential obtained by integrating along z for the sample's thickness t allows calculating the corresponding phase shift. The effect of electrostatic and magnetostatic fields can also be calculated using this approximation. An unfortunate consequence that is made evident by Equation 3.2 is that *any* source of electrostatic potential can affect the electron wave, including the buildup of charge that can occur because of contamination, or in presence of poorly conductive materials under the beam.

Now that some foundational terminology has been established, it is time to quickly recap the various beam shaping methods that are available. Essentially, as it was explained above, electron beam shaping can be conveniently done by using very thin membranes of materials, or electrostatic or magnetostatic fields. There is also a fourth method whose characterization has been intensifying in the last few years: using light to modulate the electron beam. All of the methods that will be presented in the following subsections are about the realization of some form of electron-optical phase plate, a device that is used to impart a desired phase shift to an impinging electron beam.

3.1.2 Thin-film phase plates

The first kind are the thin-film phase plates, where a thin layer of material is nanofabricated to produce a thickness profile that will impart the desired phase shift to the beam. An example are thin-film Zernike phase plates [59], that are used to enhance the contrast in TEM images by introducing a phase shift that improves the transfer function of the microscope [2][60], particularly for the low frequencies.

The main issue with this kind of phase plate for an application in Ghost Imaging is that they are obviously not programmable. Each thin-film phase plate can encode one specific design for a structured probe. That is normally fine for other applications where a specific structured illumination is required, like for electron holography, but it is not the best solution for GI. Also, this kind of phase plate is easily subject to contami-

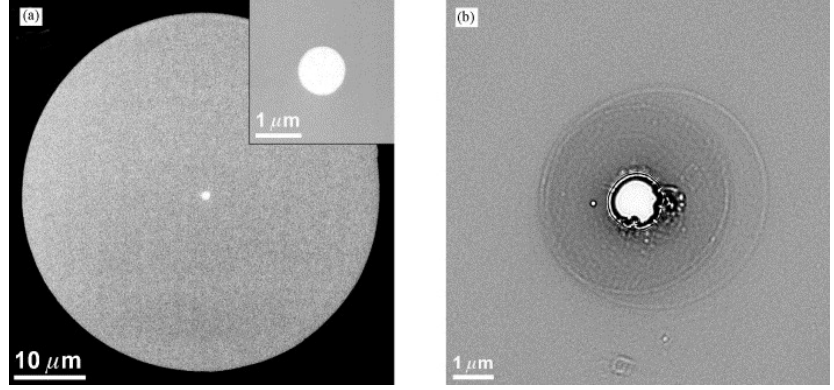


Figure 3.2: (Left) Focused Ion Beam machine of a Zernike phase plate used to mill the central hole. Inset shows the freshly-milled hole. (Right) Detail of the damage and contamination of the thin film caused by the electron beam irradiation, both in focus (specles) and defocused (darker gray area). Adapted from [61].

nation and damage, as seen in Figure 3.2. Charging is an issue as well, because these phase plates are typically made with semiconductor membranes and adding a thick enough metallic layer on top of the device to dissipate the charging would either make the device less transparent to the beam or completely opaque, rendering it less efficient or completely useless.

3.1.3 Electrostatic phase plates

Electrostatic phase plates are similar to thin-film phase plates, because the source of the phase modulation is still the electrostatic potential. However, electrostatic phase plates are made with combinations of one or more electrodes, and the electron beam is mostly (or exclusively) modulated by the electrostatic potential that extends in the vacuum between them. Some designs are meant to be operated with the electrodes exposed to the beam, in which case the device will also necessarily impart an additional amplitude modulation. Figures 3.3 and 3.4 show three examples of electrostatic phase plates.

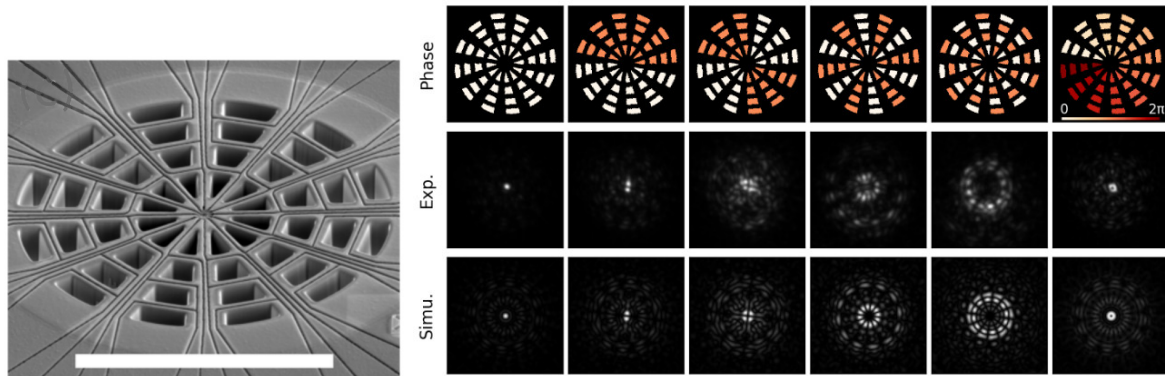


Figure 3.3: Example of a very recently proposed 48-element programmable phase plate for electrons. Left: SEM image of the device, the scalebar is 30 μm . Right: Applied phase shift in $[0, 2\pi]$, marked by the colorbar (Top row), and corresponding experimental (Central row) and simulated (Bottom row) beam intensity modulations. Adapted from [62].

The device in Figure 3.3 is an electrostatic phase plate with 48 elements that can be controlled independently. It was designed to do beam shaping for ptychography [62][28]. This device is quite interesting, as it shows a challenging many-electrode design for a fully programmable phase plate, that can generate a huge number of different structured probes, albeit with a somewhat always radially-symmetric structure. Figure 3.4 instead shows an application of two combined phase plates, that are used to make an angular momentum sorter [63]. In this application, an Electron Vortex Beam (EVB) [17] (an electron beam carrying Orbital Angular Momentum) generated with a thin-film phase plate is modified by two specifically designed phase plates, called Sorters, that perform post-selection of the Orbital Angular Momentum (OAM) states of the beam, mapping them into an OAM spectrum in the final image plane. This application shows how beam shaping is not necessarily limited to using a single device, but it can also be done in multiple steps inside the microscope.

These examples show how electrostatic phase plates are more versatile than thin-films when it comes to programmability, as they can be realized with arrays of elements that can be biased independently from each other. Of course, these phase plates offer a trade-off between complexity of the modulation and number of independent electrodes, because to have a higher number of degrees of freedom in the electron modulation, one has to introduce additional elements that make the setup more difficult to use and potentially more susceptible to damage and contamination. These devices are usually made with semiconductors, but the issue of charging can usually be avoided or

significantly mitigated for them by applying a ~ 100 nm thick metallic coating to the electrodes. Contamination and beam damage, instead, remain an issue for electrostatic phase plates as well, albeit less than the thin films because most of the device plane is just vacuum. Nonetheless, stray specs of dust or beam-induced carbon contamination of the devices can lead to the accumulation of localized charging that perturbs the beam-structuring process in ways that are quit difficult to model.

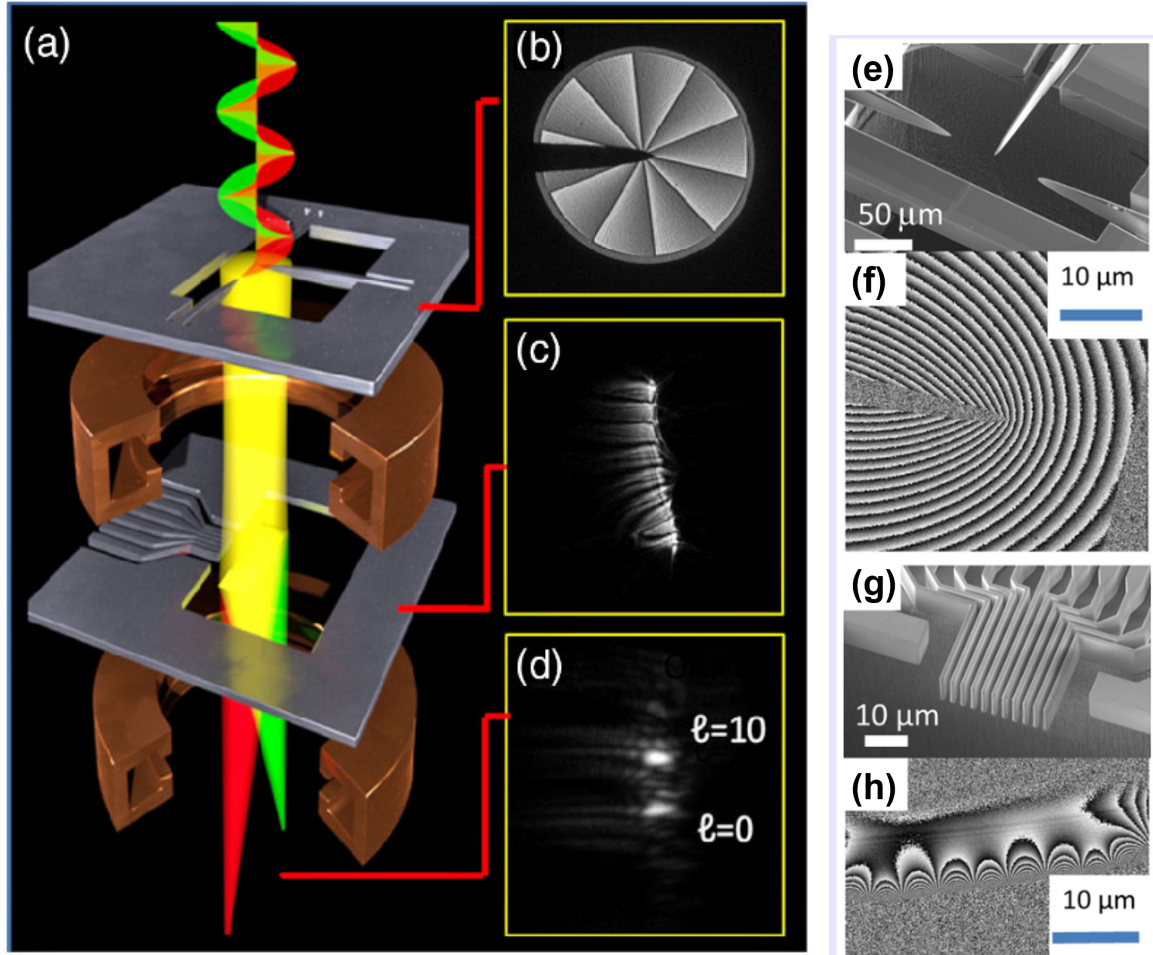


Figure 3.4: (a) Schematic of the placement of the sorter elements inside the TEM column. In the actual experiment, the "sample" phase plate was put in the second condenser aperture plane, generating an EVB with two superimposed $|\ell = 0\rangle$ and $|\ell = 10\rangle$ OAM states. The first sorter element (unwrapper) was put in the objective aperture plane, and the second sorter element (corrector) was put in the SAD plane. The OAM spectrum could then be observed in diffraction mode. (b)(c)(d) Evolution of the beam at the different positions along the column, with the spectrum separating the two OAM components. (e)(g) SEM images of the electrodes of the two phase plates. (f)(h) Electron optical phase contour maps showing projected electrostatic potentials recorded using off-axis electron holography [28]. Adapted from [63].

3.1.4 Magnetic phase plates

Magnetic phase plates produce the required phase shift through the Aharonov-Bohm effect [64], that gives rise to the magnetic vector potential dependent term of Equation 3.2. There are different ways to produce the vector potential needed for the modulation, through magnetized, but static, rings of magnetic material on thin film like in [65], to single magnetized needles that would act as magnetic quasi-monopoles [66], to the very recent Ampère phase plates [67] that use electric currents to produce tunable fields. Figure 3.5 shows an example of a magnetic phase plate realized with a magnetized microneedle, that produces an EVB carrying angular momentum.

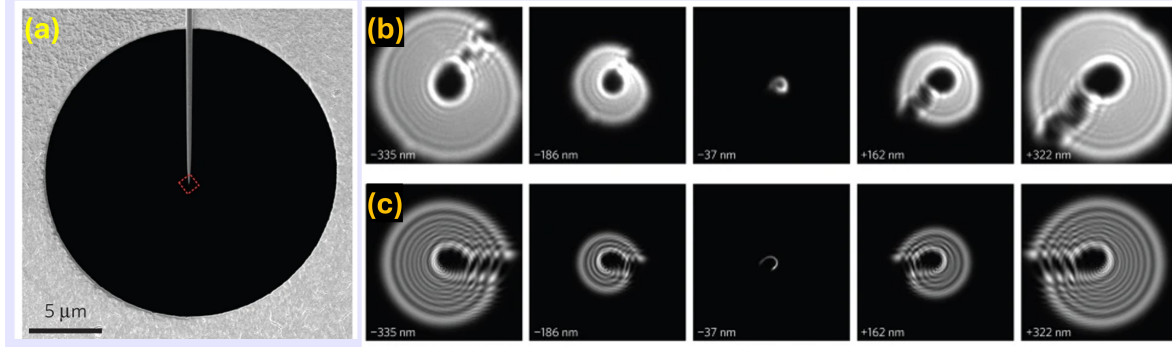


Figure 3.5: (a) SEM image of the magnetic needle phase plate. The needle was fabricated through Focused Ion Beam milling of a bulk Nickel sample and attached to a Si_3N_4 membrane. The tip is 200 nm in diameter, and is one end of a micromagnet rod a few tens of μm long. The circular aperture is coated in Au to have electrons interact only with one end of the magnet, exposing them to a quasi-monopolar magnetic field. (b) Experimental and (c) simulated through-focus series of the diffraction from the phase plate aperture. The resulting vortex beam can be clearly identified by the presence of the characteristic diffraction with a central singularity. Adapted from [66].

As it is starting to become evident in this analysis, magnetic phase plates are still in a somewhat early stage of development and are not yet comparable to electrostatic phase plates in terms of programmability. Also, manipulating the magnetic vector potential through magnetic samples or microscale circuits is more challenging than simply producing microscale electrodes, and magnetic phase plates are subject to the same issues of the electrostatic kind.

3.1.5 Optical modulation

Another possibility is to make phase plates by exploiting different forms of electron-light interaction. For instance, some proposed [68] light-based electron Zernike phase plates based on the use of the ponderomotive force that electrons feel as they traverse a region illuminated by a continuous wave laser. A very promising field for GI/SPI applications is the use of a Photonic free-Electron Modulator (PELM)[69], that exploit the Photon-Induced Near-field Electron Microscopy (PINEM) methodology [70], based on the inelastic coupling between electrons and photons in presence of a surface or a nanostructure. Essentially, the PELM requires the use of a surface to couple a structured light beam, that as we saw is much easier to produce with high accuracy and arbitrary configuration, with an electron beam inside a TEM. In this way it is possible to bypass entirely the requirement for a refined electron modulator, with a notable trade-off: the need to bring a structured laser beam inside the TEM column, and to synchronize the electron beam with the pulses of the laser. The practical realization of this setup is quite challenging, but it would be cumbersome to extend the discussion further. Figure 3.6 reports an essential description of the PELM design.

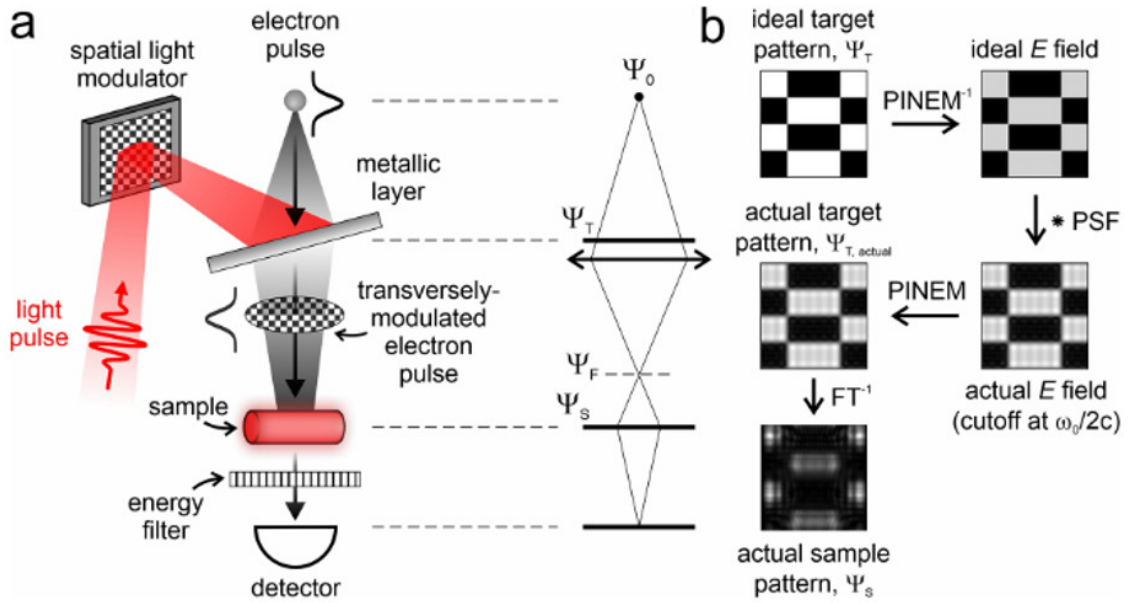


Figure 3.6: (a) Diagram of the PELM process, showing an electromagnetic field being modulated by an SLM and then coupled with the electron beam to produce a structured pattern to be set to the sample in a typical Computational GI scheme. (b) Example of the evolution of a target structured illumination pattern throughout the system. Adapted from [69].

Because of the intrinsically complex realization of an experimental setup for a PELM-based Computational GI/SPI method, it has not yet been realized in practice. However, as it was already reported, this would represent the closest approximation to an optimal electron beam modulator. Within the scope of this PhD work, the technique was still at a very early stage experimentally, so it was not possible to study a practical implementation of GI through PELM.

3.2 Our Devices and Optimal Beam Shaping for Ghost Imaging

Considering all the different kinds of phase plates presented in Subsection 3.1, it appears clear how the electrostatic phase plates are relatively easy to make and are able to meet the three criteria specified for the optimal phase plate for GI. The University of Modena and Reggio Emilia and the CNR-NANO institute have a long standing collaboration with the CNR-IMM institute in Bologna, where our group's devices are fabricated. Particularly, this PhD work is centered around two different designs of electrostatic phase plates. The basic structure consists of a thin layer of n-doped Si on a 200 nm SiO₂ substrate chip. The substrate and thin film are processed via photolithography to create the wires and contacts of the electrostatic phase plate, and an aperture is made in the substrate. In that way, the microstructured terminations of the electrostatic phase plates remain free-standing in vacuum, with a circular aperture delimiting the region of the actual phase plate. The devices are designed to be put in dedicated sample or aperture holders, as seen in Subsection 1.5 and 1.6, further enough in the TEM column that the combination of previous apertures and lenses can be used to set the convergence angle of the illumination and limit it to the desired section of the phase plate. Figure 3.7 shows the two phase plates used for this thesis:

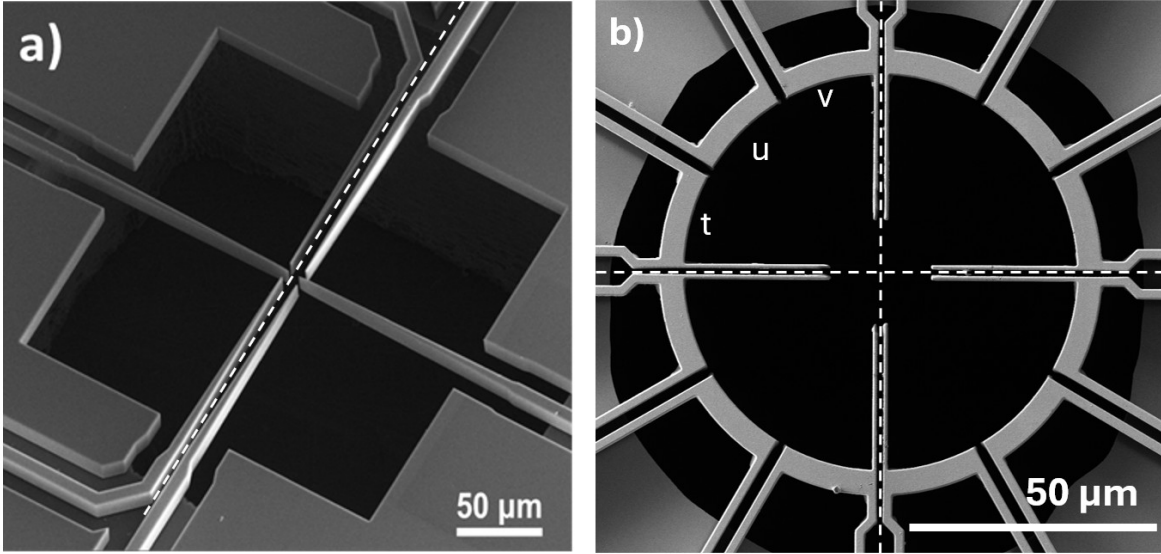


Figure 3.7: (a) SEM image of a detail of the first phase plate, made with two unwrappers from Figure 3.4e facing each other. The dotted line highlights the symmetry plane between the two mirroring sections. The resulting device has six independently biased needle-shaped electrodes. (b) Wide SEM view of the second phase plate, with the aperture and eight free-standing electrodes. There is a total of twelve elements forming the circular edge of the phase plate, connected in groups of three for each quadrant marked by the dotted lines. Each needle is connected to its own contact, and also to the other needle in its quadrant by a long wire, whose central section appears in between the two. So, elements *t* and *v* in the image are connected, and element *u* is the central part of the wire connecting them. This helps the phase to vary smoothly around the boundary of the phase plate. This second phase plate has eight independently biased contacts.

The phase plate in Figure 3.7a is derived from the unwrapper element of the OAM sorter experimental setup from [63]. By putting two of them facing each other, the

result is an electrostatic phase plate with 6 independent electrodes that we will call Double Sorter 1 (DS1). This device is capable of producing structured probes with one to three very intense spots, called caustics [73], and some fringe structures around them. Caustics are singularity points that can arise in the electron wave when the intensity is concentrated in a small region, and can be classified into different groups as described in [73]. Thanks to the superposition principle, the electrostatic potential distribution produced by each individual tip can be summed to the other five to obtain the total electrostatic potential distribution. The phase shift imparted by the phase plate can then be calculated according to Equation 3.2, considering $A_z = 0$, and integrating the total scalar potential between $z_i = -\infty$ and $z_f = +\infty$:

$$\Delta\varphi(x, y) = \frac{2\pi}{\lambda} \frac{e}{E} \frac{E_0 + E}{2E_0 + E} \int_{-\infty}^{+\infty} V(x, y, z) dz \quad (3.3)$$

The electrostatic potential distribution produced by a single uniformly charged tip has already been studied in literature [74][71], and it is possible to generalize the derivation of the potential for an arbitrary combination of linear electrodes of finite length, assuming that all of them lie in the same plane. In our phase plates, the electrodes are conductive and kept at constant potential, so that approach can be used to approximate their behavior to a good degree. For the DS1 phase plate, the formula can be simplified by assuming that all the electrodes are oriented parallel to the x or y axis, and that they lie in the $z = 0$ plane, where z is the direction of the optical axis. The potential will then be given by:

$$V(x, y, z) = \frac{1}{4\pi\epsilon_0} \left(\sum_i \int_{-c_i}^{+c_i} \frac{\sigma_i}{\sqrt{(x - h_{xi} + t)^2 + (y - h_{yi})^2 + z^2}} dt + \sum_j \int_{-c_j}^{+c_j} \frac{\sigma_j}{\sqrt{(x - h_{xj})^2 + (y - h_{yj} + t)^2 + z^2}} dt \right) \quad (3.4)$$

where σ_k is the charge density on the k -th tip, h_{xk} and h_{yk} are the displacements of the geometric center of the k -th tip from the origin and c_k is its half-length. The integration in dt is performed either along x (index i) or y (index j) depending on the direction with which the tip is aligned. Integrating the resulting total electrostatic potential according to Equation 3.3 returns the phase shift applied by the phase plate to an incident beam.

An interesting property of the design of these phase plates is that they belong to the category of Harmonic phase plates [72], because they are subject to the condition that the Laplacian of the phase must be equal to zero across their section: $\nabla^2\varphi = 0$. This is because the phase is proportional to the projected potential along z , as implied by Equation 3.2, and so the Laplacian of the phase will be proportional to the Laplacian of the potential, that is the charge density:

$$\nabla^2 V = -\frac{\rho}{\epsilon_0} \quad (3.5)$$

The charge density is of course zero everywhere except on the electrodes, where the charge is accumulated to generate the potential. This condition has two consequences: the first is that, as most of the section of the phase plate is vacuum, there is very little attenuation of the beam, so that most of the current is preserved. The second is that

the phase modulation is tunable but limited, because the phase has to vary smoothly across the phase plate, except for the discontinuities presented by the electrodes.

The phase shift calculated through Equation 3.2 is applied in the plane of the device, but its effects on the electron wave must be observed after some propagation distance along the optical axis. The standard way to calculate the propagation of an electron wave function when a sample is illuminated with a coherent electron beam in a wave-optical framework is the Kirchhoff-Fresnel integral [71]:

$$\Psi(X, Y, Z) = \frac{\exp(i\beta)}{i\lambda Z} \iint A(x, y) \exp \left\{ \frac{i\pi}{\lambda Z} \left[(x - X)^2 + (y - Y)^2 \right] + i\varphi(x, y) \right\} dx dy \quad (3.6)$$

where in the case when part of the device, or an additional aperture are limiting the illumination, $A(x, y)$ is an amplitude function that is used to take account of the opacity of the tips and the aperture. The result is correct up to a complex exponential factor $\exp\{i\beta\}$ that can be neglected, because the quantity of interest is the square modulus of the wave function in the plane of observation. The integral in Equation 3.6 is written assuming that the phase plate is in the $z = 0$ plane transverse to the optical axis, and the wave is propagated to the plane $z = Z$. Both the phase shift and the Kirchhoff-Fresnel integral can be calculated numerically, so it is possible to run simulations that predict the structured illumination at the desired distance, both in the Fresnel (near field) and Fraunhofer (far field) regimes [10]. An additional assumption is that the structured beam is observed in the back-focal plane of a lens. In that case, the electron wave function can be calculated from Equation 3.6, with additional phase terms that factor in the focal length f of the lens (because a lens images the far field of an object in its back-focal plane [10]), the phase plate term φ and the aberration function χ . Then, Equation 3.6 becomes:

$$\Psi(X, Y, Z) = \frac{\exp(i\beta)}{i\lambda Z} \iint A(x, y) \exp \left\{ \frac{i\pi(x^2 + y^2)}{\lambda} \left(\frac{1}{Z} - \frac{1}{f} \right) - \frac{2i\pi}{\lambda} [xX + yY] + i(\varphi(x, y) + \chi(x, y)) \right\} dx dy \quad (3.7)$$

As it was mentioned in Subsection 2.4, there was an application of SPI that used light caustics [53] for imaging, but in that case the caustic patterns were much more extended. The structured probes generated by the DS1 are typically more localized. Despite this very crude modulation, the DS1 was meant to be used for a proof-of-concept work on microscale imaging, and it provided enough control and variability for that purpose.

Moving on to the second phase plate in Figure 3.7, it is evident how it still uses needle-like electrodes, but there has been a change in their arrangement. The second phase plate was designed for more refined imaging at atomic resolution. For this application, the probes should be much smaller than those needed for microscale imaging, but this condition can be achieved even with the DS1, in principle. One must simply set the appropriate condition of the illumination system to make nanoscale probes. The key feature of the second device is that it has the ability to produce structured probes that resemble atomic lattices and lattice fringes. This is quite relevant, as other works on SPI with light optics pointed out how the effectiveness of the imaging method also

depends on the choice of the set of structured illumination most suited for the kind of target scene or application [41][51][75].

In STEM-HAADF imaging, crystalline samples imaged at atomic resolution show lattice fringes, with an intensity directly determined by the positions of the atomic columns, as discussed in Subsection 1.7. So, having a phase plate that can produce structured probes with similar features is very useful. We called the second device Quadrichopsticks (QC) phase plate, as it resembles an assortments of four chopsticks placed parallel to two perpendicular axes. Each "pair of chopsticks" is actually a Spiral Phase Plate (SPP)[17], a kind of phase plate that can generate an Electron Vortex Beam. The QC is made with four coplanar SPPs, each one working independently from the others. The following paragraph will illustrate the mechanism at the origin of the characteristic structured probes made by the QC phase plate.

3.2.1 Electron Vortex Beams and the QC phase plate

Among the possible types of structured electron waves, there is a kind called electron vortices [17]. The peculiar feature of electron vortices is that they possess helical phase fronts, that spiral around the propagation direction of the wave, and circulating current density. The presence of a spiraling phase is associated with the existence of a corresponding phase singularity. That is because if the phase gradually varies around the propagation direction, it must have a discontinuity in a plane transverse to the propagation direction. This in turn implies the existence of a point where different equiphase lines meet, where the phase is undefined. That point is the phase singularity point in that plane, and in 3D the singularity extends along a line. There are two properties associated to the singularity. First, as the phase is undefined, the electron wave function must vanish there: $|\Psi(\vec{r}_s)| = 0$. Second, in general the circulation of the phase around the singularity is different from zero:

$$\oint \nabla \text{Arg} \Psi(\vec{r}) \cdot d\vec{r} = 2\pi\ell \quad (3.8)$$

where $\ell = 0, \pm 1, \pm 2, \dots$ is an integer number. This means that the modulo 2π of the phase will have ℓ jumps from 2π to 0 by circulating around the propagation direction. ℓ is called the *topological charge* of the vortex, and its sign indicates the left- or right-handedness of the phase winding. Using cylindrical coordinates (r, φ, z) , and assuming propagation along the z direction, it is usually possible to express the wave functions of electron vortices in a form $\Psi \propto f(r) \exp(ik_z z + i\ell\varphi)$. The dependence on the topological charge suggests that these states carry some form of angular momentum, and indeed it can be verified [76] that they are eigenstates of the z -component of the quantum Orbital Angular Momentum (OAM) operator, $\hat{L}_z = -i\hbar \frac{\partial}{\partial \varphi}$, with associated eigenvalues $\hbar\ell$.

An Electron Vortex Beam (EVB) is simply an electron beam in which the individual electrons are in a vortex state, possessing a well defined *intrinsic* OAM component along the propagation direction. The word *intrinsic* here is used to distinguish the quantum OAM, that is independent from the frame of reference of choice, from the *extrinsic* classical OAM $\vec{L} = \vec{r} \times \vec{p}$ that instead depends on it. As seen already with some applications in Subsection 3.1, EVBs can be generated in a controlled way using Spiral

Phase Plates, that are able to impart to an incident beam the azimuthally-varying phase gradient typical of EVBs. A possible strategy to do so is to use an electrostatic phase plate made of two oppositely charged linear electrodes. As demonstrated in [77], a line of charges of constant density σ , placed in the $z=0$ plane between the points $(0, -a)$ and $(0, 0)$ generates an in-plane electrostatic potential that can be calculated with Equation 3.4 as:

$$V(x, y, 0) = \frac{\sigma}{4\pi\epsilon_0} \left[\log \left(\frac{\sqrt{(a+y)^2 + x^2} + a + y}{\sqrt{x^2 + y^2} + y} \right) - \frac{a}{\sqrt{(x+x_D)^2 + (y+y_D)^2}} \right] \quad (3.9)$$

with (x_D, y_D) being the in-plane coordinates of a neutralizing charge, that in this case can be seen as located on a nearby conductor, like the side of the holder carrying the SPP. Using this potential in Equation 3.2, yields:

$$\begin{aligned} \Delta\varphi(x, y) = C_E \frac{\sigma}{4\pi\epsilon_0} & \left[- (a+y) \log \left((a+y)^2 + x^2 \right) + y \log(x^2 + y^2) + 2a \right. \\ & \left. + 2x \arctan \left(\frac{y}{x} \right) - 2x \arctan \left(\frac{a+y}{x} \right) + a \log \left((x+x_D)^2 + (y+y_D)^2 \right) \right] \end{aligned} \quad (3.10)$$

where C_E is the interaction constant, the prefactor of the integral in Equation 3.2. Putting two oppositely charged line electrodes at a distance α gives a combined resulting phase shift $\Delta\varphi_{TOT}(x, y) = \Delta\varphi(x - \alpha, y) - \Delta\varphi(x + \alpha, y)$ like the one shown in Figure 3.8a:

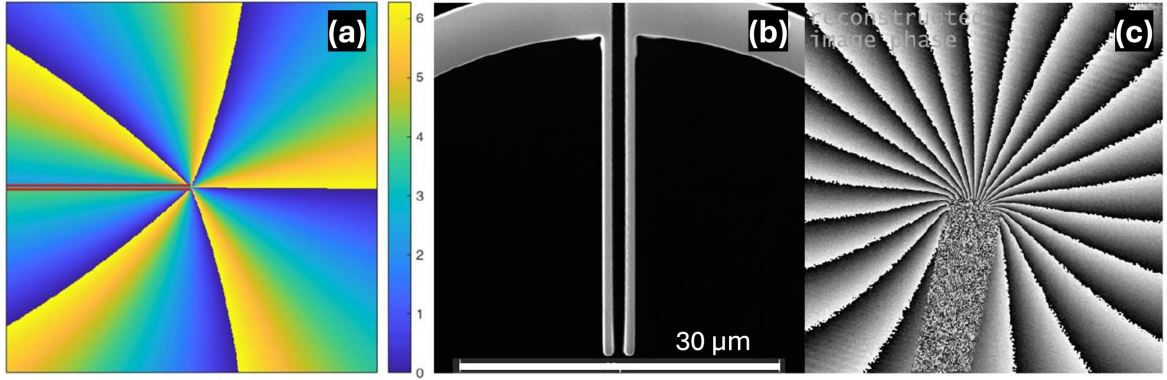


Figure 3.8: (a) Simulated phase shift for an SPP that is applying an $\ell = 5$ topological charge. The line charges in the central-left part of the pael (in red) have a length $a = 200 \mu m$, their separation is $\alpha = 0.2 \mu m$, and the electron beam energy is assumed to be $300 kV$. The phase is wrapped over 2π , as evidenced by the colorbar. (b) SEM image of a real electrostatic SPP. (c) Experimental phase profile observed through Off-Axis Electron Holography [28] by applying a $\Delta V = 2V$ between the two needles, $+1V$ on one and $-1V$ on the other. As expected, the phase varies azimuthally around the device. Adapted from [78].

The QC phase plate combines four of these SPPs, all identical, on two perpendicular lines on a circular aperture. The QC can thus generate four separate vortex beams, each one with its own topological charge, that can be tuned in real time by changing the biases applied to the contacts of the device. This can be understood by considering that

Equation 3.8 tells us that the OAM imparted by each SPP depends on the circulation of the phase around the SPP, and Equation 3.3 tells us that the phase produced by the SPP is proportional to the projected potential in the direction perpendicular to the phase plate. So, it is possible to identify a proportionality between the voltage difference across the needles of one SPP and the topological charge it applies to the electron wave:

$$\ell \propto \Delta V_{SPP} \quad (3.11)$$

The four EVBs are generated near to each other and interfere to produce a controlled structured phase landscape, that under the appropriate conditions present features similar to lattice fringes. A more detailed description of the structured probes generated by the QC and the DS1 phase plates is given in Section 7, where we present the results of some numerical simulations made to model the two electrostatic phase plates.

Some final remarks before moving to the next Section. Both the phase plate designs presented here have electrodes that extend approximately towards the center of the phase plate, meaning that the tips of these line electrodes are separated by only a few μm . This could lead to a slight cross-talk between the different electrodes, but that is a price that must be paid to comply with the need to make a more interesting structured phase. This is especially true for the QC phase plate, where there must be interference between the EVBs generated by each pair of electrodes. If the needles were far apart the interference would be much weaker, as the singularities (the centers of the phase vortices) are located close to the tips of each pair.

Another aspect to bear in mind is that in principle, any structured probe generated with these phase plates is biunivocally associated to the configuration of biases applied to the electrodes. For the DS1, that has a lower symmetry, this simply means that the 6 applied potentials define the structured probe. The QC is instead effectively an octupole, though not perfectly radially symmetric. If the electrodes are biased with a random set of biases, the resulting probe will be generically structured. If instead the biasing is uniform, with positive or negative sign, there will be a defocusing effect on the probe. Similarly, a configuration where the opposite "chopsticks" are biased with the same sign, and with opposite sign to the other two, can introduce some two-fold astigmatism effect. And it is also possible to apply opposite biases to opposite chopsticks, leaving the other two electrodes at 0 V, creating a tilt.

All these effects become relevant because they can be used to better understand the properties of a set of structured probes made under a given illumination condition and biasing range. The 8 degrees of freedom associated to the available biasing channels of the QC phase plate can also be seen as 4 OAM degrees of freedom, for the biasing of each individual "chopstick", plus 2 degrees of freedom for two-fold astigmatism, and 2 more degrees of freedom for tilting along two perpendicular directions. A detailed explanation is given in Subsection 10.2 where code to generate the QC probes is explained. These degrees of freedom allow generating a variety of structured probes, by combining the "basic" interference of the four EVBs with the astigmatism and tilt that can move the structured probes around a central position and stretch them. Naturally, the tilting and astigmatism effects arise even with the DS1 phase plate, but it is more difficult to characterize them because the device is less symmetric.

4 Electron Computational Ghost Imaging

This section is devoted to the introduction of the method at the core of this thesis, that we called Electron Computational Ghost Imaging (ECGI). The method is proposed as a direct adaptation of the Computational Ghost Imaging method discussed in Subsection 2.3 in a Transmission Electron Microscope. Figure 4.1 shows a diagram with the principal information about the method:

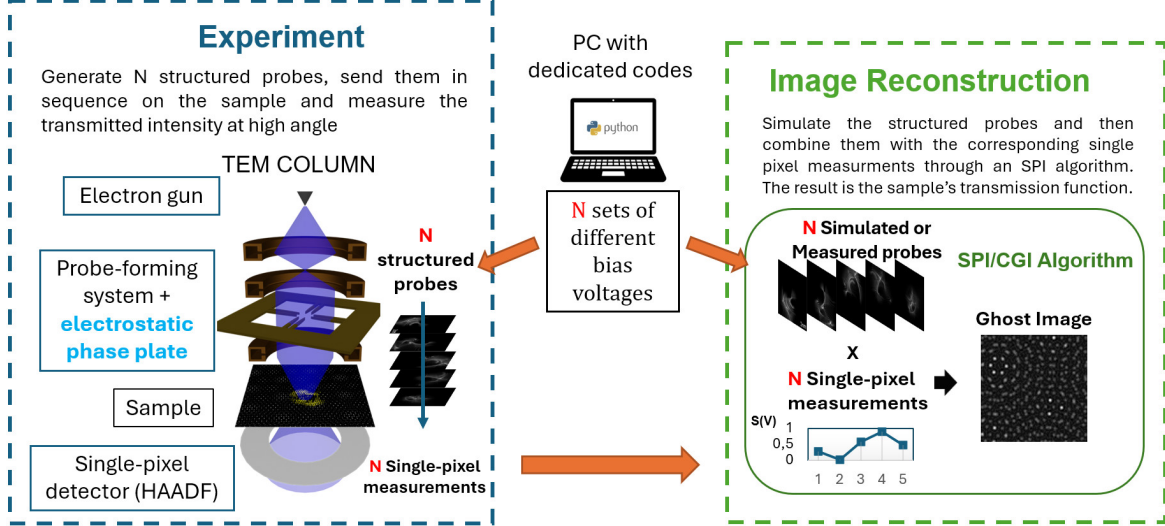


Figure 4.1: Diagram of the Electron Computational Ghost Imaging method. The method can be divided into two main parts. The first one is the experiment, performed at the TEM, where the goal is to collect the signal from the high-angle scattered electrons, with one measurement for every structured probe. The second part is the image reconstruction, where the measured signal and the simulated images of the corresponding structured probes are combined using an SPI algorithm to recover the sample's object function in the form of the Ghost Image. The PC is needed both to control the electrostatic phase plate and store the measurements during the experiment, and to run the SPI algorithms for image reconstruction.

There are two main parts of the method. The first one is carried out at the TEM. The *Experiment* box in Figure 4.1 shows a simple diagram of the TEM column with the phase plate inserted in the last aperture position available before the sample. The microscope is operated in STEM mode, with a convergent illumination, to have a micro- or nanoscale probe at the sample plane (depending on the desired scale for imaging). The combination of source size, apertures and lens excitations determines the convergence angle of the probe and the region of the phase plate that is effectively used while forming the structured probes, so it is crucial for the experiment. Notice how in the scheme there is not a probe corrector. Our goal is to incorporate aberrations as well in the image reconstruction scheme, so having a corrector would only make it impossible to use our device by complicating exponentially the task of characterizing the phase plate.

The Control System generates a list of N sets of different biases to apply to the electrodes of the phase plate, and then synchronizes the switching between different probes with the signal measurement of the transmitted intensity at the HAADF detector. The data produced during the first part is the list of N sets of biases applied to the phase

plate and the corresponding N measurements of the HAADF signal. In case that the structured probes can not be simulated numerically, as it is expected for Computational Ghost Imaging, they can be measured experimentally and stored in a database. Under those circumstances, the only additional requirement is that there must be no change in the illumination system and in the device positioning between the step where the probes are imaged and that in which the signals are measured. The details will be discussed later in Section 8.

The second part of the method is found in the *Image Reconstruction* box, where the numerically simulated (or experimentally measured) structured probes are combined using an SPI algorithm. Depending on the number of structured probes used for a measurement, the size of their images, and the kind of algorithm being used, the image reconstruction part can take from a few minutes to a few days of calculation, and a considerable amount of RAM memory and computing power. The resolution of the final image is lower-bound by the pixel size of the images of the structured probes used to run the algorithm, but it does not necessarily coincide with it. The spatial information in the final image is obtained from the one present in the structured probes. That means that the set of structured probes used for a measurement should sample the spatial frequency space as uniformly as possible, because the finest features in the probes will determine the finest features found in the final image.

The proposed method can be classified within the domain of electron computational imaging methods, similarly to other techniques like the already mentioned Holography and Ptychography. The proposed setup is designed to produce optimal results on thin crystalline samples, where its operating conditions are met, as discussed later in Section 7.2. Then, the ECGI method should be able to recover amplitude information at high resolution by sampling the specimen's characteristic spatial frequencies through a set of structured illuminating probes. As it was already stated in Section 2.5, and repeated here, the advantage of ECGI should be the possibility of incorporating the aberrations in the probe-generation scheme. That implies the realization of high-resolution imaging with convergent illumination beyond the usual limit imposed by aberrations on uncorrected microscopes.

Thus, if the proposed method could be implemented fully, it would open the way to high-resolution amplitude imaging even on relatively cheaper microscopes, where a probe corrector and/or high-resolution pixelated detector is not installed or can't be installed. Finally, as already stated in Section 2.5, the long-time perspective for the development of the method is not only to see it fully implemented through the proposed setup, but also to explore the effect of using other sources of signal between the numerous products of the beam-sample interaction.

4.1 Hybrid Scanning-ECGI Method

The ECGI method already makes use of the STEM mode of the microscope, to produce a nanoscale convergent structured probe at the sample plane. Being in STEM mode also gives access to the scanning mode of the microscope, that is not normally considered in the light optics versions of the technique, where the beam is stationary. This opportunity offers three additional advantages that can further improve the image reconstruction process. Figure 4.2 shows the use of the scanning system in the context of ECGI:

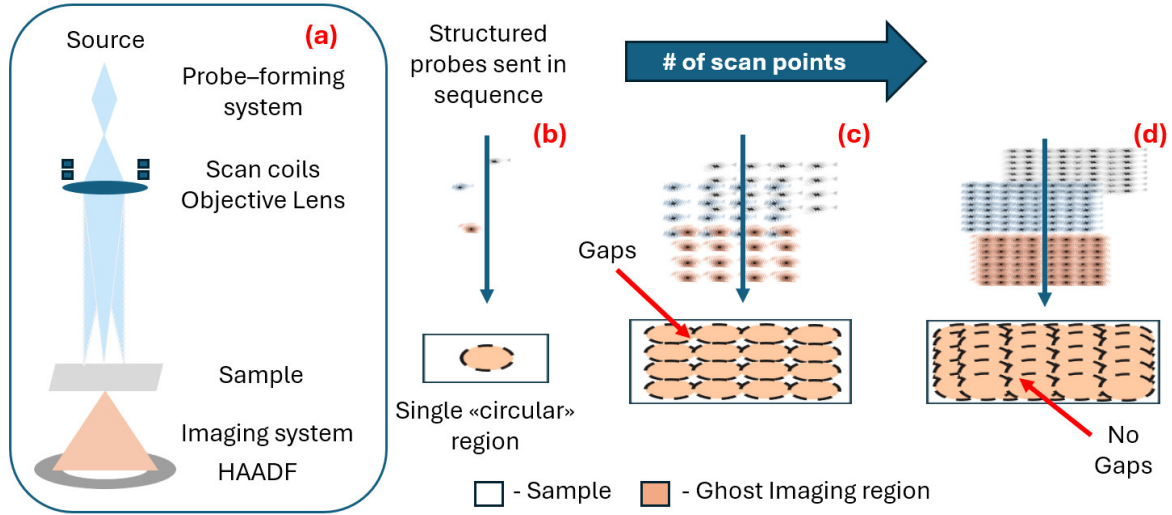


Figure 4.2: (a) Diagram of the TEM column with the detail of the scanning system. (b)(c)(d) Schematization of the effect of adding probe scanning to the ECGI method. Notice how as the number of scan points increases, the total reconstructed area becomes rectangular and uninterrupted.

The first advantage of introducing the scan is to reconstruct square images instead of circular/elliptical ones. The structured probes generated by our phase plates fall either in a circular area, in the case of the QC, or in an elliptical one, for the DS1. This is due to the symmetry of the phase plates: the DS1's electrodes have a two-fold symmetry and are arranged in a more "rectangular" disposition, while the QC's electrodes have a four-fold symmetry. If the beam is not scanned, the regions illuminated by the two phase plates follow their geometry, as seen in Figure 4.2b. When the STEM raster scan is turned on, the probes are rastered over a square grid, so the illuminated region becomes a square. If the scan points are too far apart, the final image will have gaps between adjacent areas illuminated by the device. This problem can be solved in two ways: either the number of scanning point is increased at the same STEM magnification so that the gaps are filled, or the STEM magnification is increased for the same number of grid points. The STEM magnification just changes the size of the step for the raster scan. Of course, in the first case there will be a greater number of data points to use for the reconstruction, so this approach might be helpful or detrimental depending on the available computational resources. The second approach keeps the same number of scanning points, but scans a smaller area, reducing the field of view of the final image.

There is also the possibility to change the size of the area scanned by the probes at fixed number of scan points and STEM magnification by expanding the interval spanned by

the electrodes' biases, as discussed in Section 7.1. However, this third strategy is more troublesome to adapt, as the size of the region explored by the structured probes increases rapidly with the applied bias, possibly leaving unmapped areas within a single scan point if the number of patterns used is not large enough. In any case, the hybrid scanning-enhanced ECGI can make better use of the final image's field of view by reconstructing a larger area.

The second advantage can be directly deduced from the first: if the probes are scanned, there is a less strict requirement on the differentiation between probes generated by a device. With a $P \times Q$ raster scan and N_P different probes, the total number of different probes used is $N_{TOT} = N_P \cdot P \cdot Q$, because a shifted probe counts as a different one, as it is probing a different region of the sample. So, for instance, to have 40 thousand probes on the device, one could use $N_P = 100$ different probes on a $P \times Q = 2020$ grid instead of having to use 40000 different probes. The other aspect to consider is that using the STEM scan system the HAADF signal for each scanned probe can be acquired and stored with the microscope's detector as a HAADF image, simplifying the acquisition process. Finally, to address a possible concern, the size of the raster scans that we wanted to use was between 32×32 and 64×64 , not enough points to retrieve a high-resolution STEM-HAADF image on their own, especially considering that the structured probes are extended and not point-like as a STEM probe should be.

The last advantage about the use of the scanning comes from a hypothesis about one of the SPI algorithms selected for the image reconstruction process. Essentially, the equation at the base of that algorithm mirrors one of those used for another scanning-based imaging technique: Ptychography. A more detailed explanation will be presented in Section 5, but the key concept is that in Ptychography, the image reconstruction process recovers both the sample and the probe's functions. So, we considered the possibility to write a dedicated algorithm that would work on the sample and the structured probes simultaneously, to correct for any error in the simulated/measured probe images used for the reconstruction, when compared to the real ones used to measure the HAADF signal. The probe reconstruction or refinement becomes possible because the same probe is used in different scanning positions so as to make the measurement more redundant. This is indeed the same redundancy that permits to refine the probe in Ptychography.

5 Single Pixel Imaging Algorithms

So far, it was stated several times that Computational GI requires an algorithm to perform the actual image reconstruction, combining the simulated structured illuminations and the single-pixel detector signal. This section provides the mathematical background for the algorithms used during the *image reconstruction* part of ECGI. Particularly, the algorithms were selected among those that have been already studied for light optics, both in the context of Ghost Imaging and in that of Single-Pixel Imaging, because the mathematical description is completely equivalent for the two approaches.

All the algorithms that will be described here are based on a specific assumption: the linearity of the measurement process. In mathematical terms, this constraint can be modeled by saying that the scalar signal that is measured at the detector, say S , is simply proportional to the overlap integral between the probe $\vec{P}$ and the sample's transmission function $\vec{T}$. Within this framework, $\vec{P}$ and $\vec{T}$ are both one-dimensional vectors representing images of n pixels, while the signal $S = \vec{P} \cdot \vec{T}$ is the result of their scalar product. Using this notation, the entire experimental measurement process with m probes can be rewritten as a linear system of equations using matrix notation:

$$\hat{P}\vec{T} = \vec{I} \quad \Leftrightarrow \quad \hat{A}\vec{x} = \vec{b} \quad (5.1)$$

where $\hat{P}$ is the measurement matrix whose m rows are the structured probes, $\vec{T}$ is the sample's transmission function and $\vec{I}$ is the one-dimensional array of m single-pixel measurements. A more common notation [79] for this kind of problem is the second form in Equation 5.1, where $\hat{A} \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^{n \times 1}$, and $\vec{b} \in \mathbb{R}^{m \times 1}$. The constraint on the linearity of the measurement process is quite relevant, and its validity for ECGI is studied in Section 7. The ultimate goal of an algorithm applied to a problem like Equation 5.1 is to solve for $\vec{x}$. All the Computational GI/SPI algorithms can be classified into two main categories: iterative methods and non-iterative methods. As the nomenclature suggests, the first kind converges to a solution by using the whole set of structured probes over multiple iterations, while the second kind performs a full reconstruction without iterating multiple times over the dataset. The discussion will begin with the non-iterative methods, as they are the ones that are easier to understand intuitively.

5.1 Non-Iterative methods

Non-iterative methods solve the linear problem directly, without iterating over the dataset multiple times. The single pixel measurement are obtained from the overlap between the sample's transmission function and the probes, so they can be interpreted as weighting coefficients, that quantify how similar the probes are to the imaged scene. To get some insight into the method, it is useful to imagine what would happen when trying to image a completely opaque target with a central round hole. If a probe is almost completely centered on the hole it will have a high signal, while if it is almost completely out of the hole it will have a very low signal. Then, the simplest strategy to recover the transmission function of the sample is to calculate a weighted sum of all the probes, with weights given by the corresponding single-pixel measurements:

$$\vec{x} = \frac{1}{m} \sum_{i=1}^m (b_i - \frac{1}{m} \sum_{i=1}^m b_i) \vec{a}_i = \{b_i \vec{a}_i\} - \{b_i\} \{\vec{a}_i\} \quad (5.2)$$

where b_i is the i -th measurement of the signal and $\vec{a}_i$ is the i -th row of the measurement matrix, i.e. the i -th probe used. The terms in curly brackets are the averages over the entire measurement set, so the final image $\vec{x}$ can be seen as the difference between the average of the weighted probes and the product of the average signal for the average probe. The average of the probes can approach a uniform distribution over the illuminated area if the probes are independently identically distributed and a large number is measured. Then, the final image will have a flat average background, and the more a probe is blocked by the sample (i.e., the lower its overlap with it), the darker it will appear, while if it is largely transmitted it will appear bright. Summing all the contributions of the probes reveals the contrast produced by the sample's transmission function. This algorithm is the most basic non-iterative one, and since it has also been used often in literature on GI, we will refer to it as Traditional Ghost Imaging (TGI) algorithm.

In Subsection 2.4 it was pointed out that some works of Computational GI and SPI were focused on optimization strategies, like selecting the most appropriate set of probes for the scene of choice. Several well-established probe sets, like the Hadamard matrices, efficiently sample the frequency space and are orthogonal, so each probe gives information on a specific subset of spatial frequencies. In the case of our phase plates, the orthogonality constraint is not easily imposed because they do not grant such a high-level control on the modulation. So, it was interesting to try to implement a form of orthogonalization of the probe set during the image reconstruction process. We introduced this idea first in [80], as a method inspired to the Kaczmarz algorithm [81][82][83], a solver for linear systems of equations. This variation of TGI algorithm was called Orthogonalized Ghost Imaging (OGI):

$$\vec{x} = \frac{1}{m} \sum_{i=1}^m [(b_i - \{b_i\}) - (c_i - \{c_i\})] \vec{a}_i \quad (5.3)$$

where $c_i = \vec{a}_i \cdot \vec{x}_{i-1}$ is the estimate of the i -th single-pixel signal value obtained from the overlap between the i -th probe and the latest update of the reconstructed image. The core of this method is to subtract a correction term (the second round bracket, $(c_i - \{c_i\})$) to the TGI algorithm. The correction is mitigating the effect of summing non-orthogonal probes by removing their contribution to parts of the image that have already been covered by other probes. Normalization become quite relevant in this case, because the real single-pixel signal $\vec{b}$ and the estimated one $\vec{c}$ must have similar magnitudes on average, or one contribution will dominate on the other, preventing proper convergence. If the normalization is curated carefully, the quantity $[(b_i - \{b_i\}) - (c_i - \{c_i\})]$ should tend to zero as the number of combined probes increases. This is because the non-orthogonalized probes progressively have a smaller and smaller chance of falling in an area that has not been explored by the previous ones.

5.2 Iterative Methods

The two methods seen so far were based on a single iteration over the measured dataset, but the problem of solving the linear system of Equation 5.1 can also be visualized as an optimization problem to be solved iteratively, by applying some constraints set by the experimental data. There are several possible linear and non-linear iterative algorithms, but for this thesis work I concentrated on the Alternating Projection (AP) algorithm [79], that was first proposed in [84]. It is a method derived from previously existing methods dedicated to phase retrieval [85][86], like the algorithm used for Ptychography [28] that is derived from the Gerchberg-Saxton [87] algorithm. The idea behind this algorithm is to begin with an estimate of the reconstructed image of the sample $\vec{x}$, and an estimate of the transmitted intensity after the interaction with the sample, $\vec{a}_i \cdot \vec{x}$, and update both alternatively. The AP sample update step is:

$$\vec{x}' = \vec{x} - \frac{\vec{a}_i^T}{\max(\vec{a}_i)^2} \frac{\vec{a}_i \cdot \vec{x} - b_i}{n} \quad (5.4)$$

The estimated signal $\vec{a}_i \cdot \vec{x}$ is calculated. Then the experimental signal b_i is subtracted from it to compute a coefficient (second fraction). The probe $\vec{a}_i^T$ is multiplied by this coefficient and everything is normalized to the maximum value of the probe and the number of pixels n . This correction is then used to update the final image. At each iteration the algorithm uses all the probes, one at a time. In this way, the algorithm uses the experimental signal and the experimental probes as constraints on the transmitted intensities and on the illumination. As the algorithm is run for multiple iterations over the experimental dataset, the reconstructed image converges to the sample's transmission function. Intuitively, as the reconstruction gets closer and closer to the real sample, the difference $\vec{a}_i \cdot \vec{x} - b_i$ grows smaller and smaller, giving progressively milder corrections to the reconstruction.

There are other iterative methods, both linear and nonlinear, but I limited myself to the use of the AP method because, as reported in [79], it offers a good compromise between the required running time and the required number of different measurements, and it is also less memory-intensive because it uses the probes one at a time instead of all at once like other methods. This became particularly important for the application of the hybrid scanning-ECGI method proposed in Subsection 4.1. Hybrid scanning-ECGI requires to use $P \times Q$ shifted versions of each probe when running the reconstructions. Storing many probes in working memory, at a decent resolution of a few hundred pixels, for all the shifted replicas of the same probe would have exceeded the resources available at the time. So, I modified the AP algorithm (and also the TGI one) to include the scanning motion of the probes in the reconstruction scheme. The shifting motion was introduced by moving the reconstructed image within a larger space. The shift was done either with pixel-wise movements, when it was commensurate with the reconstructed image pixel size, or through Discrete Fourier Transform shifting [88] if the shift was sub-pixel.

When talking about iterative methods, it is also often mandatory to introduce some sort of criterion to establish when the algorithm has reached a sufficiently accurate solution. This condition is called *convergence*, and it is usually quantified by a function that estimates the correctness of the result, called the *loss* function. For the AP algorithm,

the choice of the loss function fell on the square root of the average square of the difference between the correct and the estimated signal. In symbols:

$$loss = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{a}_i \cdot \vec{x} - b_i)^2} \quad (5.5)$$

As the error between the recovered image and the real sample gets smaller and smaller, the estimated signals $\vec{a}_i \cdot \vec{x}$ get closer and closer to the experimentally measured ones b_i , making the loss drop to zero. Depending on the data set, the loss can decrease in just a few iterations or over several tens of iterations.

5.3 Ptychography-inspired variation of Alternating Projection

As it was previously mentioned, the AP algorithm was adapted from a preexisting family of iterative algorithms, and shares some similarities with the algorithms developed for Ptychography [28]. Those similarities brought us to wonder whether it would be possible to modify the AP algorithm to exploit some of the advantages of Ptychography.

Recalling Equations 5.6 and 5.7 presented in Subsection 1.10, it is easy to trace a parallel with the form of Equation 5.4. The equations for the update of the object function $O(\vec{r})$ and the probe function $P(\vec{r})$ are respectively:

$$O_{j+1}(\vec{r}) = O_j(\vec{r}) + \alpha \frac{P^*(\vec{r} - \vec{R}_{s(j)})}{|P(\vec{r} - \vec{R}_{s(j)})|_{max}^2} (\psi'_j(\vec{r}) - \psi_j(\vec{r})) \quad (5.6)$$

and:

$$P_{j+1}(\vec{r}) = P_j(\vec{r}) + \beta \frac{O^*(\vec{r} + \vec{R}_{s(j)})}{|O(\vec{r} + \vec{R}_{s(j)})|_{max}^2} (\psi'_j(\vec{r}) - \psi_j(\vec{r})) \quad (5.7)$$

The parallel is better understood by comparing two equations together and considering the ePIE object update at a single position, because AP does not natively include probe shifting:

$$O_{j+1} = O_j + \alpha \frac{P_j^*}{|P_j|_{max}^2} (\psi'_j - \psi_j) \quad \Leftrightarrow \quad \vec{x}' = \vec{x} - \frac{\vec{a}_i^T}{\max(\vec{a}_i)^2} \frac{\vec{a}_i \cdot \vec{x} - b_i}{n} \quad (5.8)$$

The object function is updated by multiplying the probe function (normalized to its maximum squared) by the difference between the updated guess of the exit wave function and the previous exit wave function. In the case of AP, the difference is between two scalars instead of two functions, but the concept remains roughly the same. One of the terms depends on the reconstructed object function at the previous iteration ($\vec{a}_i \cdot \vec{x}$) and the other depends on the constraint in Fourier space (b_i). The similarity between the two methods is even apparent when comparing the experimental setups. AP is based on using multiple structured illuminations in the form of probes that sample information about the object in real space, and then uses a simple single-pixel detector. ePIE is based on using a simple probe to measure a lot of information through a structured detection from the diffraction of the sample, so it samples information in reciprocal space.

Thus, we postulated that it would be possible to write a probe update step for the AP algorithm as well, assuming to be using our hybrid Scanning-ECGI setup. The probes would be scanned in a small raster compared to their size, so that most of the information used to reconstruct the sample would come from regions where the probes self-overlapped. Hopefully, this would be enough to allow the transfer of some information from the reconstructed sample to the probes. The probe update step is:

$$\vec{a}_i^T ' = \vec{a}_i^T - \frac{\vec{x}}{\max(\vec{x})^2} \frac{\vec{a}_i \cdot \vec{x} - b_i}{n} \quad (5.9)$$

The advantage of Ptychography is that it is collecting most of the information in reciprocal space, after the interaction with the sample, and it is using a single probe. So, there is only one probe to correct, and the correction can be done by enforcing a reciprocal space constraint based on a lot of information (J diffraction patterns of $M \times N$ pixels). For our scanning-ECGI method instead, the only strict constraints are those on the J measured signals, because it is difficult to make sure that the probes at the sample level are exactly like the simulated/experimentally measured ones. It would be possible in principle to invoke a relation between different probes but it requires some accurate knowledge of the beam shaping device. The algorithm would be therefore less general. Finally, I must add that if our modified version of the AP algorithm proves effective in recovering the structured probes, some suggestions for further improvement have already been discussed for the ePIE algorithm [89], and they could be used as a starting point for the development of our algorithm. I did not further explore these paths for improvement.

6 The Multislice Method

The basic assumption of the SPI algorithms that have been presented is that the interaction between the structured probe and the sample is well represented by Equation 5.1. That is to say, the wave exiting the sample is simply the product of the structured illuminating wave and the sample's transmission function. This linear behavior is intuitively approximately true for the microscale imaging condition that will be discussed later, because in that case we measured the total transmitted intensity of the sample, so all the intensity that was lost upon interaction was either absorbed by the thick sample or backscattered. For the nanoscale/atomic resolution imaging, instead, it becomes critical to verify that the measurement is linear, because the single-pixel signal is only determined by the high-angle scattering, ignoring completely what happens at lower angles.

The Multislice method [90][2] is a versatile simulation technique that can be used to verify whether that condition is satisfied. The Multislice method computes the propagation of an incident wave through a sample in very small steps, so that it accounts for all the scattering effects due to the atoms of the sample. In this way, the Multislice method is able to simulate experiments done in a variety of conditions, both in terms of illumination of the sample and for the method of detection. The Multislice method can simulate perfect crystals, but it can also be used to study point or line defects and more complex configurations. Also, it is not used purely for simulations. Recently, it has been employed to improve the results of experimental imaging methods. For instance, it was proposed to improve Ptychography in 2012 [91], and after some development it led to remarkable results in the last two years in [92] and [93]. Another example is in [94], where there is an application of Multislice to Electron Tomography in STEM.

The origin of the name Multislice comes from the modeling of the sample that is necessary for the application of the method. The sample is sectioned into many 2D slices perpendicular to the propagation direction. The slices can be as thin as an atomic layer, or even thinner. The algorithm takes in input a model of the sample and the wave function representing the illumination (in most cases, a plane wave for TEM or a convergent probe for STEM). Then, it calculates the interaction between the entrance wave with the first slice, giving an estimate of the exit wave. That exit wave is then propagated in vacuum to the position of the next slice. The process is iterated through-

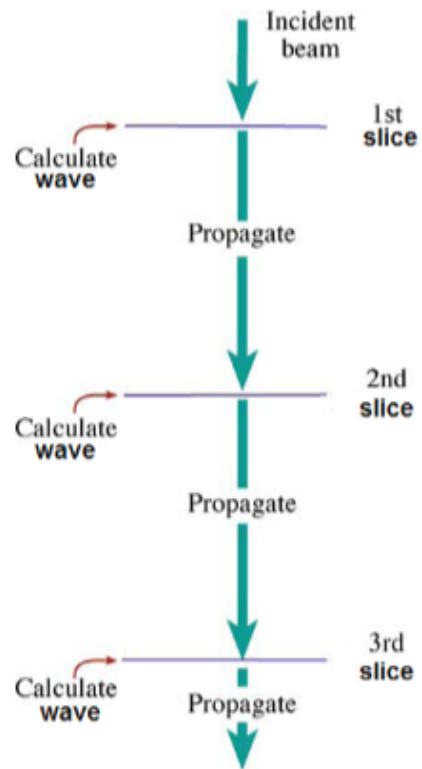


Figure 6.1: Diagram of the Multislice method. The incident wave is propagated through the sample between ever slice. At every slice position, the wave is multiplied by the slice's transmission function. Adapted from [2].

out the entire thickness of the sample, for all slices. The final exit wave can be used to calculate the high-resolution TEM image of the sample (through its square modulus), while the square modulus of its Fourier transform provides the diffraction pattern of the sample. For STEM, it is possible to integrate the diffraction pattern between two values of scattering angle to obtain the STEM BF/ADF/HAADF signal.

There are a few possible implementations of the Multislice method from the computational standpoint. Our code uses the well-known Kirkland implementation [90]. The sample is divided into slices, of either equal or variable thickness. Given a known structure for the sample, the transmission function for each slice at position z along the propagation direction can be computed as:

$$t(x, y, z) = \exp \left[i\sigma \int_z^{z+\Delta z} V(x, y, z') dz' \right] \quad (6.1)$$

where $\sigma = 2\pi me\lambda/h^2$, and Δz is the thickness of the slice. Indexing the slices with $n = 0, 1, 2, \dots$, each slice's transmission function will be t_n . The propagation between the slices is obtained by using the Fresnel propagator p_n along the propagation direction:

$$p_n(x, y, \Delta z_n) = \frac{1}{i\lambda\Delta z_n} \exp \left[-i\pi \frac{(x-x')^2 + (y-y')^2}{\lambda\Delta z_n} \right] \quad (6.2)$$

The illumination wave function $\psi_0(x, y)$ at slice 0 can be chosen to be a plane wave or a convergent probe. Then, at every step, the wave function update is given by:

$$\psi_{n+1}(x, y) = t(x, y)[p_n(x, y, \Delta z_n) \otimes \psi_n(x, y)] + \mathcal{O}(\Delta z^2) \quad (6.3)$$

where $\otimes$ is the convolution product. The error depends on the square of the slice thickness squared Δz^2 , but for a large enough number of slices $N_s \propto 1/\Delta z$, so the overall error goes as $1/\Delta z$. The product with the slice transmission function t_n and the propagated wave function is simply a product between complex functions because each slice should be thin enough that the WPO approximation holds true for it.

Part II

Materials and Experimental Methods

This Part covers the development of the methods introduced in the previous Part, and the necessary steps taken to move from theory to real experiments. In order, this Part will discuss the most relevant simulation work, the methods used to characterize the experimental phase plates, the implementation of the control software, the samples, and finally the experimental procedure of the real experiments.

7 Simulations

The preparation and design of complex experiments always requires some degree of practical tuning of the experimental setup. In most cases, it is convenient (when possible) to run simulations, so that some aspects of the experiment can be tested in a controlled environment. Simulations allow to decouple the different effects that determine the final result, and to spot potentially problematic sources. In particular, this Section will cover three main aspects: simulations of the beam shaping effect of the two phase plate models, that provide some insight on the expected modulation; multislice simulations, used to determine the applicability of ECGI for high-resolution imaging; and finally, some simulated experiments, done to test the various algorithms. Unless stated otherwise, all simulations were made in Python[95].

7.1 Expected Far Fields of the Phase Plates

The two phase plates used for this thesis have been developed at different stages of the thesis work, and with different intended uses. The Double Sorter 1 (DS1) phase plate was, as the name suggests, a variation of the design of a different phase plate, used in [63]. It was the first phase plate with which I was introduced to the topic of Computational Ghost Imaging, and it was supposed to provide the simple but systematic modulation of the electron wave necessary for the technique. The Quadrichopsticks (QC) phase plate was instead designed after we realized the first experimental proof of concept of Computational Ghost Imaging with the DS1 phase plate. Thus, its design was specifically oriented towards high-resolution imaging. In both cases, the preparation of the experimental setups involved making simulations of the effect of the phase plates. Particularly, the beam-shaping effect of these two phase plates is observed in their far field, so for long propagation distances of the shaped wave front (or equivalently, in the back-focal plane of a lens).

The far field propagation of the shaped wave fronts is calculated numerically through wave optics simulations. A simulation begins with a calculation of the phase shift generated by the given distribution of electrodes within the desired field of view of the device. Assuming that the charge on each needle is uniform, it is possible to calculate the phase shift for a single needle using Equation 3.10. Particularly, the phase shift can be calculated for the case of an applied bias voltage of 1V and then rescaled linearly for any other bias value. That is because the potential and the phase shift for a single

needle scale linearly with the charge density. This, in combination with the superposition principle, is quite convenient to compute the phase shift of the phase plate. It is possible to model the phase shift of each needle in its position, biased at $1V$. The individual phases are stored separately in an array. Then, any arbitrary combination of applied biases can be simulated by multiplying the phases calculated numerically by the appropriate set of biases, and summing the resulting phases to get the total phase plate phase. This avoids having to compute the phases from scratch when simulating a different configuration of the device.

Another method to calculate the phase shift imparted by the phase plate is to simulate the projected potential of Equation 3.3 through finite elements calculations, and derive the phase shift accordingly. In that case, the entire device has to be modeled at once, but the previous strategy can be adopted in roughly the same way. Because the potential scales linearly with the applied charge density, it is possible to simulate the full device with just one needle biased at $1V$ and all others biased at $0V$. Then, the phase can be rescaled linearly to other applied biases. Since in this case the whole device must be simulated at once, the best approach is to exploit any device symmetry to reduce the number of different finite elements calculations to be made. The phase of symmetrically arranged electrodes can be derived by rotation/axial symmetry of the phase of one needle.

Once the phase plate shift has been calculated, it is possible to add the presence of an aperture or of the opaque electrodes, if they fall inside the field of view of the simulation in the device plane. This is done by multiplying the complex exponential of the phase by an amplitude term containing a binary mask, set to zero on the regions shadowed by the opaque elements and to one outside. The phase of the device is then propagated to the back focal plane of a lens through a Fourier transform. During the propagation step, it is also possible to include the effects of device charging, spherical aberration, defocus and 2-fold axial astigmatism in the simulations. It is sufficient to add the corresponding phase terms to the device phase before evaluating the Fourier transform. Our model for charging was simple, only accounting for uniform charging of the individual needles. Modeling complex localized charging phenomena would have been beyond the scope of the thesis, as shall be discussed in Section 9.

7.1.1 Double Sorter 1

The simulations for the far field of the DS1 phase plate were done with a C++ code, adapted from preexisting codes written by a colleague in my research group. The algorithm required as inputs the set of biases to be applied to the electrodes and the finite elements calculations of the phases of the electrodes. The phases were numerically calculated in COMSOL Multiphysics [96], a software that is widely used for finite elements calculations of the properties of user-defined systems. It can effectively model electrical, mechanical and thermal properties of a system, along with other aspects that are not relevant for our analysis like fluid-dynamics, acoustics or chemistry.

The device was modeled in a finite elements calculation of the electric properties, that took account of the material of the electrodes as well ($30\mu m$ free standing rods of n-doped Si). The goal of the COMSOL finite elements calculations was to recover the phase shift of each needle at $1V$, so that they could be used as the starting point for the

calculations of the phase plate's phase later on as explained before. Figure 7.1 shows the COMSOL model of the device and the phase of two needles not mutually symmetric. As mentioned before, the remaining four phase images for the other electrodes can be recovered through axial symmetry of the two calculated ones.

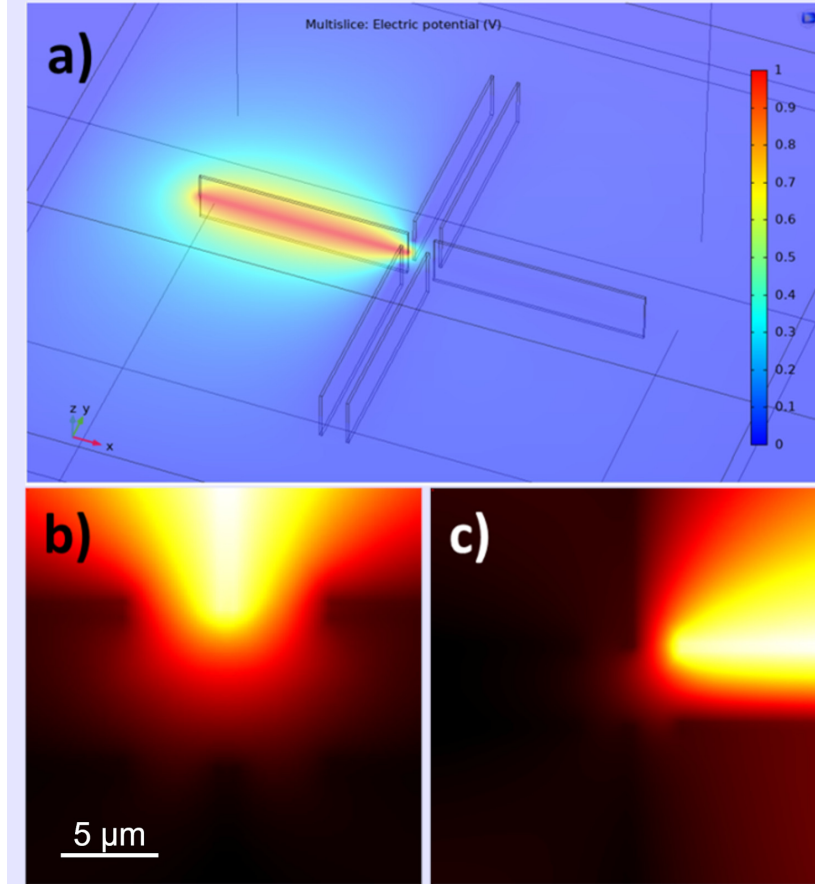


Figure 7.1: (a) Comsol model of the device, displaying one of the electrodes biased to 1V. The colorbar shows the potential varying from 0V to 1V, going from blue to red. (b)(c) phase images for the phase shift imparted by two non-mutually symmetric electrodes. As it can be seen, the remaining phases can be derived by axial symmetry operations on these images.

Figure 7.1a shows the device simulated in vacuum. Each electrode is simulated to be $1\mu\text{m}$ wide and $80\mu\text{m}$ long. There are no boundary electrodes in the simulation, but in our case that is not strictly a problem as the DS1 phase plate is meant to be used under parallel illumination and with a small aperture that only selects the central part of the device, leaving only the very tips of the electrodes inside the region used to structure the probes. Any additional phase effects arising near the phase plate frame are assumed to be too weak to perturb the phase in the central region of the device. The diameter of the aperture in the simulations was adjusted to match the condition observed in the first tests at the microscope.

The simulation procedure is resumed in pseudocode in the following Listing 1. The simulated structured probes are generated starting from a set of randomly generated potentials within a fixed range. It is also possible to load a list of sets of biases from a file, so that they can be used later to generate real probes in the lab and compare

them with the simulations, or vice-versa. These biases, and the simulated phase images of the electrodes, are the input of the simulation algorithm. The algorithm computes the square modulus of the Discrete Fourier Transform of the complex wave function modeled in the plane of the phase plate. The output comes as a 512x512 pixels binary (.bin) image file that contains the aforementioned square modulus. Thus, each image represents the intensity of the structured probe in the plane of the sample. The algorithm can generate multiple structured probes in a single run. The resulting images can be used for simulations that test applications of the SPI algorithms to measurements made using the DS1 phase plate. The choice of the 512x512 pixel resolution was made as a compromise between the need to maintain sufficient detail in the simulated images and the computational costs of using large arrays for the simulations.

```

1  n_elctds = 6      number of electrodes
   patterns = ##### number of patterns to simulate
3  elctd_phases = 0  #initialize electrode phases

5  for (int i=0; i < n_elctds; i++)
   {elctd_phases[i] = single electrode finite-elements calculated phase}

7  electrodes_mask = array(1 on vacuum, 0 inside electrodes)

9  for (int n=0; n < patterns; n++)
11 {biases = load from file/generate a set of 8 potentials
   tot_phase_array = 0 (plane wave)

13     for (int i=0; i < n_elctds; i++)
15     {tot_phase_array += elctd_phases[i] * biases[i]}

17 tot_phase_array *= electrodes_mask
   wavefunction = (electrodes_mask * aperture) * exp(1j * tot_phase_array)

19 propagate to lens back focal plane by Fourier Transform
21 calculate square modulus

23 save image (and potentials) on file}

```

Listing 1: DS1 structured probe simulation algorithm in pseudocode.

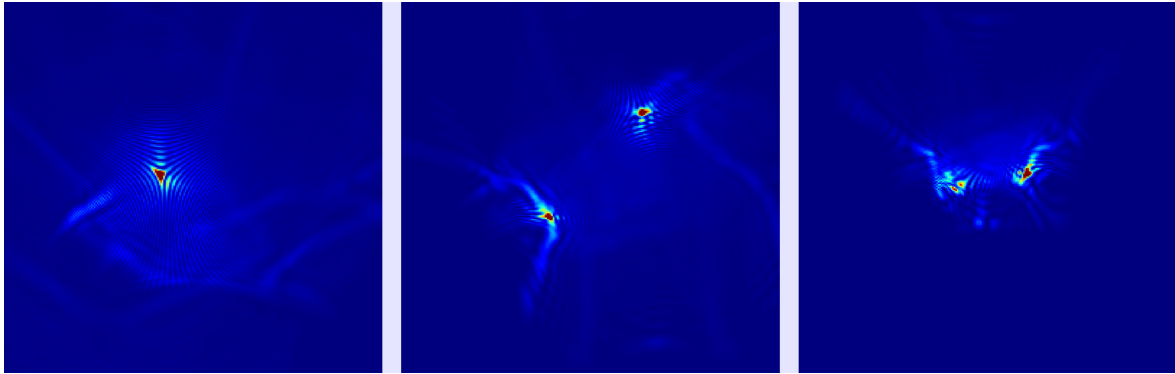


Figure 7.2: Three examples of simulated probes. The images are obtained with sets of randomly generated sets of six biases in the interval $[-2.5, +2.5]V$. The structured probes have been cropped out of the 512x512 simulated images to improve their visibility, normally they occupy about 1/4 of the field of view.

The probes generated by the DS1 are essentially caustic patterns, with a number between one and three spots of high intensity and a sequence of low intensity fringes that extend from the caustic positions to the surrounding area. The position of the caustics depends on the applied biases, but they are generally distributed in an elliptical region

that is related to the geometry of the phase plate. Some examples of simulated probes are shown in Figure 7.2.

It is worth mentioning that the probes generated by this kind of phase plates tend to become larger and larger as the range of allowed biases expands, because of the proportionality between the phase and the projected potential. Another effect is that the extent of the region explored by the structured probes expands as well. This is easily understood if we consider that some randomly generated configurations of applied biases might give an approximate phase gradient across the phase plate, that produces a tilt effect on the wave front. Thus, the larger the range of the applied biases, the greater the chance that some probes are shifted towards the edge of the simulation's field of view. It is best to avoid that condition, because it poses some issues with the normalization of the probes that is required to execute the SPI algorithms. At the same time, if the range of the applied biases is too narrow, the probes will not change much and they will be mostly localized around the center of the field of view, and the reconstructed images will mostly display a flat background. As mentioned in Section 5, one must also remember that when generating the structured probes it would be ideal to have the complete set of probes illuminate a given area in a uniform way, when superimposed. So, the choice of the range of biases used to make the probes becomes quite a relevant step of the simulation/measurement process.

7.1.2 Quadrichopsticks

The Quadrichopsticks (QC) phase plate is the second model of phase plate produced for experiments of Computational Ghost Imaging. It is displayed in Figure 7.3.

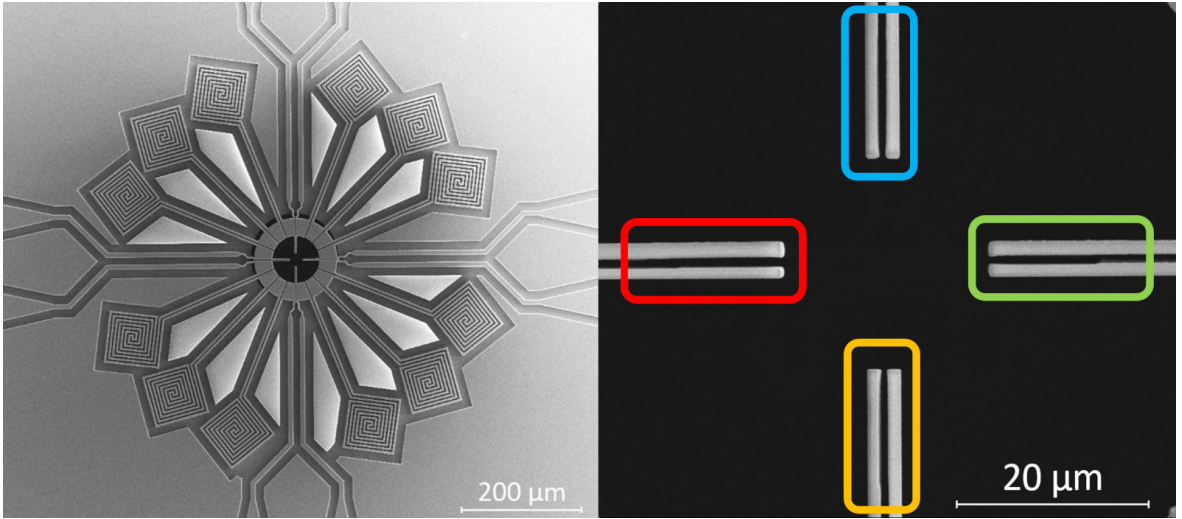


Figure 7.3: (Left) Wide view of the Quadrichopsticks device, showing the central narrow aperture and the SPPs, surrounded by the photolithographed wires and resistive spirals. (Right) Detail of the four SPPs of the QC, each marked with a different color.

In the beginning of this Section it was remarked how the DS1 phase plate was only used for the proof of concept experiments, while the QC is the phase plate designed for atomic resolution imaging. The main issue with the DS1 phase plate is that the structured probes it generates do not resemble atomic lattice fringes. As we explained in Subsection 3.2, that is the optimal condition to be in while trying to make atomic

resolution images. Another issue of the DS1 phase plate is that there is little variability in the structured probes compared to those of the QC, and this would further complicate the measurement process at high resolution. This Subsection will show the simulations of the structured probes that led to the choice of the QC phase plate for the high-resolution measurements.

The QC phase plate is supposed to generate probes with a size of a few nm , that efficiently sample different spatial frequencies of the order of those found in the sample. The QC phase plate was designed with the goal of placing it in the third condenser aperture of a three-lens TEM condenser system. In that way, each lens would perform a separate task for probe structuring. The first two condenser lenses (C1, C2) and their apertures would decide the spatial coherence of the illumination, necessary because the QC works by interference of the vortex beams from each SPP, and its convergence angle, effectively deciding the size of the aperture seen by the device and the consequent diffraction-limited resolution. The third condenser lens' (C3) aperture would house the phase plate and the lens itself would project the structured probes at the sample level. Because of the diffraction limit, and the fact that the structured probes are supposed to be at the nanoscale, with atomic resolution features, the convergence angle becomes a critical parameter of the experimental setup. In the initial simulations, the probes were generated using an aperture large enough to include part of the electrodes. This aspect has relevant implications that became more significant during the characterization of a real device at the microscope, so they will be discussed later, in this Subsection and in Section 9.

One of the greatest advantages of the QC phase plate on the DS1 is that the QC can be biased with a centrosymmetric phase distribution. As a consequence, the phase shift of the QC is almost completely translating into an amplitude effect, because after the propagation the probe wave function becomes real-valued. So, the QC generates probes that are structured in amplitude and not in phase, exactly the kind of modulation needed to make Computational Ghost Imaging, that is an amplitude imaging technique. Of course, to generate several different structured probes there must be a degree of perturbation of the centrosymmetric phases, but it is possible to do so in a controlled way that mostly grants amplitude modulation. In principle, the phase perturbation could also be due to aberrations of the optics in the illumination system, that might alter the phase shift but are not strictly detrimental to the technique as long as the resulting structured probes are known. Of course, aberrations become more relevant when dealing with high-resolution measurements, so it is best to add the effect of the most common ones to the simulation algorithm. In our case, we added the possibility to introduce Spherical aberration, defocus and 2-fold axial astigmatism, that are the ones most commonly found in uncorrected STEM probes.

The simulation code used for the DS1 was modified to be more flexible and include the possibility to update the simulation parameters while running it. I adapted the code to make it usable in Jupyter notebook [97], a web-based interactive Python [95] environment, because it made it easier to change the simulation parameters and plot the results of the simulations. I first wrote a .dll library that contained a modified version of the DS1 simulation algorithm. It allowed importing and executing the C++ code in Python. Later, for ease of editing, the entire code was rewritten directly in

Python. The new code obtained some new features, like:

- a tunable aperture size, to account for the different possible convergent illumination conditions
- the possibility to choose whether to use COMSOL-simulated phases, or compute them according to Equation 3.10
- a simple model of stationary device charging effects
- an additional step to compute the effect of spherical aberration, defocus and 2-fold axial astigmatism (C_3, C_1, A_1)

Listing 2 shows the modified algorithm used for the QC phase plate simulations. Bear in mind that the code was written in Jupyter notebook, so the following pseudocode is not necessarily executed all at once but it can be easily broken down in sub-sections to check intermediate results, like the QC total phase that is sent in input before the propagation.

```

1  define real space and reciprocal space parameters
   (pixel number and size of the field of view)
3
4  r_ap = R          #set aperture size and aperture array
5  aperture = array(1 if r<r_ap, 0 if r>=r_ap)
6
7  n_elctds = 8      #number of electrodes
   elctd_phases = 0  #initialize electrode phases
9
10 create model of the device with either:
11   for (int i=0; i < n_elctds; i++)
   {elctd_phases[i] = single electrode finite-elements calculated phase}
12 or:
13   for (int i=0; i < n_elctds; i++)
14   {elctd_phases[i] = computed i-th electrode phase with analytical formula}
15
16 electrodes_mask = array(1 on vacuum, 0 inside electrodes)
17
18 uniform_charging = 0      #charging phase shift array
19
20 for (int i=0; i < n_elctds; i++)
   {uniform_charging += single electrode phase for the beam-exposed section}
21
22 biases = load from file/generate a set of 8 potentials
23
24 tot_phase_array = 0 (plane wave)
25 tot_phase_array += uniform_charging
26
27   for (int i=0; i < n_elctds; i++)
   {tot_phase_array += elctd_phases[i] * biases[i]}
28
29 aberrations_phase = shp_ab_phase + defocus_phase + astig_phase
30 tot_phase_array += aberrations_phase
31
32 tot_phase_array *= electrodes_mask
33 wavefunction = (electrodes_mask * aperture) * exp(1j * tot_phase_array)
34
35 propagate to lens back focal plane by Fourier Transform
36 calculate square modulus
37
38 save/plot image}

```

Listing 2: QC structured probe simulation algorithm in pseudocode.

Figure 7.4 shows two examples of simulated phases and the corresponding far-field amplitude modulations, obtained during an exploratory study of the QC phase plate. The model used for the phases is the one based of the theoretical formula and not

on the COMSOL simulations. The rounding of the electrode tips that is visible in the images was simulated to model more accurately the real devices, that often end up with rounded tips after Focused-Ion-Beam-based cleaning procedures. Also, aberrations were included, while charging was added at a later step, so it is left out here. The aperture diameter is $50\ \mu\text{m}$, and it is including a large portion of the device. In a realistic setup, it would correspond to large convergence angles, greater than $10\ \text{mrad}$.

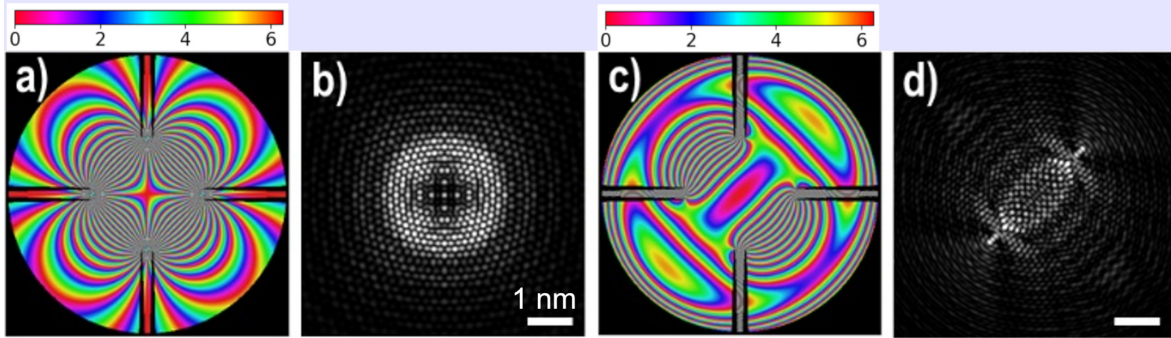


Figure 7.4: Two simulated phases plotted in modulo 2π (color) and the corresponding far-field amplitude modulations (b/w). (a) A centrosymmetric distribution of alternating positive and negative biases of equal modulus on the eight needles produces (b) a finely amplitude-structured probe, with spots that resemble atomic lattice fringes. (c) When spherical aberration is introduced, the amplitude modulation is disturbed but maintains the lattice-fringe-like features (d) in different orientations, thus sampling a different set of spatial frequencies.

When the biases are set to a fixed centrosymmetric distribution and small perturbations of about 15% of the applied biases are randomly applied, the simulations predict that the structured probes would exhibit a lot of fringes and explore a region concentric to the device, like those seen in Figure 7.5. Defocus and astigmatism can alter the extent and shape of the explored region, because they can change the shape of the probes and their features.

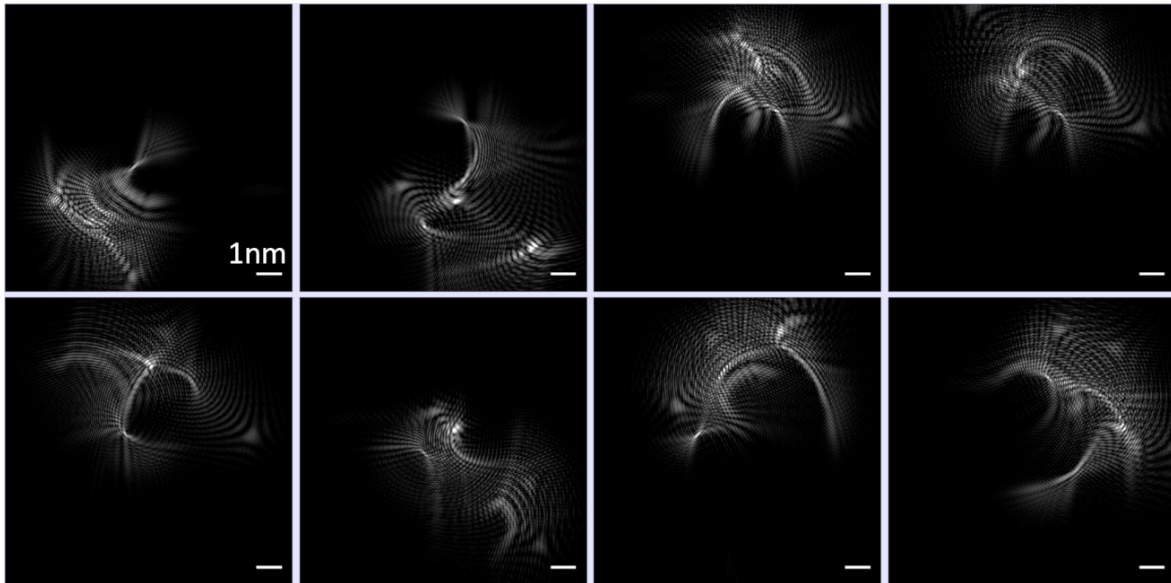


Figure 7.5: Simulated structured probes obtained by random perturbation of the symmetric applied biases. The probes explore a roughly circular region centered on the optical axis of the phase plate.

The probes seen here were simulated with a large portion of the device illuminated by the beam. However, as it was said earlier in this Subsection, the aperture that limits the illuminated region of the device is the C2 aperture. When a C2 aperture is selected, the intensity of the C2 lens can change the convergence angle of the beam, that consequently illuminates only a fraction of the device. In practice, it is the projection of the C2 aperture on the C3 aperture that is deciding the convergence angle of the illumination, thus setting the numerical aperture of the optical setup and the maximum available resolution. The level of detail in the structured probes will change accordingly, and this is also visible in the simulations, as depicted in Figure 7.6.

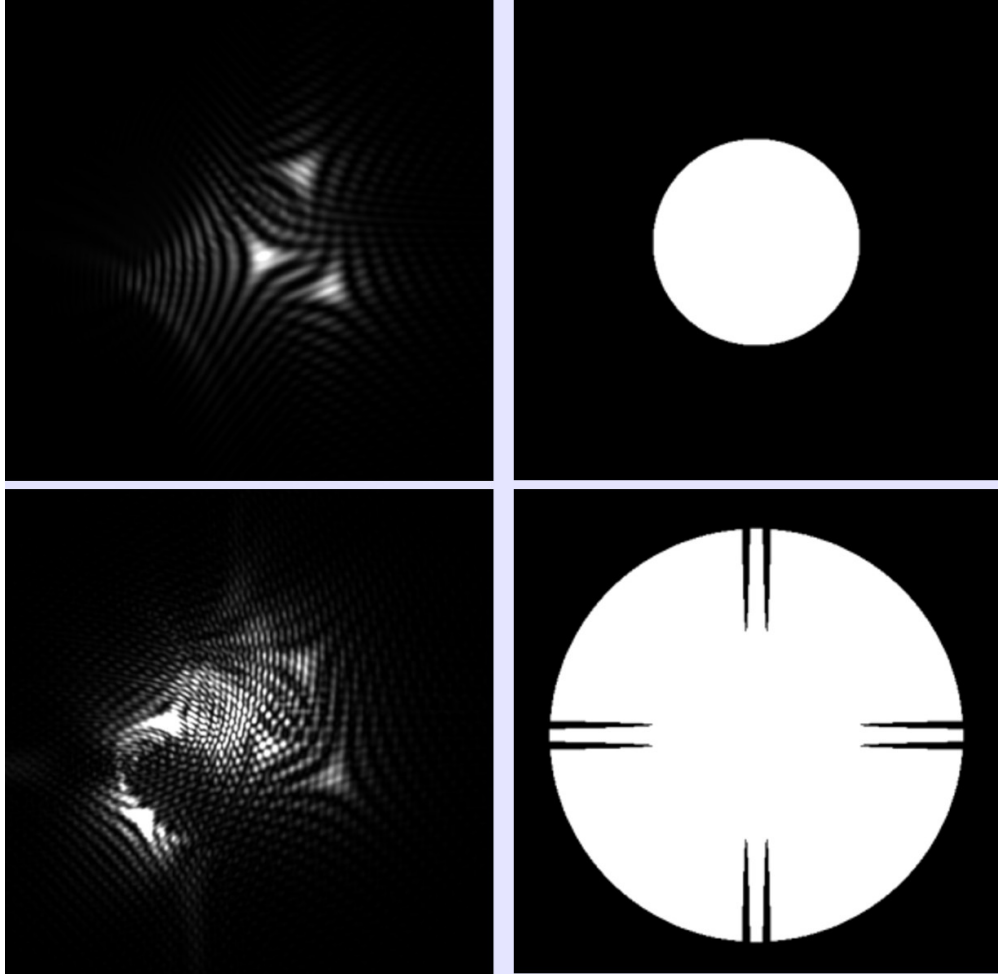


Figure 7.6: Simulated structured probe, in the same exact conditions of applied biases and without aberrations. The size of the aperture is halved, dramatically reducing the number of features and the size of the finer ones. Notice that the use of a larger aperture that includes the electrodes has also brought two additional triangular caustics with large intensity.

As a final remark for this subsection: remember that the QC phase plate can also be described in terms of applied topological charge (or OAM value). Thus, the structured probes can be identified equivalently by the electrodes' biases and the applied OAM value. In terms of simulations, it is easier to describe the probe generation as the process of applying a fixed base OAM value to all the spiral phase plates, and then an additional OAM perturbation on top to make different probes.

7.2 Multislice

The Single-Pixel Algorithms of Section 5 require that the single-pixel signal is obtained through a linear measurement process that satisfies Equation 5.1. For a binary sample, either completely opaque or transparent, the linearity constraint is easily met, because the transmitted intensity is simply the fraction of the total intensity that is not blocked by the sample, so the signal is directly proportional to the overlap of probe and sample. Such binary samples do not exist for high resolution TEM imaging, where electron wave scattering effects become relevant and complicate the wave transmission processes. The proposed setup for Computational GI is based on the use of the HAADF detector as the single-pixel measurement tool, so it was quite important to assess in which conditions the HAADF signal would satisfy the linearity constraint. In the specific context of high-resolution imaging, it was necessary to test the range of sample thickness and detector integration angle for which the HAADF signal would be proportional to the overlap integral of the probe wave function and the sample's function.

The most straightforward way to verify the operation range of the technique was to run multislice simulations of the propagation of the probe inside a model sample. The multislice was run on a thick Si crystal oriented on the $[110]$ direction [98], showing its characteristic pairs of atomic columns. The crystal slices were modeled using the software STEM_CELL [99][100]. The crystal was modeled to occupy a 2048×2048 grid in the (110) planes, with a 17×12 supercell containing equally many in-plane repetitions of the unit cell. Slices were equally spaced, with a separation $t_0 = 1.92 \text{ \AA}$. The Si crystal model had a total of 400 slices, for a total thickness of 76.8 nm , covering a good fraction of the usual TEM sample thicknesses. Figure 7.7 shows the Si FCC cell with a biatomic base, and a part of the crystal model seen along the propagation direction.

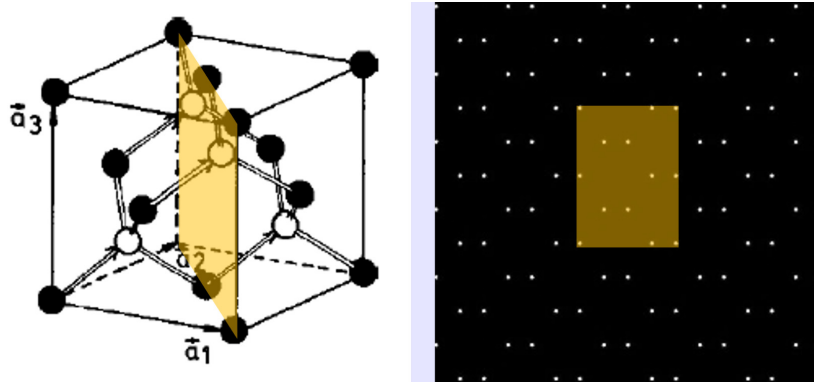


Figure 7.7: (Left) Schematic of the Si FCC lattice with biatomic base, generating the typical diamond crystal structure. In orange, the (110) plane. Adapted from [8]. (Right) Cropped section of the crystal model used for the multislice simulations, seen along the $[110]$ direction.

The structured probes used for the multislice were also simulated on a 2048×2048 grid to match the crystal model's size. The probes were generated from a fixed topological charge configuration with $\ell = 10$ on each spiral phase plate and alternating sign (opposite SPPs having the same sign). The different probes were obtained by adding a randomized perturbation of $\delta\ell \in [-5, +5]$. The probes simulations were ran at zero defocus and astigmatism, and with a spherical aberration $C_3 = 0.5 \text{ mm}$. The simulation generated a small database of 100 probes in total, selected from a larger number.

The selection was limited to those probes that have less than 10% of their total area-normalized intensity within 7/8 of the radius of the simulation's field of view. This criterion was adopted by empirical observation of the probes, to ensure that the simulated probes would remain entirely confined within the field of view of the simulation. Otherwise, part of a probe intensity might fall outside the field of view and alter the analysis. The complex probes' amplitudes and phases were saved separately, because they were required as inputs for the multislice. The probes were ordered by increasing value of overlap with the squared atomic potential of the first slice:

$$OL = \sum_{i,j=0}^{N_{px}} |\psi|_{i,j}^2 V_{i,j}^2 \quad (7.1)$$

Multislice simulations for large images and for many slices are computationally expensive, so the code was run on a high-performance computing server, but only for ten different probes. Figure 7.8 shows the overlap calculation process and a plot of the behavior of the HAADF signal vs travelled thickness for the ten probes.

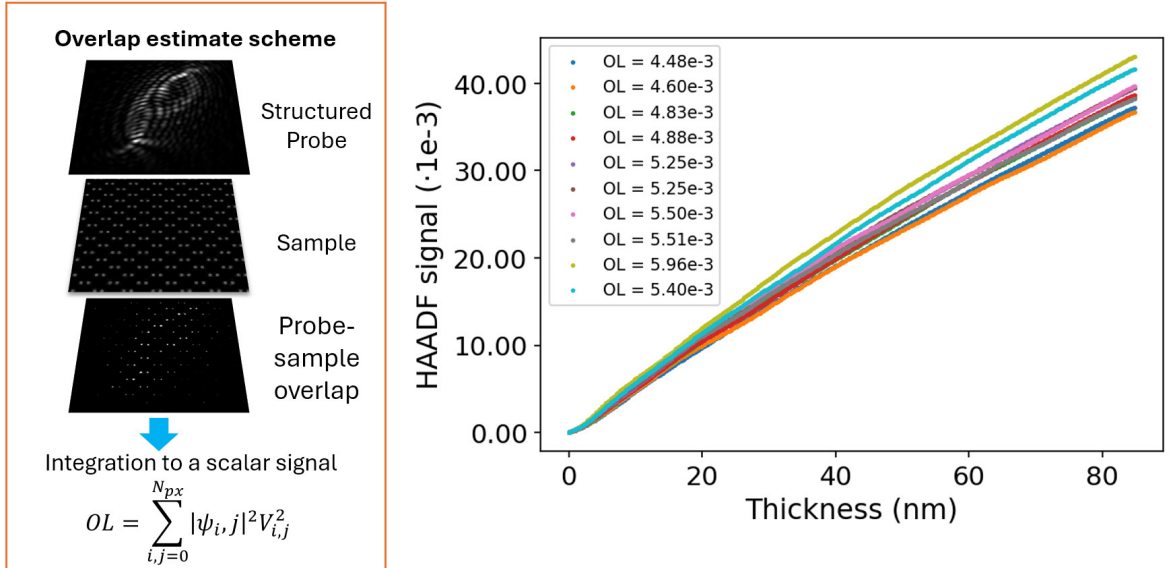


Figure 7.8: Overlap measurement scheme and plot of the HAADF signal resulting from the multislice runs for ten different probes. The HAADF signal is obtained for a collection angle between 50 and 200 *mrad*.

Of the ten probes, eight had different values of overlaps roughly distributed between the maximum and the minimum. The remaining two had values of overlap similar to other two probes, but were visually different and were selected to test for differences in the propagation associated to the different intensity profiles. The multislice algorithm calculated the propagation of the probes for each one, starting from 9 slightly shifted configurations on a 3x3 square grid centered on the original overlap calculation position. The runs include the effect of thermal diffuse scattering effects, averaged over 5 different lattice configurations for each grid position. So, in total, the multislice simulations involved making a total of 10 (probes) * 9 (shift positions) * 5 (thermal diffuse scattering configurations) = 450 propagations of the probes through the 400 slices of the sample. The result is a set of data cubes from which it is possible to extract plots like the one in the right panel of Figure 7.8.

By selecting an integration range, the corresponding HAADF signal can be calculated. The value of the HAADF signal for each probe at fixed thickness can be recovered, producing plots of the HAADF signal vs probe-sample overlap as a function of thickness. Performing linear regressions on each curve recovers the Pearson linear correlation coefficient r vs sample thickness, that can be plotted to reveal the range within which the linear HAADF-signal-to-overlap relation is valid. Figure 7.9 displays some linear fits (again for 50-200 $mrad$ collection angle) and the final plot that identifies the linearity region, where $r > 0.95$, for different collection angles.

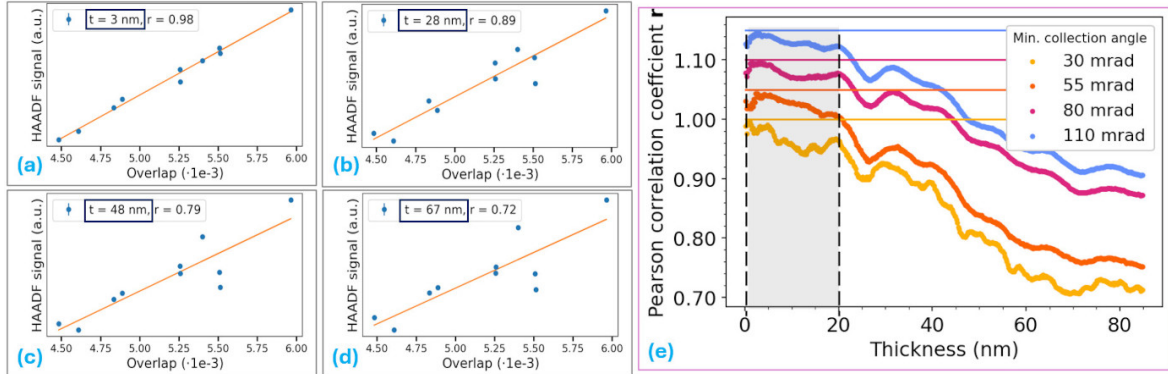


Figure 7.9: (a)(b)(c)(d) Linear regressions of the HAADF signal vs the probe-sample overlap integral at different thicknesses. For each regression, a value of the Pearson linear correlation coefficient r can be extracted. Notice how as the sample's thickness increases, r becomes lower and lower. (e) Linear correlation coefficient r extracted from many linear fits like those in (a-d), plotted versus sample thickness. The plot shows four different curves for four different values of the inner collection angle of the HAADF detector. The outer collection angle is set at 200 $mrad$ for all the curves. The gray box highlights the region where the curves maintain a mild slope, indicating the region of validity of the linear approximation. The curves have been shifted apart by 5% for clarity. The $r = 1$ lines are plotted in the same colors.

The results of the multislice simulations are summed up in Figure 7.9e. It is interesting to see that there is a dip in the value of the correlation coefficient at a thickness of about 20 nm for all the values of inner collection angle, where r drops below 0.95. As a consequence, we identified a maximum allowed sample thickness of 20 nm for optimal validity of the linear measurement approximation. Of course, in the presence of thicker samples the linear correlation coefficient does not get excessively lower, but still it would be best to limit the applications to thin or very thin samples. Another conclusion is that it is best to collect the HAADF signals for larger angles, because the slope of the r vs thickness plots appears to be less pronounced for larger angles, and the curves are smoother. It would appear that the best condition is to stick to inner collection angles larger than 55 $mrad$, because the curve at 30 $mrad$ is oscillating quite a lot.

7.3 Simulated Experiments

Before the realization of the practical experiments we made some simulations of the image reconstruction process for the different experimental setups. Here, we present some of those simulations. Please bear in mind that the simulations are gathered here because of thematic affinity, but they were carried out at different times during the PhD, usually in between experimental sessions. Therefore, they were not conceived all at the same time, but they rather served as tools for a better interpretation of the expected effects of the experimental setup as it was being developed.

7.3.1 Effect of the number of different probes

The first simulations were made in preparation of the proof-of-concept experiment with the DS1 phase plate. The structured probes made with the DS1 are spot-like, as seen from the Subsection 7.1, so it was interesting to estimate a rough number of total different probes generated at random that would produce a reasonably good image in the reconstruction of a binary mask. The chosen target was a 6-spokes Siemens star, a target with alternating circular sectors that are either completely transparent (white) or completely opaque (black). The image's pixel size was not calibrated.

The simulation used a database of 10'000 simulated probes, and the signal needed for the image reconstruction was simulated as the sum of the element-wise product of the simulated images and the target. When the probe images are normalized to their total intensity, this calculation mimics the transmission of a probe through a real binary sample, where the total transmitted intensity is just the sum of the intensities of the parts of the probe that impinge on the transparent sections of the target. The simulated signal and the images of the probes were used to run the TGI algorithm. The results are visible in Figure 7.10. The Figure clearly shows that the basic structure of the sample can be

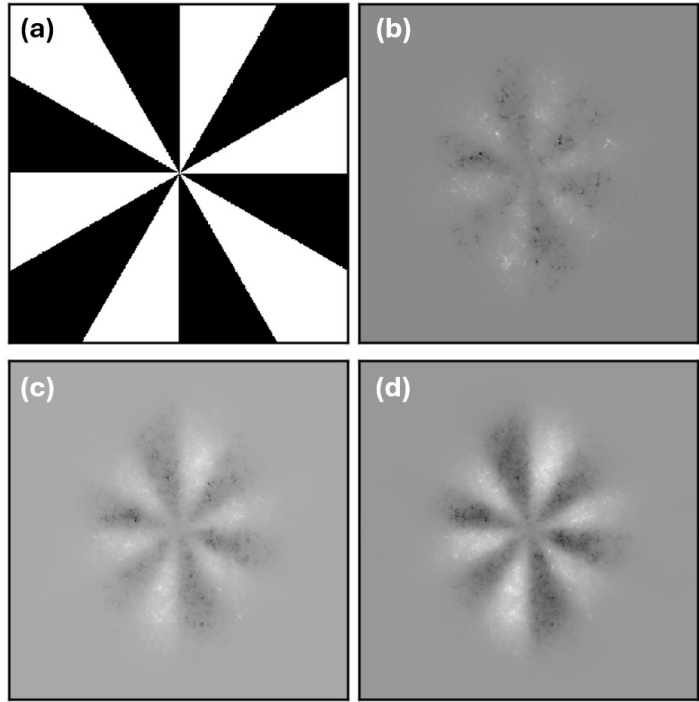


Figure 7.10: (a) Target binary Siemens star. (b) TGI reconstruction with 1k randomly generated probes, (c) 5k probes and (d) 10k probes. easily identified even just after the first 1k probes, but the contrast between the opaque and transparent sections gets better after a few thousand more probes. The details are already visible after 5k probes, the additional 5k only provide a better contrast. These effects are best understood by calculating the radial Michelson contrast:

$$C = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (7.2)$$

If we plot the radial contrast for the three reconstructions, it becomes easier to see that the central region gets better contrast between the 1k and 5k images, while the average contrast is not significantly improved, but then between 5k and 10k probes the contrast almost doubles everywhere. Notice that the contrast for the sample is 1 everywhere, because the sample is binary. Considering all this, we set the goal to the acquisition of a number around 10k different experimental data points for the first experiment. This result also showed

how the highly localized probes made by the DS1 phase plate were not optimally covering the imaged region, providing one additional reason to switch to a different phase plate for the future more ambitious experiments.

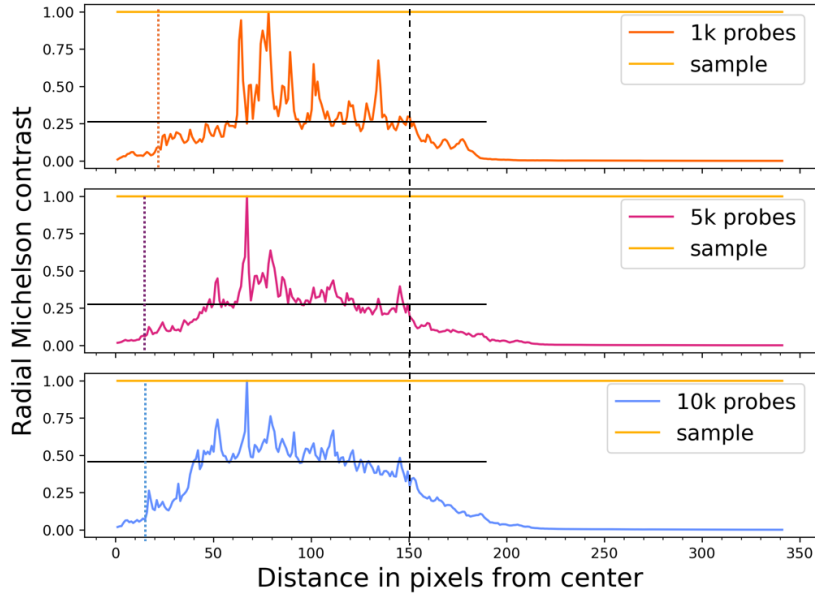


Figure 7.11: Plots of the radial Michelson contrast for the three reconstructed images of Figure 7.10. Notice the increase in average contrast as the number of probes increases, marked by the horizontal lines. The dotted lines mark the onset of contrast in the central region of the images. The dashed line marks the distance at which the contrast begins to fade due to the imaged region elliptical form on the outer edge.

7.3.2 Effect of the scanning on the Traditional-GI and AP algorithms

The hybrid scanning-ECGI method required to modify the code written for the SPI algorithms so that it would include the scanning of each probe on a square raster. Before testing it on real data, we simulated the effect of using different scan steps on a set of low convergence probes generated by the QC phase plate. The probe images pixel size and the scan step were chosen to be consistent with realistic experimental conditions observed in a previous session of phase plate characterization (more details on the calibration method are found in Section 9). The probe pixel size was chosen to be $d_{pr} = 0.177nm$, while the raster scan pixel size was chosen to be $d_{RS} = 0.9808nm$ for a 32x32 raster scan, and it was set to multiples of that value for less fine scan grids (i.e., $4d_{RS}$ for an 8x8 scan).

The database was assembled with 600 different probes. The simulations used a noiseless simulated HAADF signal calculated just like in the previous case, by computing the numerical overlap integral of the probe with the sample at every scan point. The signal and probes were used to run the TGI algorithm with an added scanning procedure in the image reconstruction code. Essentially, the size of the total reconstructed image is larger than the probe images, and can be shifted laterally using Fast Fourier Transforms to make sure that each contribution to the weighted sum of the probes is added at the right coordinates. The result can be seen in Figure 7.12.

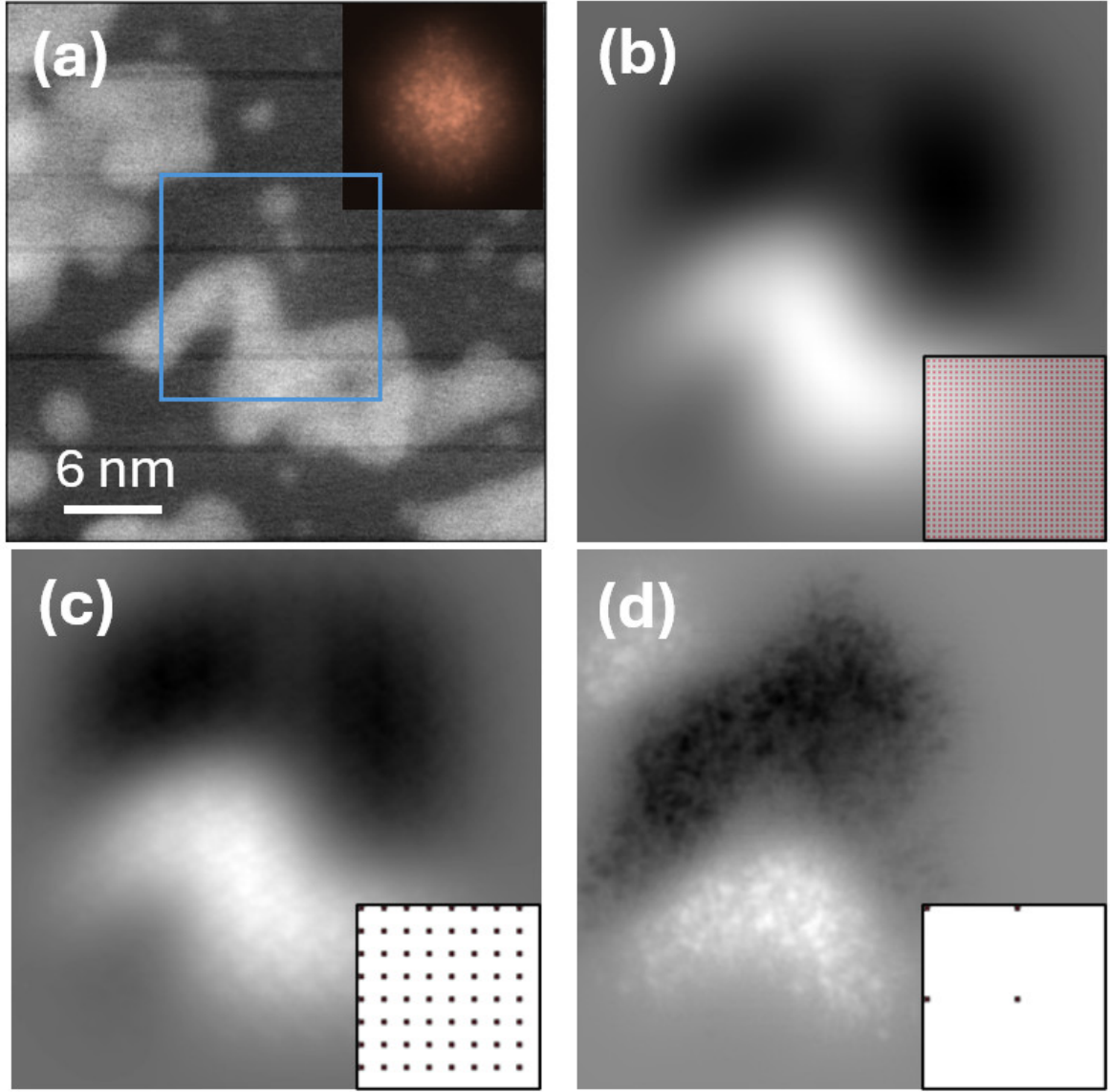


Figure 7.12: (a) Test sample image, an experimental STEM HAADF image of Au islands on an amorphous carbon film. The inset shows the sum of all the images in the database. The blue square represents the region explored by the center of mass of the database during the raster scan. (b) TGI reconstruction with the 600 probes scanned over a 32×32 scan grid, schematized in the inset, and over subsets of that grid of (c) 8×8 points and (d) 2×2 points. Notice how as the number of grid points decreases, the separation between the explored regions becomes more apparent.

The image reconstructions were made over grids of 32×32 , 16×16 , 8×8 , 4×4 , and 2×2 , using the TGI algorithm. This choice was made to save time, because the TGI algorithm is non-iterative and it requires only one cycle to produce the final image. The Figure shows only three of these images, to get a sense of the procedure. Notice that the distribution of the probes in the inset of panel (a) is more or less circular, as expected for the symmetry of the QC phase plate, even though it seems to be slightly vertically elongated. The effect is completely accidental, and it is simply due to the probe generation from random configurations. Notice that, as expected from the hypotheses made in Section 4, as the number of scan points is increased, the gaps between adjacent regions vanish and the reconstructed area becomes square. From the images,

it becomes clear that to avoid gaps in the images, the scan step size for these imaging conditions can be picked at a value between that of the 32x32 scan and that of the 8x8 scan. Of course, the finer the acquisition step is, the larger the number of scan points becomes, and the acquisition time follows. That is true unless the dwell time per scan point is reduced, but then the noise on the signal measurement will increase (just like in ordinary STEM). Another question that became relevant was if it would be possible to select the probes from the randomly generated set and order them according to a parameter that optimized the image reconstruction process. If the reduced database provided similar results to the full database, it would be possible to use a reduced number of different probes for a measurement. Therefore, it would be possible to either make a shorter measurement, or use the time more efficiently, for instance by increasing the dwell time per scan pixel to reduce noise or use a finer grid to acquire more experimental points overall. The criterion selected for the optimization of the probe database was the average self-overlap of each probe with itself for both horizontal and vertical shifting. Figure 7.13 provides a visual explanation of the criterion.

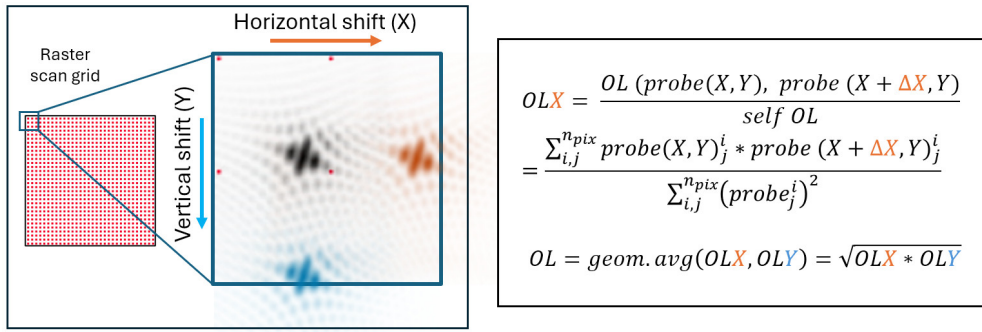


Figure 7.13: Schematic of the horizontal and vertical scan of a probe. The formula for OLX is the calculation of the normalized self overlap upon horizontal shift of the probe. The geometric average OL of the horizontal and vertical shifts self-overlaps provides the final value to use for comparison with the other probes.

The criterion is not very strict, as it breaks down easily if the scan step is too small or too large compared to the average probe size, because in the first case all the probes are obviously overlapping with themselves almost perfectly, while in the second they are so far apart that the self-overlap is always close to zero. However, this criterion worked quite nicely for the experimental conditions of the first scanning-ECGI measurements. Some examples of its applications are shown in the following Figures, where the reduced database was used to run the AP algorithm. The use of the AP algorithm in place of the TGI algorithm for these examples was motivated by the need to have more details in the reconstructed images.

The first test made on the application of the criterion was to determine if a reduced database would actually provide a similar level of detail to that of a complete database. The test was made on the same set of probes and simulated signal used for the images of Figure 7.12. The probes were filtered to those that had a self overlap greater than 60% over the smallest simulated scan step d_{RS} . This reduced the number of probes from 600 to 300, so almost half of them. Running the AP algorithm for 200 iterations both on the full database and on the reduced one for a reconstruction on an 8x8 database produced the results shown in Figure 7.14.

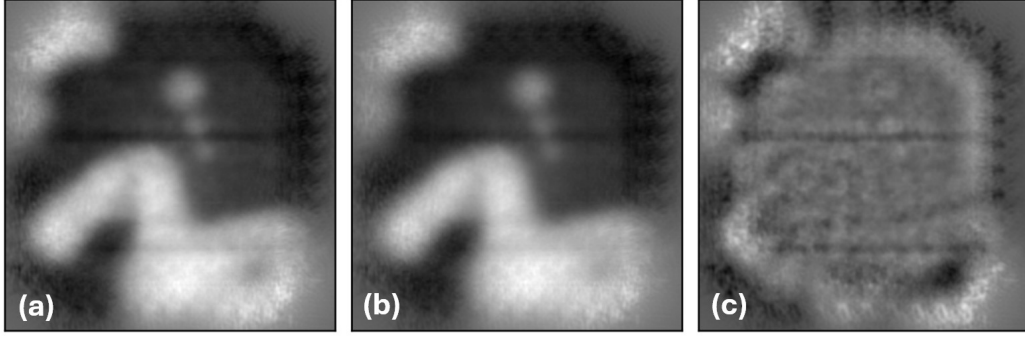


Figure 7.14: (a) AP reconstructed image using the full database of 600 probes and (b) the reduced database of 300 probes for the same number of iterations. Their difference (c) is mostly diverging in the outer parts of the reconstruction, but closer to the center they look visually very similar. The edge features, like the two horizontal lines and the borders of the three aligned particles are slightly sharper in the image obtained with the full database, as testified by the fact that they appear also in the difference of the images, meaning that they have not been recovered to the same level of detail.

First of all, a quick comparison of these images with the previous ones made on the same region makes it evident that the AP algorithm outperforms the TGI one. The comparison in Figure 7.14 is interesting, it shows that most of the low frequency information in the images can be recovered even by using a lower number of probes, in the right conditions. Also, technically speaking, the comparison is not completely fair to the image made with the reduced database, because for iterative algorithms like AP, it should be made between images obtained at an equal number of reconstructed image updates = $(N \text{ probes}) \cdot (\text{scan points}) \cdot (\text{iterations})$. In this case, the probe number was halved in one case, but the two other parameters remained the same. Given this consideration, it became interesting to see if, after reducing the database, it was more advantageous to use less probes and more scan points or vice-versa for an acquisition. So, we produced a second reduced database, containing only probes with a self overlap larger than 70% for a shift of d_{RS} . The second reduced database only counted 100 probes. We compared the previous reconstruction of $(300 \text{ probes}) \cdot (8 \times 8 \text{ scan points}) \cdot (200 \text{ iterations})$ with a new reconstruction of $(100 \text{ probes}) \cdot (16 \times 16 \text{ scan points}) \cdot (150 \text{ iterations})$, for an equal number of total reconstruction updates. The results are presented in Figure 7.15.

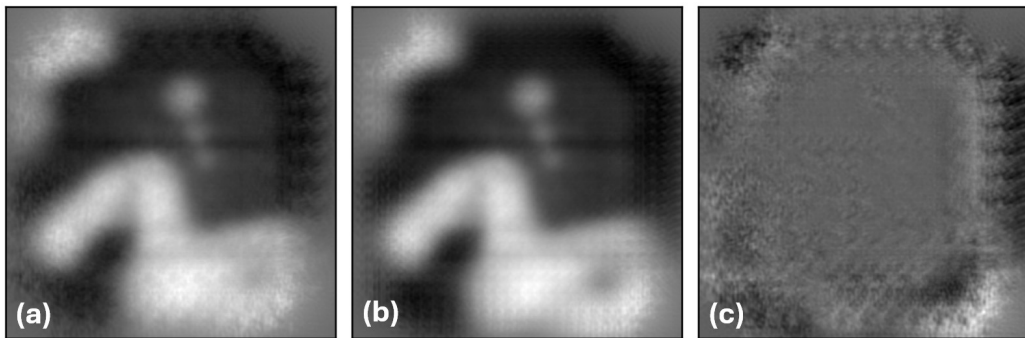


Figure 7.15: (a) AP reconstructed image using the 60% self-overlap reduced database of 300 probes and (b) the second reduced database of 100 probes for the same number of reconstructed image updates. Their difference (c) is close to zero only in the flatter parts of the image, but everywhere else the shift artifacts of the 8×8 scan are quite prominent.

In this Figure it becomes more evident that the introduction of the scanning also produces shift artifacts in the reconstructed images. This problem is actually quite common for iterative methods that process data acquired over periodic scanning, like in Ptychography [28]. There the issue is usually solved by introducing non-periodic scan, but in our case the issue is still of second order compared to the challenge of producing an experimental setup, so it could be investigated in a later work. The two reconstructed images in panels (a) and (b) look once again quite similar. The only conclusion from this analysis was that the use of a lower number of probes and a finer grid would possibly allow to save time for the acquisition, while simultaneously pushing the shift artifacts to higher spatial frequencies. For very fine scans, the shift artifacts might end up to frequencies large enough to filter them out of the reconstructed images by applying a low-pass filter on their FFT, without losing any other relevant information.

Finally, we made a comparison between a 10-iteration AP reconstruction and the TGI one using the 70% reduced database for both and a 64x64 raster scan, to show how for the probes of the QC phase plate it is essentially better to just use the AP algorithm for image reconstruction, as the TGI performs worse. Figure 7.16 shows this comparison, made once again on a Siemens star target like the first simulated experiment.

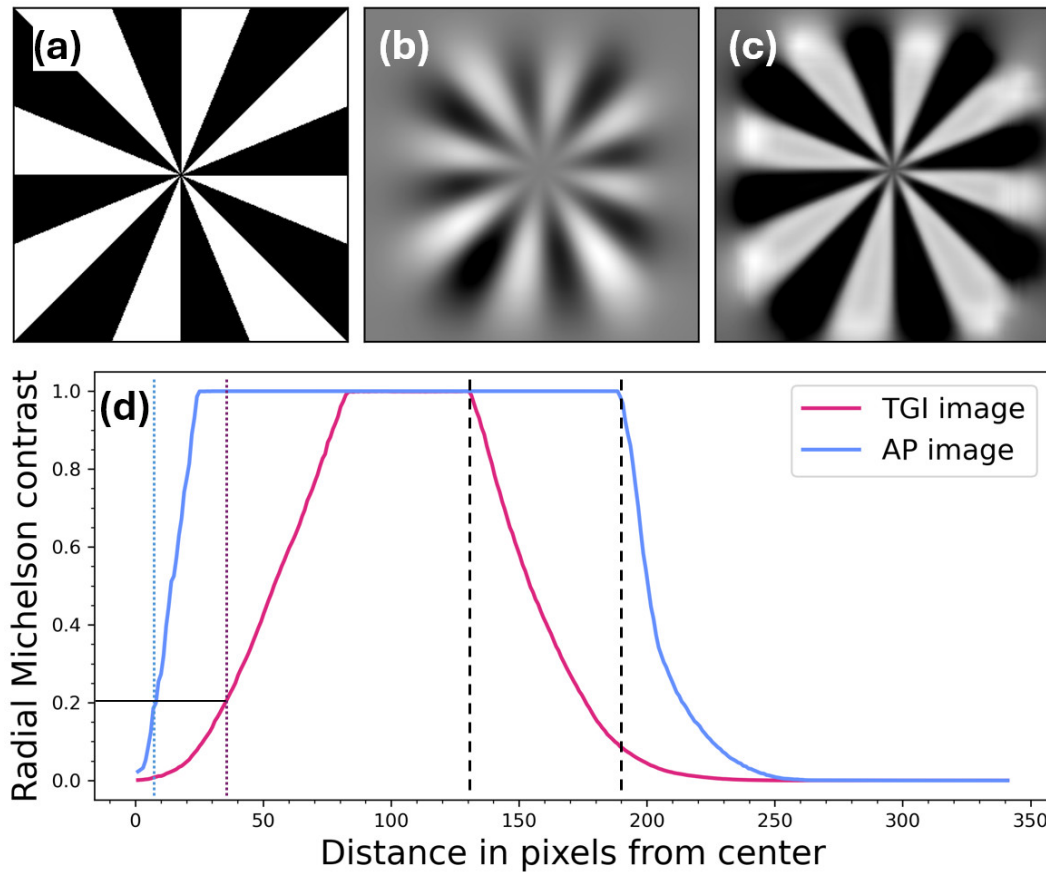


Figure 7.16: (a) Siemens star target and its reconstructions with the (b) TGI and (c) AP algorithms. The radial Michelson contrast is plotted in (d), where it is evident that although both algorithms can reconstruct the sample at medium frequencies, the AP outperforms the TGI both at low frequencies (far from center) and high frequencies (close to the center). It even keeps a non-zero contrast at the central pixels, which is remarkably good. The horizontal line marks the point where the high-frequency contrast drops below 20%.

7.3.3 Simulation of probe reconstruction with ptychography-inspired variation of AP

In Subsection 5.3 it was mentioned that the similarity between the sample image update of the Ptychography algorithm and the sample image update step of the AP algorithm are very similar, so it was possible to write a probe update step for the AP algorithm as well. Some simulation work was devoted to testing this hypothesis by running an algorithm that implemented the formula for the probe update. The input data necessary for a full implementation of an AP-ptychography algorithm are:

- a simulated HAADF signal produced using the "correct" probes to raster scan the target image of choice;
- a set of distorted probes produced by altering the "correct" ones;
- a fine grid for the raster scan, because the information to update the probes comes only from the data points scanned by that probe.

The initial tests were not made on a complete code that implemented both the sample and probe update, because the interest was only on verifying that the probe reconstruction step worked, and the combined effect of both the probe and sample update steps would be more complex to analyze. Thus, the first version of the code only performed the probe update step, using the "perfect" target image in place of the intermediate sample reconstruction. The necessary input data remained the same, but it could be reduced to the use of a single distorted probe and the corresponding raster scan signal produced with the correct one, because there was no need to reconstruct the sample. Ideally, if the algorithm worked properly under these conditions, it would be possible to introduce the sample image update as well.

The simulations were made by using a low-convergence probe from the QC phase plate, distorted to erase most of its features. The perfect and distorted probes are shown in the two panels of Figure 7.17.

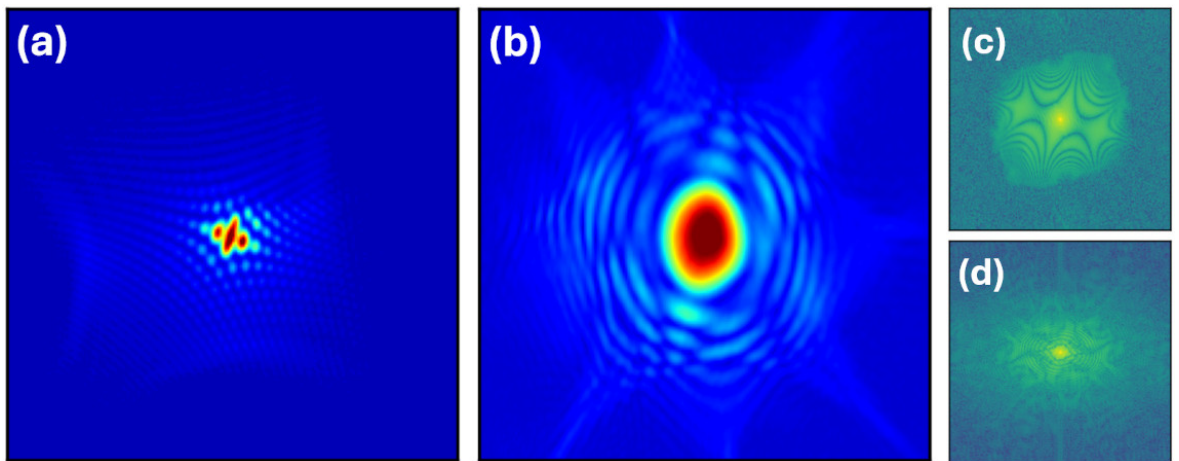


Figure 7.17: (a) Image of the "perfect" initial low-convergence probe. (b) Same probe, after being heavily distorted by adding defocus and astigmatism to it. The insets show the FFTs of the two probes for comparison.

The first attempts at probe reconstruction were made using a 32x32 raster scan with step d_{RS} on the **8-spoke Siemens star** target from Figure 7.16(a), and then a 64x64

scan with step $d_{RS}/2$ made **on the target image of Figure 7.12(a)**. The results are shown in Figure 7.18.

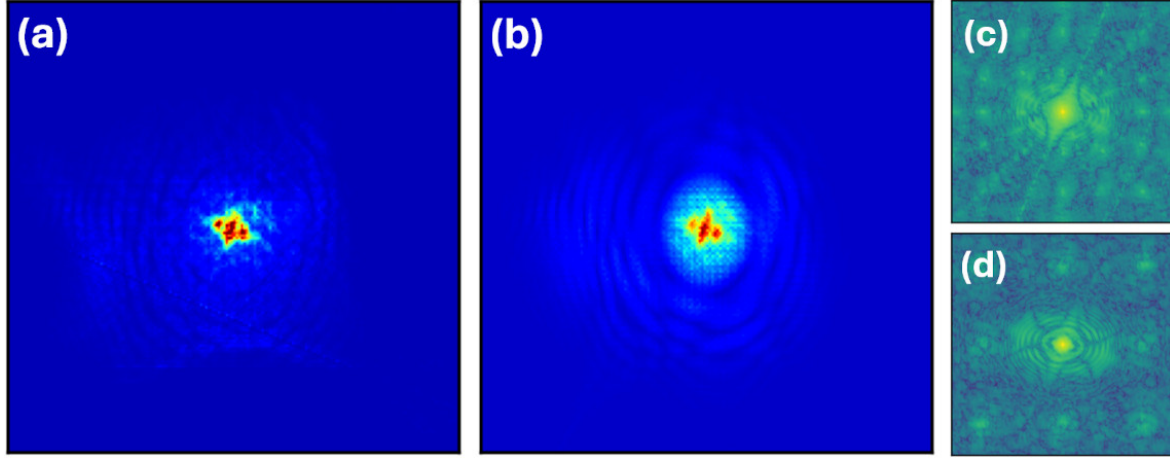


Figure 7.18: (a) Reconstructed probe after 200 iterations of the hybrid AP-ptychography algorithm, starting from the condition of 7.17(b) and using the signal from a 32x32 scan grid on a Siemens star target. (b) Another reconstruction of the probe, over 200 iterations with signal from a 64x64 scan grid on a STEM haadf image. The insets show the FFTs of the images.

Panel (a) shows the result of the first test, after 200 iterations of probe updates on a 32x32 scan grid, for a total of $\sim 800k$ updates, when the algorithm strongly approached convergence. Some of the main features of the original probe were correctly recovered, but the low-intensity fringes of the original probe were not, still showing the ripple-like pattern of the distorted probe. Notice that in the FFT there are shifting artifacts at frequencies compatible with the main features of the original probe, and a second order of artifacts at higher frequencies. Despite these flaws, the result was encouraging. Panel (b) shows the result of the second test, where the sample was changed to the real STEM-HAADF image, thus altering the distribution of the spatial frequencies in the sample used for the reconstruction. The scan step was also halved, to increase the density of the sampling. The reconstruction after ~ 3.27 million probe updates seen in panel (b) is worse than that in panel (a). The algorithm reached convergence in this case as well, but the low intensity features of the distorted probes had not been corrected, and were even more pronounced than in the previous case. Several attempts at code editing were made at this point, checking for possible mistakes in the normalization factors, or other issues. Ultimately, however, the most likely explanation was that the loss of high-frequency information in the sample due to the switch from Siemens star to STEM HAADF image was limiting the recovery of the correct probe, even despite the increased sampling density.

To test this hypothesis, we produced four different versions of the Siemens star target by applying progressively stronger low-pass filters, to make the features of the sample become less and less sharp. Figure 7.19 shows the four versions of the Siemens star used, obtained at four different frequency cutoffs based on the aperture-limited resolution for the illumination conditions of that probe. Specifically, the cutoffs were set at 3, 4/3, 2/3 and 1/3 of the frequency corresponding to the aperture limited resolution of the probes f_{AL} , to progressively deplete the sample of the high-frequency information content that

was presumably needed to recover the finest features of the probes.

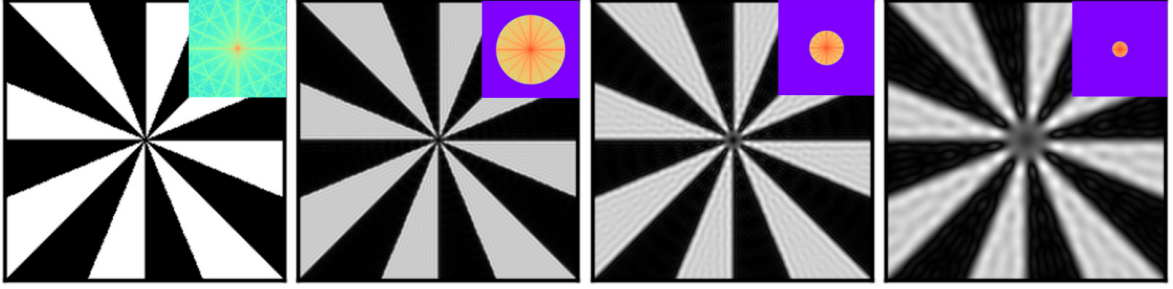


Figure 7.19: Detail of the central sections of the Siemens star target for the four different cutoffs used: (a) $3f_{AL}$, (b) $\frac{4}{3}f_{AL}$ (c) $\frac{4}{3}f_{AL}$ and (d) $\frac{1}{3}f_{AL}$. The insets show the FFTs of the full images.

When the probe reconstruction algorithm is applied to the distorted probe, using a simulated signal made with the perfect Siemens star on a scan grid 64×64 and a sample image from Figure 7.19, produces the results seen in Figure 7.20 after $(50 \text{ iterations}) \times (4096 \text{ scan points}) \simeq 200\text{k}$ probe updates.

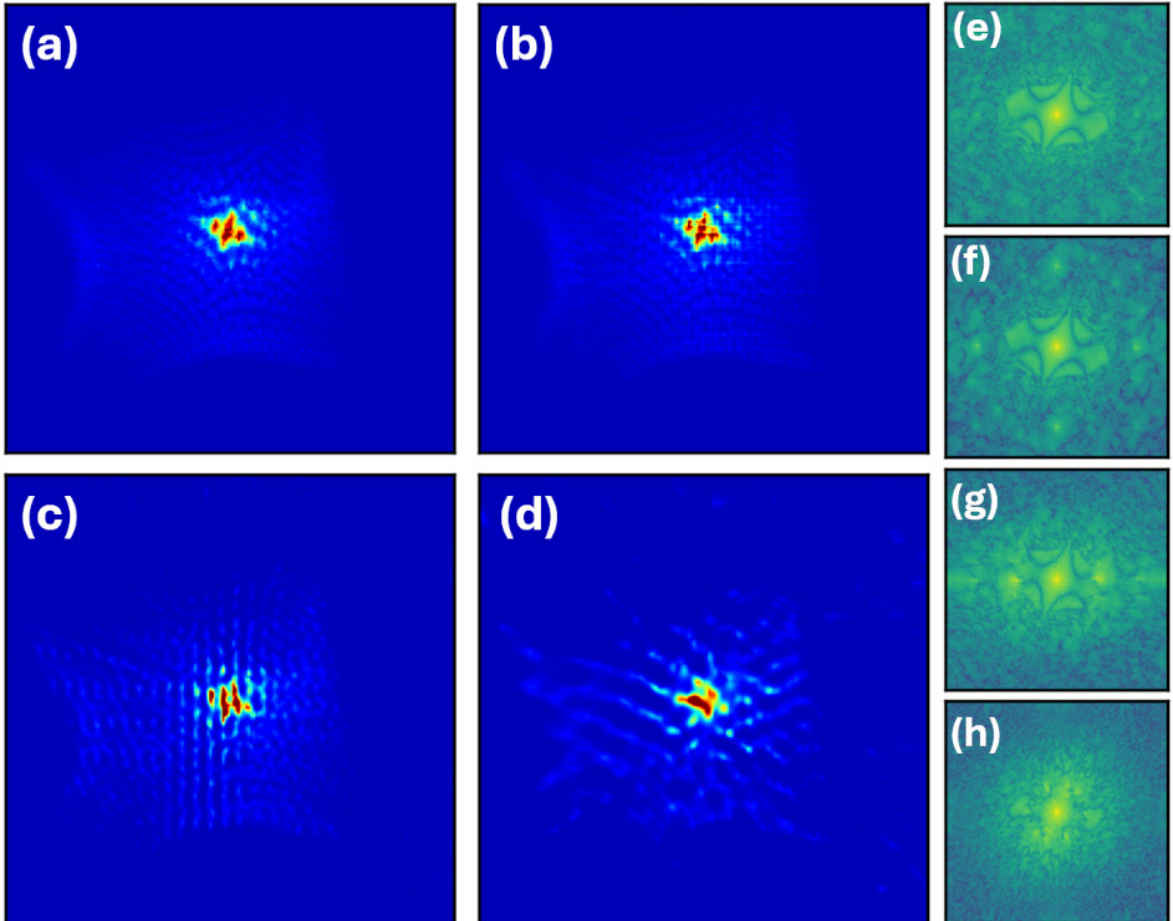


Figure 7.20: Detail of the central sections of the Siemens star target for the four different cutoffs used: (a) $3f_{AL}$, (b) $\frac{4}{3}f_{AL}$ (c) $\frac{4}{3}f_{AL}$ and (d) $\frac{1}{3}f_{AL}$. The insets show the FFTs of the full images.

The results show how the high-resolution features are recovered correctly if the sample

image contains features at those spatial frequencies. Basically, the best reconstruction in panel (a) is obtained with the perfect Siemens star, but panel (b) is also still close to the original. The reconstruction fails when the sample image does not contain details with frequencies up to the probe's resolution limit. That would be a significant problem for an algorithm that simultaneously updates probes and sample, because even for the ordinary AP method, in order to reconstruct fine features in the sample they must already be present in the probes, and the vice-versa is also true. So, during the first iterations, if the probes lacked high-frequency features for any reason, the recovered image would also lack those frequencies, and the algorithm would enter a negative feedback loop, being unable to draw the information it needs from either the probe images or the sample image. This result showed us that the implementation of a ptychography-like method that reconstructs both probe and sample simultaneously is not straightforward, and put a swift end to our attempts in that direction.

7.3.4 Simulation of atomic resolution imaging

The last simulated experiment was done to verify that the high-convergence structured probes made by the QC phase plate could reconstruct the image of a sample at atomic resolution. The sample to reconstruct was a pair of Au nanoparticles simulated in vacuum. That sample was chosen because the first real sample to be used for high-resolution imaging would be the standard TEM alignment sample of Au nanoparticles on a thin film of amorphous carbon. The simulated sample was then a reasonable approximation of the realistic experimental conditions, because the thin-film of amorphous carbon is usually only a few *nm* thick and scatters very little. The two nanoparticles were simulated in two different orientations, on the [111] and [100] crystallographic directions, showing the typical hexagonal and square lattice fringes. I simulated a STEM HAADF image of the nanoparticles using the STEM_CELL software, assuming illumination conditions of 300 *kV* of accelerating voltage and 14 *mrad* of convergence angle. The simulated image is displayed in Figure 7.21.

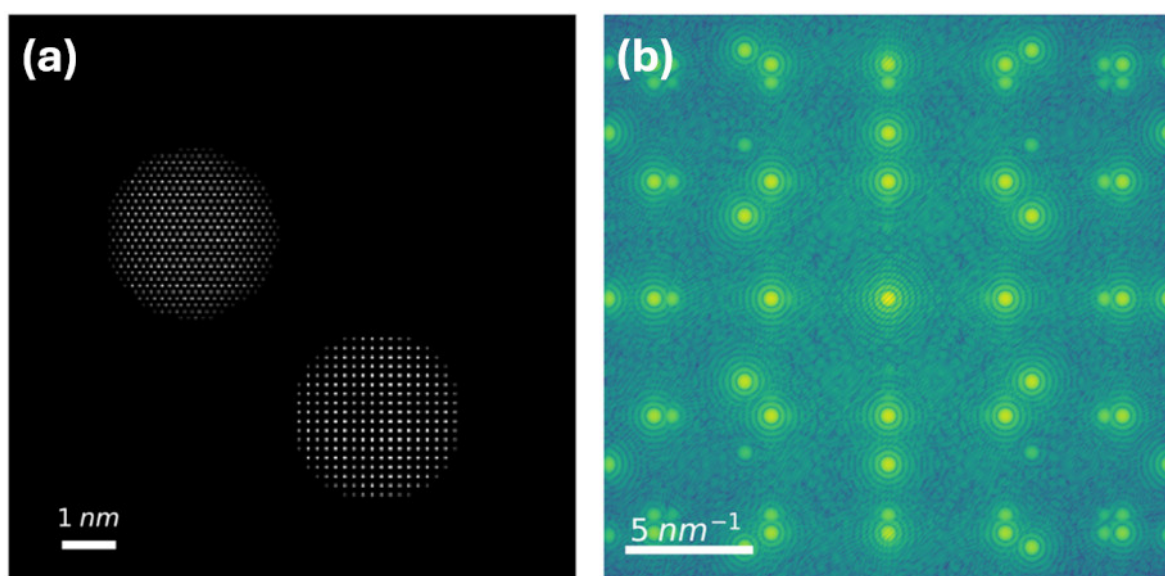


Figure 7.21: (a) Detail of the two simulated Au nanoparticles and (b) their FFT, showing the spots associated to the two different orientations.

Using this image, the HAADF signal needed for the image reconstruction process was simulated as usual, by scanning 250 different probes on the image on a fine 64×64 raster. The scan step size was equal to the simulated image's pixel size, $0.42 \text{ \AA}$. This choice can be replicated experimentally, and corresponds to a condition in which the probes are moving very little compared to the nanoparticles. Using a much larger raster scan would have been pointless, because the surrounding area was just vacuum, so there was no additional information to recover.

The high-convergence probes used for the measurement were generated with a convergence angle of illumination of 7.23 mrad , so the maximum spatial frequency present in the probe images was just large enough to probe the first few spatial frequencies of the Au nanoparticles of the sample. The AP reconstruction algorithm was run for a total of $(100 \text{ iterations}) \times (4096 \text{ scan points}) \times (250 \text{ probes}) \simeq 102 \text{ million}$ image updates. The result appears in Figure 7.22.

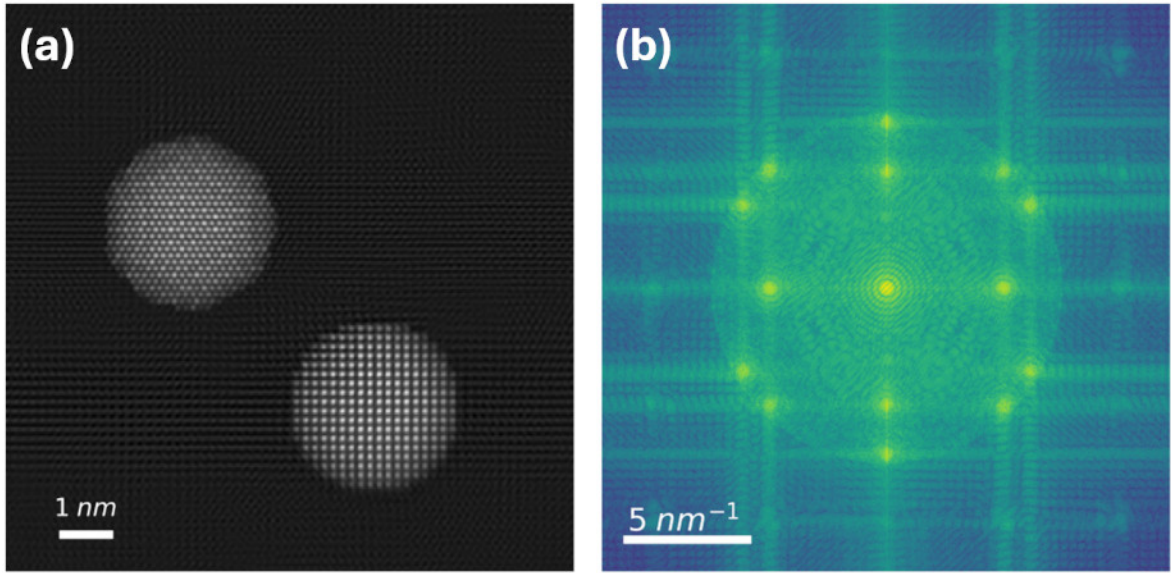


Figure 7.22: (a) reconstructed image of the Au nanoparticles and (b) its FFT, that helps identifying the amount of information recovered from the sample. Notice the horizontal and vertical stripes associated with shift artifacts, and the frequency cutoff associated to the maximum spatial frequency of the probes.

The recovered image is correctly reproducing the lattice fringes of the two nanoparticles, as it is evident even from the FFT. However, there are also delocalized fringes appearing all over the vacuum region. This is possibly an artifact associated to the use of a periodic scan, or maybe a weak point of the AP algorithm when it is used to reconstruct periodic objects. The result of this last simulated experiment showed that under these realistic experimental premises it would be possible to image features at atomic resolution. Of course, in a real experimental setup there would be additional effects contributing to the degradation of the image reconstruction process, like for example noisy data and sample thermal drift. However, the purpose of this specific analysis was to verify that in optimal conditions the QC phase plate would allow atomic resolution imaging.

8 Experimental Setups

This section presents the experimental setups used for the two types of experiments: the microscale imaging setup with the DS1 phase plate, and the nanoscale/atomic resolution imaging setup with the QC. The setups are presented now so that in the upcoming Sections it will be easier to explain the reasons motivating a certain approach or the use of a particular device.

8.1 Microscope Setup - Double Sorter I

The setup that used the DS1 phase plate was planned as a direct transfer of the Computational GI method on a TEM. The microscope available at the time of that work was the FEI Tecnai F20 ST TEM of the CNR-IMM in Bologna. Unfortunately, we did not have a model of biasing aperture holder that would fit in the condenser aperture positions of that microscope, so it was not possible to mount the phase plate in place of the last aperture (C2). Luckily, we also had a single-tilt biasing sample holder (virtually identical to the aperture holder in Subsection 1.6, just with a thicker body). So we were able to perform the experiment by putting the device in the sample position, where a small objective aperture could be used to select only the central part of the phase plate, to match the simulation conditions. This made the setup more flexible in terms of selecting the illumination condition, but also more complex in terms of signal measurement.

Because the phase plate was put in sample position, the sample had to be placed at a different point along the column, in the imaging system, and before the HAADF detector. The natural choice was the Selected Area Diffraction (SAD) aperture position. The SAD aperture is in the first image plane of the microscope, so in order to see the far field of the device in that plane it would be necessary to introduce a large objective defocus. Finally, the setup was designed thinking about using a binary sample like the Siemens star seen in the simulated experiments of Section 7, so the HAADF detector was used to collect the total transmitted intensity, as there was no need to select only the high-angle scattering. This meant setting the detector off-axis, shifting the beam onto it manually through the controls of the imaging system. In that way, the beam would always be completely on the annulus, for any bias configuration of the DS1. Figure 8.1 shows the main components relevant for the DS1 setup.

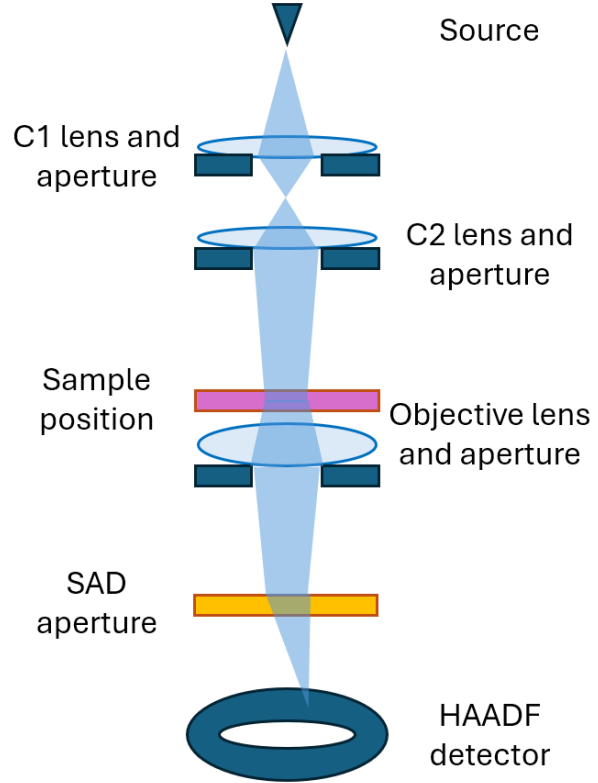


Figure 8.1: Setup for the microscale imaging with the DS1 phase plate. The phase plate (in purple) is placed in the sample position, while the sample (in orange) is in the SAD aperture position.

As one can imagine from the description given so far, this setup was relatively unusual for a TEM, and it had a few drawbacks. First, the system had to be set in STEM mode to be able to read the signal from the HAADF detector, but this automatically turned off the camera, forcing the operator to rely on visual comparison alone between the experimental probes and the simulated ones used for the reconstruction. The large defocus needed to bring the phase plate's far field in the SAD image plane was also another troublesome aspect of the measurement, requiring a very patient alignment of the microscope. The situation was further exacerbated by the use of the HAADF detector in the off-axis configuration, that relied on manually tuning some of the elements of the illumination system (intermediate lens and projector beam shifts). Given the context of proof-of-concept experiment, we endured the hardships of experimental work, but it became clear that the setup had to be revisited to make it more accessible and manageable.

8.2 Microscope Setup - Quadrichopsticks

The setup using the QC phase plate was realized on the FEI Titan G2 60-300 HOLO microscope in the Ernst-Ruska Centrum for Electron Microscopy and Spectroscopy of the Forschungszentrum Jülich, in Germany, that is equipped with three condenser lenses instead of two, and an Image Corrector. The presence of the corrector is not particularly relevant for the measurements of the signal, as it is not affecting significantly the measurements at the level of the HAADF detector, but it can alter the images of the probes seen on the camera. The QC phase plate was placed in the C3 aperture position. The setup of the illumination system then became analogous to the one used for the DS1. However, thanks to the third condenser aperture, the phase plate could fit in the probe-forming system, and the sample could be loaded in its usual position. Another reason for which having the device in the C3 aperture position was good, was that the microscope's scan coils could be used to scan the probes on the beam. With that setup, the alignment of the microscope was much more straightforward than with the DS1, as it was basically a standard STEM alignment made with a large C3 aperture. To make the characterization of the probes and the measurements, the setup had to

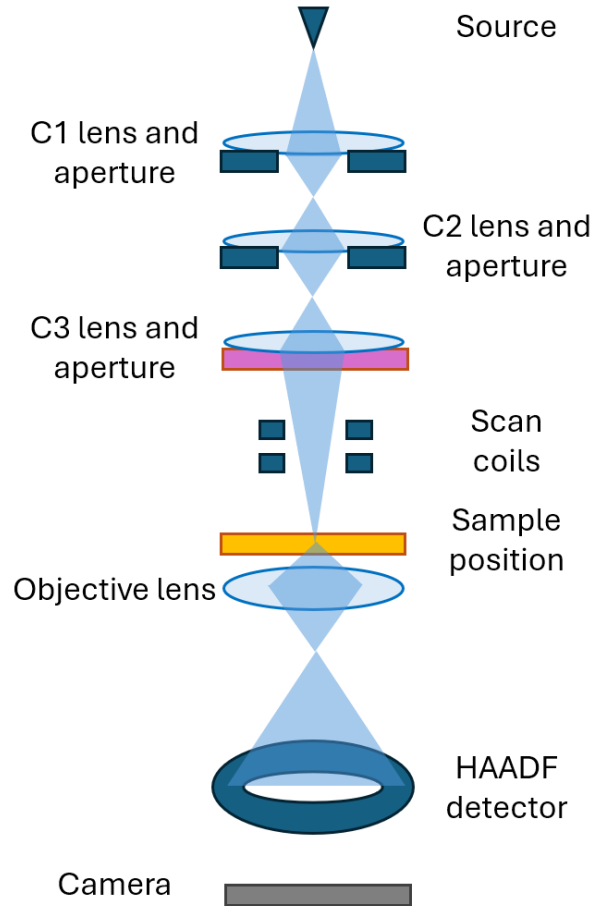


Figure 8.2: Setup for the nanoscale and atomic resolution imaging with the QC phase plate. The phase plate (in purple) is placed in the C3 aperture position, while the sample (in orange) is in the sample position. The scan coils are also relevant, for their use in the hybrid scanning-ECGI approach.

be slightly adjusted after switching the C3 aperture to the phase plate slot. For this configuration, the HAADF detector was used in its proper on-axis aligned condition, because the high-resolution measurements relied on the linearity criterion established with the multislice simulations of Subsection 7.2. An additional component available in this setup was the high-resolution camera of the microscope, that allowed to acquire images of the probes when the microscope was brought to image mode. Being able to see the finer features of the structured probes became quite relevant, because for the high-resolution imaging we could not simply visually confirm that the experimental probes matched the simulations.

9 Characterization of the Phase Plates

An important step towards the full implementation of the procedure is to compare the simulations made for the phase plates with the actual structured probes generated by the real devices. It is necessary to understand whether the simulations offer an accurate model of the structured probes, so that they can be calculated numerically for the reconstructions. This Section describes the characterization process adopted for the two devices. The process basically consists in having the device placed in the aperture position of choice for the experimental setup, and then applying a few combinations of biases to generate the structured probes. The probes have to be imaged and stored for later analysis, which will involve making a comparison between the simulations and the experimental images.

9.1 Double Sorter I

The DS1 phase plate makes structured probes that are mostly point-like, as seen in Figure 7.2. It was the first device to be characterized, so the entire procedure was limited to two main steps. First, finding an optimal range of applied biases that would have the probes spread over a more or less uniform region, not too small. Second, compare the simulated images to the experimental ones, to verify the correspondence. This is a complex task, because it involves accurately modeling a wide range of phenomena that might alter the structured probes, like localized charging or defects in the fabrication of the electrodes. Note that for the DS1, the characterization and the experiment were realized on two different microscopes. The characterization was carried out on the Thermo Fisher Scientific Talos F200S G2 microscope at the University of Modena and Reggio Emilia.

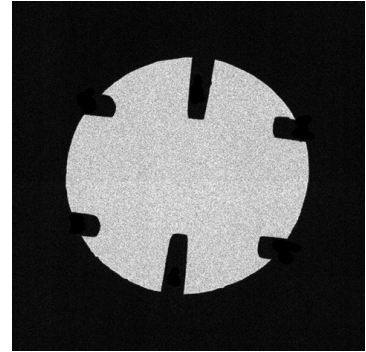


Figure 9.1: TEM image of the DS1 phase plate in the sample position, limited by a $20\ \mu\text{m}$ objective aperture.

The first step was the selection of the operation bias range of the phase plate. The DS1 was put in the microscope's sample position, as indicated in Section 8. The microscope was aligned for a Low Magnification (LM) STEM measurement, because the illumination had to cover the entire device (a few tens of microns) at the *sample* level. The C1 lens was set at spot size 6, to increase the coherence of the source. The intensity of the C2 lens was set to 100% to have a planar wave front at the level of

the device. We used a $10\ \mu\text{m}$ objective aperture to select only the central region of the phase plate. The camera length was set to a large value (few hundred meters) to see the phase plate's far field on the camera. Figure 9.1 shows an image of the central portion of the device, limited by a $20\ \mu\text{m}$ objective aperture to show the presence of the device's tips, that would otherwise be hidden.

The optimal range of applied biases was identified by setting the camera acquisition time to a large value and generating a large number of structured probes in a short time, to see the range covered by the superimposition of all the patterns. The probes are generated by randomizing the values of the applied biases for each electrode within a range $[-V_0, V_0]$. This method gives a rough idea of the extent of the illumination area on the camera as a function of the maximum applied bias. The procedure can be repeated by looking at the individual probes, to make sure that there are not any features that escape the field of view. With careful tuning, the optimal range was established to be $V_{bias} \in [-2.5, +2.5]V$.

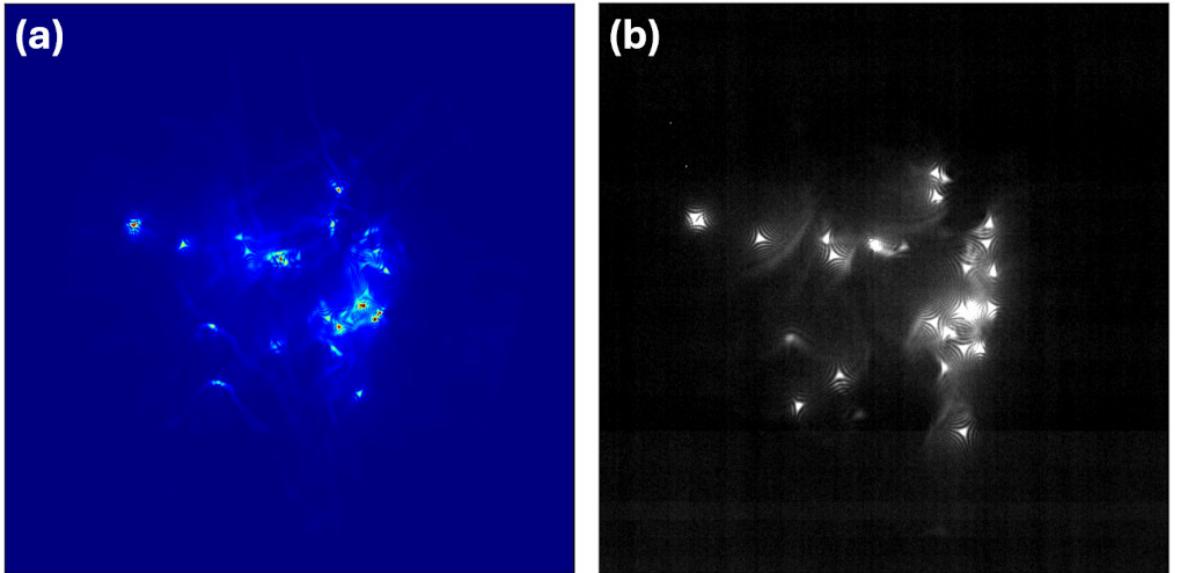


Figure 9.2: (a)Sum of 15 simulated probes and (b) its counterpart, sum of the corresponding 15 experimental images. It is evident that the general distribution of the probes is the same, but there are differences. For instance, the "triangle" in the bottom-left corner of both images has a different shape, because the three caustics are not exactly in the positions predicted by the simulations.

The second step is to make a comparison with the simulations. It is quite difficult to get an exact match between the simulated and experimental probes, because there might be secondary effects that distort the finer features of the probes, or slightly change their position. There might also be a rotation between a simulated probe and an experimental one, partly due to the rotation effect of magnetic lenses and partly due to the natural angle between the sample holder (so, the phase plate's reference horizontal axis) and the camera used for the acquisition of the experimental probes. A direct comparison of each probe with its simulated counterpart would be time consuming, and it would make it more difficult to account for effects that are present on all the simulated probes at the same time (like the aforementioned rotation). A quicker way to make the comparison is to sum the experimental images of a few probes in a

single image and do the same with their simulated counterparts. For the DS1 phase plate, the result is a group of caustics that form a "constellation". Comparing the salient features of these two images reveals the presence of differences between the simulations and the experimental images in a more immediate way. Figure 9.2 shows the aforementioned comparison.

At the stage of the realization of the first experimental setup, these tiny discrepancies between the simulations and the experimental images were dismissed, as the setup of the real experiment with the DS1 phase plate we did not have the possibility to record the structured probes used for the experiments, so we had to rely on the imperfect simulations to complete the image reconstruction process. Given the nature of the structured probes, that were just one or two caustics moving around a central position, it was not necessary to be too accurate with the simulation process. However, once we moved on to the QC phase plate, it quickly became apparent that the task of simulating the device would be much more difficult in that case, because of the need to control the fine features of the structured probes.

9.2 Quadrichopsticks

The characterization of the QC phase plate was carried out on the same microscope used for the experimental setup, the Titan-HOLO in Jülich. The procedure involved similar steps to that of the DS1, with the addition of the initial electron holography imaging of the device. Remember that the QC is meant to work by interfering four vortex beams generated by the Spiral Phase Plates. Ideally, the four SPPs would be identical, generating the same topological charge for a given applied bias difference between the electrodes, as expressed in Equation 3.11. In practice, small differences in the fabrication process introduce some variability in the strength of each one. So, the first step for the QC characterization was to do some Low Magnification Off-Axis Holography to measure the phase of the vortices generated by each SPPs.

To make off-axis holography on the SPPs, the QC was mounted in the sample position, and the microscope was prepared by going into Low Mag TEM mode. The objective was turned off to achieve a field free condition at the level of the sample, and the microscope was aligned with the usual operations. Then, the intensity of the C2 lens was decreased to spread the illumination. The sample was put in focus. Finally, the electron biprism was inserted and polarized to produce the interference field. Holography requires to interfere an object wave, that has interacted with the sample, in this case one SPP, with a reference wave that propagated through vacuum. This already presented an issue, because when the device was inserted there was not any vacuum region that was far from another section of the phase plate. So, the recovered phase might have been affected by some perturbations of the reference wave. There were, of course, empty regions of the device that were large enough to be overlapped with the SPPs regions using the biprism, but there was no way to have a pure vacuum reference wave. For this reason, the phase difference across each SPP was estimated just a few microns from the tips of the pair of electrodes, in a position that was brought to overlap with the vacuum region in front of the opposite SPP. This was the only possibility due to the geometry of the device.

I collected a series of eight holograms for each SPP, for eight values of biasing, four for "direct" bias and four for the "reverse" bias condition (opposite sign on the electrode pair). This produced a total of 24 holograms. I also collected reference holograms of the unbiased SPP, with all the electrodes set to the ground voltage $V_{gnd} = 0V$, and the device kept in the same position as when the holograms were measured. This meant acquiring an additional set of 24 holograms. In principle, four would have been enough, one for each SPP, but the device was drifting over time and it was best to acquire a few more holograms for safety.

The use of these 0V bias reference holograms removed the perturbations of the reference wave due to the presence of the device, as per Equation 1.28, but in doing so they only allowed measuring the phase shift due to the biasing of one of the SPPs, and not the actual total phase shift relative to pure vacuum. So, the phase recovered from the processed holograms are not actually representing the phase shift imparted by the phase plate as a whole, but only the phase shift from the grounded condition. It is a "relative" phase measurement, not an "absolute" one, meaning that the phase holograms of the device might actually exclude some distortions due to localized charging, impurities on the needles, or other effects. They can inform us about the effectiveness of each SPP to impart a given topological charge when biased, but they can not be taken as the actual phase shifts of the device. Figure 9.3 shows a hologram and its 0V reference for one of the SPPs.

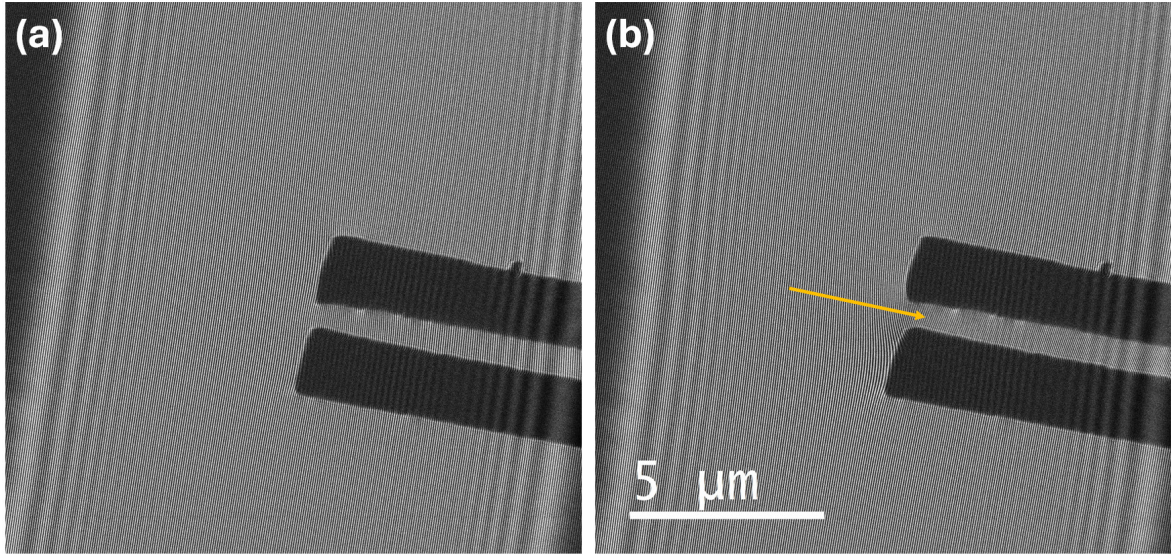


Figure 9.3: (a) Low mag hologram of one of the SPPs, unbiased. (b) Same hologram, but with an applied bias. It is possible to notice a slight bending of the fringes of the hologram across the direction marked by the arrow. It is also possible to notice a tiny spec of dust tat deposited on the side of one of the needles. It was removed before inserting the device in the C3 aperture holder.

The holograms in Figure 9.3 were used, together with the other 23 pairs, to recover the phase shift produced by each SPP when biased with a certain ΔV_{bias} . The topological charge ℓ is simply proportional to the number of steps of 2π seen on a closed loop across the SPP. The easiest way to determine the topological charge is then to measure the phase difference across the phase plate. The phase shift was measured from the images

of the unwrapped phases, in a position a few microns away from the tips, exactly at the same distance for all the SPPs. A couple of examples are visible in Figure 9.4.

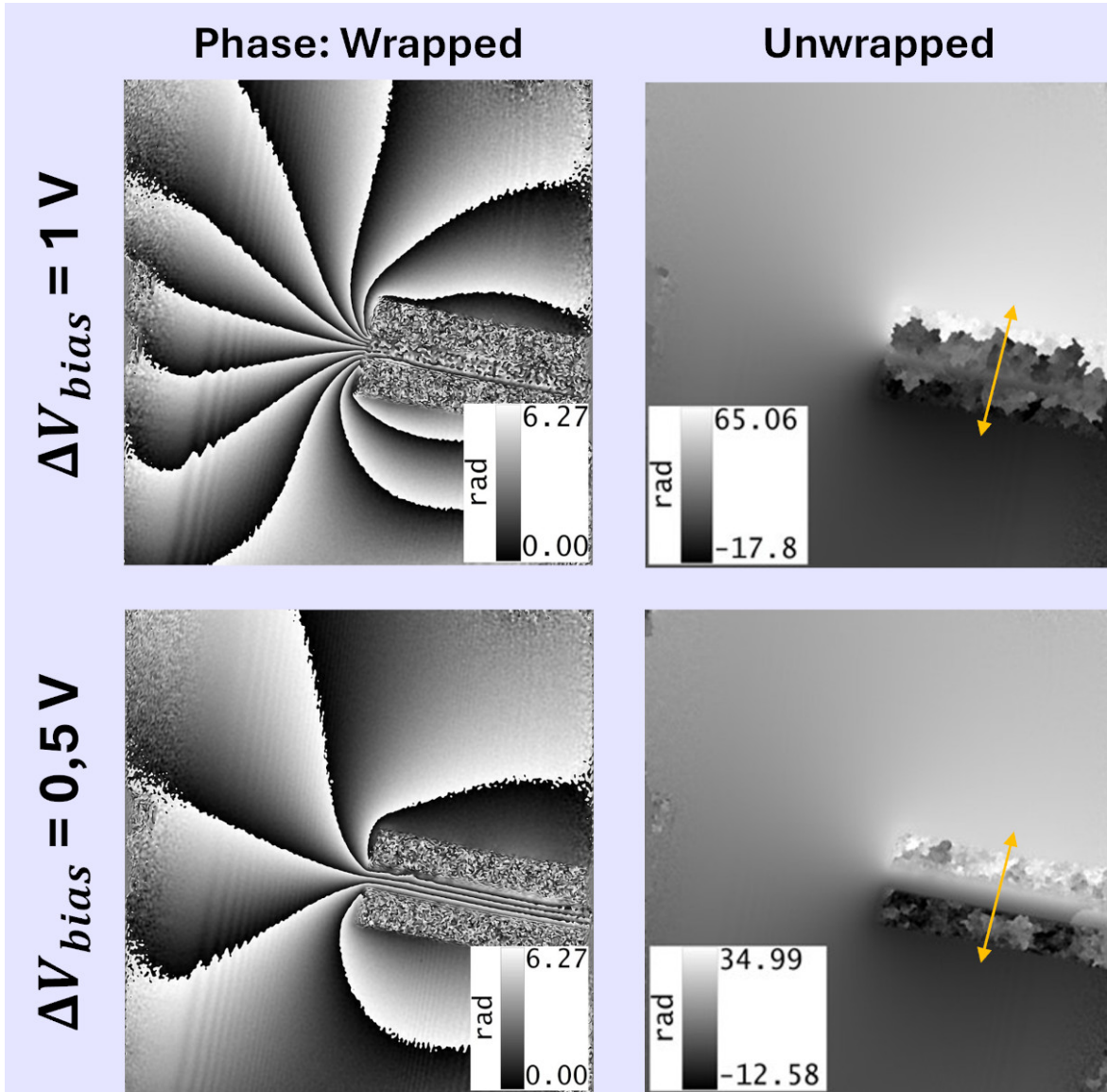


Figure 9.4: (a) Wrapped phase of the SPP in Figure 9.3. (b) Unwrapped phase, with the position to measure the phase difference marked by the orange arrow. (c)(d) Same, but for a different value of bias voltage. Notice that the phase lines are not radiating perfectly straight from the tips, and are not too uniform close to them. This can introduce errors in the measurement of the phase shifts.

The values of the phase shifts and biases were noted, and then used to make linear regressions of topological charge ℓ vs applied bias ΔV . The slope of the linear regression would represent the ℓ -to- V proportionality coefficient. Figure 9.5 shows the fits and corresponding results. On average, the SPPs of the device can apply $\sim 13.9\hbar/V$ topological charge per Volt. Each SPP behaves in a slightly different way, as expected, with some being more efficient in generating ℓ than others. This shows once again that, just like for the double Sorter I, electrostatic beam shaping is very sensitive to the device conditions, and matching the experimental results with simulations can be very challenging.

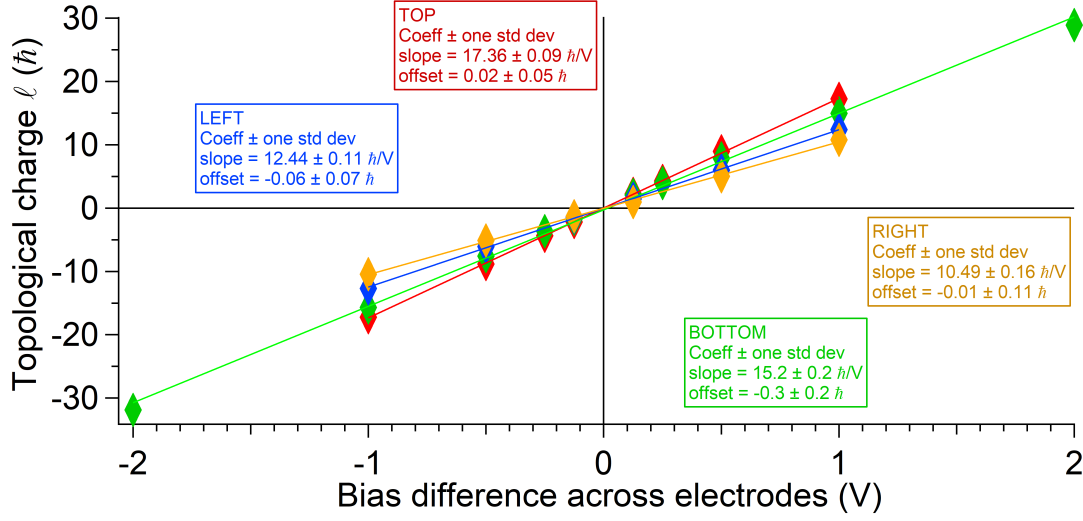


Figure 9.5: Plot of the linear regressions of ℓ vs ΔV_{bias} for all four SPPs. The names are associated to the position of each SPP in the device, relative to the holder. The *bottom* SPP is the closest to the holder handle, the others follow.

After finding the proportionality factor between topological charge and applied bias, the next step was to characterize the structured probes themselves. The phase plate was mounted in the C3 aperture holder of the microscope, following the setup of Figure 8.2. The QC requires high spatial coherence to ensure that the vortex beams of the four SPPs can interfere, so the alignment of the microscope used the highest spot size value available (i.e. smallest source size), spot size 11. Also, the alignment was prepared with a $50 \mu m$ aperture in the C2 position, for the same reason. Under these conditions, with the microscope set to STEM nanoprobe mode, it was possible to change the convergence angle of the illumination of the device in the range $\alpha \in [1.4 mrad, 7.23 mrad]$, but the effective limits were set by the geometry of the device. In fact, under these conditions, it was not possible to go below $1.8 mrad$, as that was the limit to see the tips of the device at the very edge of the aperture. Also, the ring of wiring of the electrodes that are visible around the device started appearing inside the ring at $7.23 mrad$, so that was the outer limit to avoid seeing amplitude effects from other elements different from the SPPs.

Changing the convergence angle changes the features of the structured probes, as seen in the simulations. The difference in illumination convergence angle also determines a different diffraction limited resolution. The characterization was made for two different values of convergence angle: *low* convergence for $\alpha_{low} = 1.8 mrad$ and *high* convergence for $\alpha_{high} = mrad$. Using Equation 1.11, the diffraction limited resolution is $R_{low} = 6.64 \text{ \AA}$ and $R_{high} = 1.65 \text{ \AA}$. The setup of the QC actually allowed a precise calibration of the size of the field of view explored by the probes. The method used for calibration consisted in preparing a conventional STEM alignment in the desired condition, on the standard TEM grating of Au nanoparticles on amorphous carbon, and then inserting the phase plate. Turning the microscope from diffraction mode to image mode allowed to image the probes on the microscope's camera. Then, by tuning the applied biases, it was possible to generate a spread out structured probe, with some regions of flat

intensity where the details of the Au lattice fringes appeared. These features could then be used to calibrate the pixel size of the camera for any image magnification step where they were visible, by means of crystallographic characterization. An example is shown in Figure 9.6.

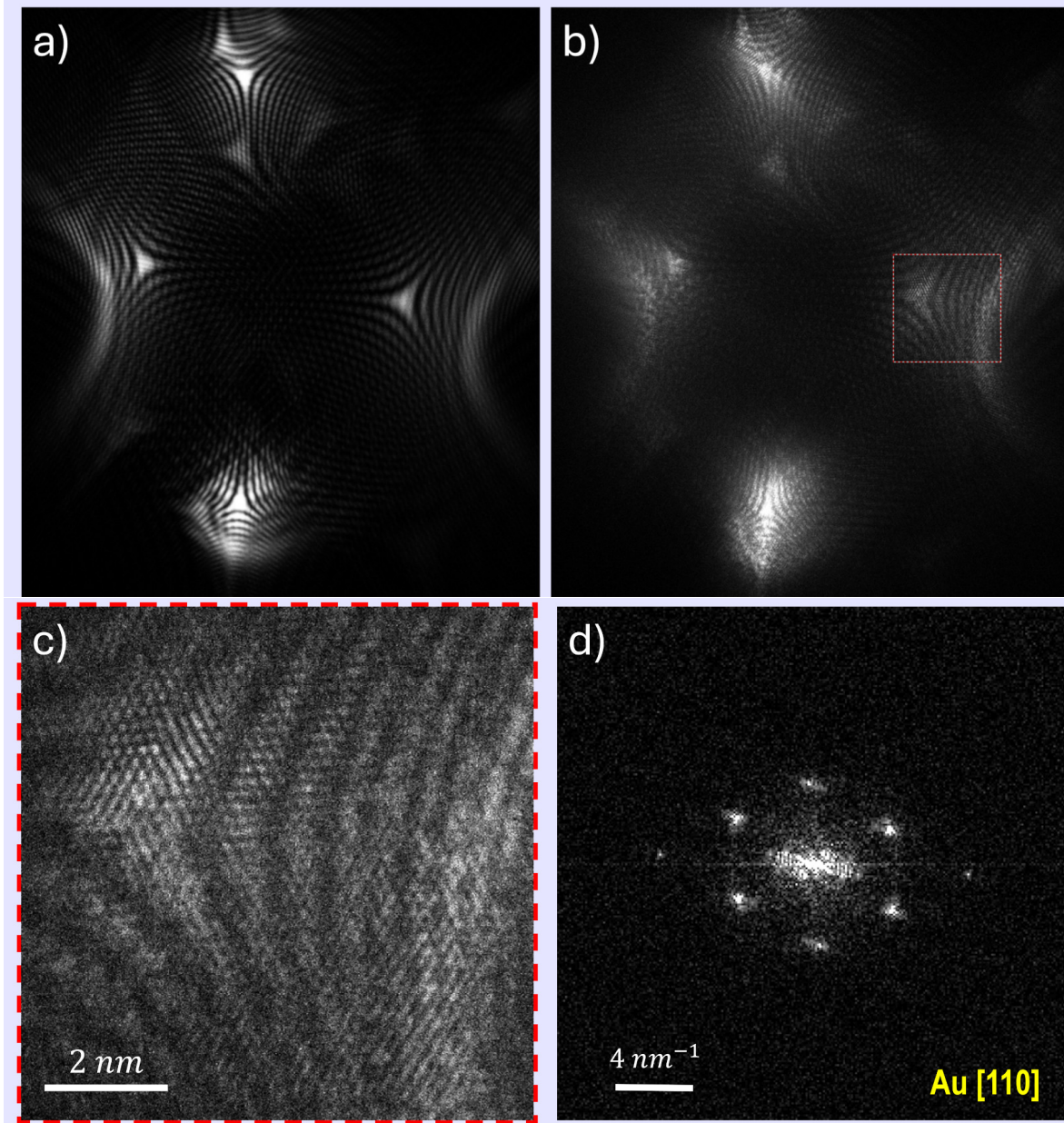


Figure 9.6: (a) Extended probe used for the calibrations, on vacuum and (b) on the carbon grating. The region in red was extracted (c) and Fourier-transformed to characterize the lattice fringes visible inside. The identification of the lattice fringes orientation (d) allows calibrating the pixel size of the image.

With this method, the rest of the characterization process involved the same steps used for the DS1 phase plate: identification of the range of applied biases that would grant a good balance between making visually different probes and having them explore a region of the desired size. In the end, the low-convergence probes were acquired at a probe magnification $M = 270k$ and the high-convergence probes were instead measured at $M = 430k$, with pixel sizes $d_{low} = 0.012nm$ and $d_{high} = 0.0083nm$, with

a pixel resolution of 3838x3710 pixels (the native format of the Titan-HOLO's Gatan K2 camera). Some examples are shown in Figure 9.7.

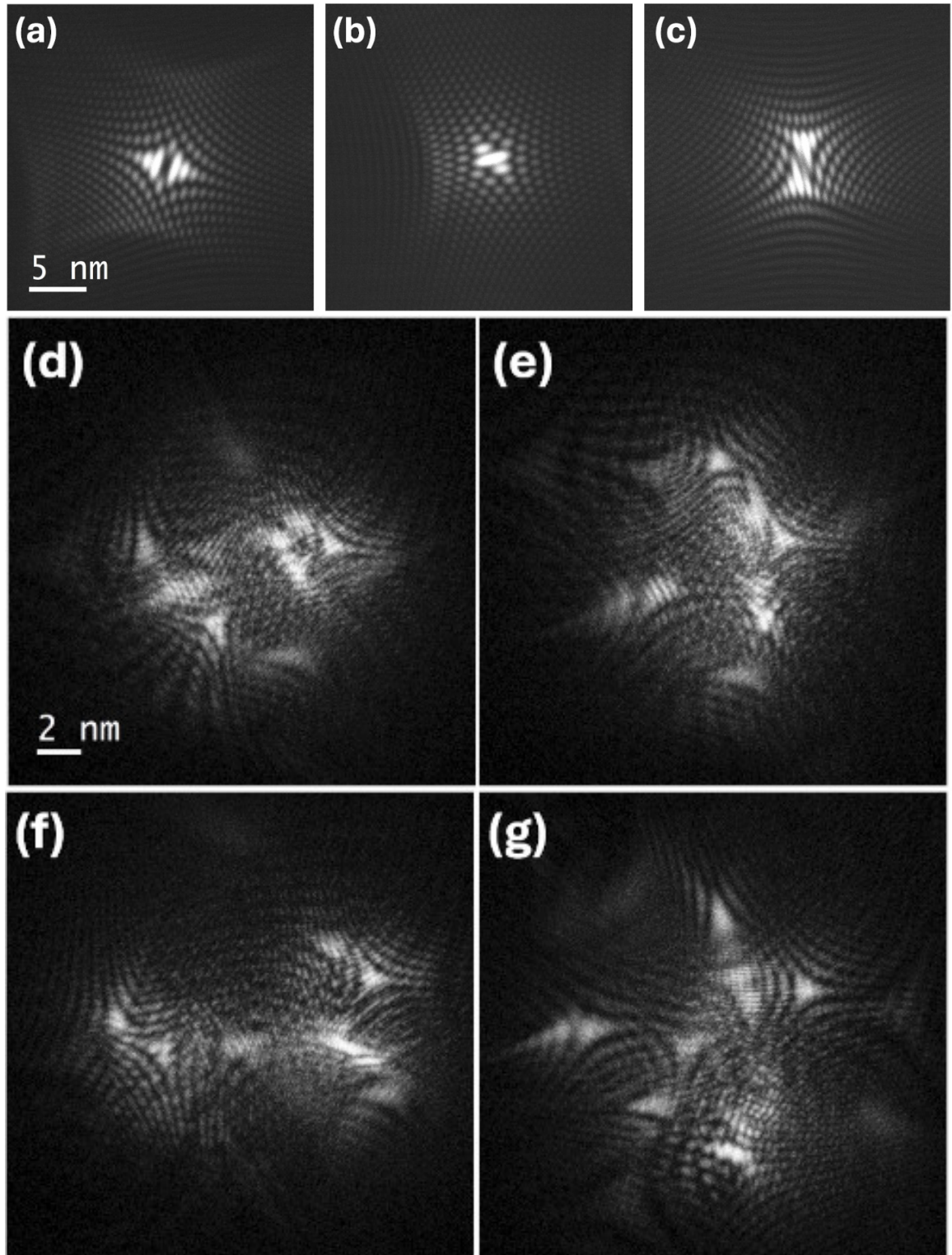


Figure 9.7: (a-c) Examples of low-convergence probes and (d-g) of high-convergence probes, extracted from the experimental data sets. Notice how the high-convergence probes have much finer details, as predicted by the simulations. The images are cropped around the probes to improve the visibility of the details, but they fit within a slightly larger frame.

During the characterization process, we also repeated the same test done for the DS1, to compare the simulated and experimental images. It was made using the low-convergence probes, because their lower number of features allowed to sum the probe images and still distinguish some features. Unfortunately, the test revealed immediately that the device had some kind of intrinsic asymmetry, because the experimental probes were distributed in a far more spread out configuration than the simulated ones, as seen in Figure 9.8.

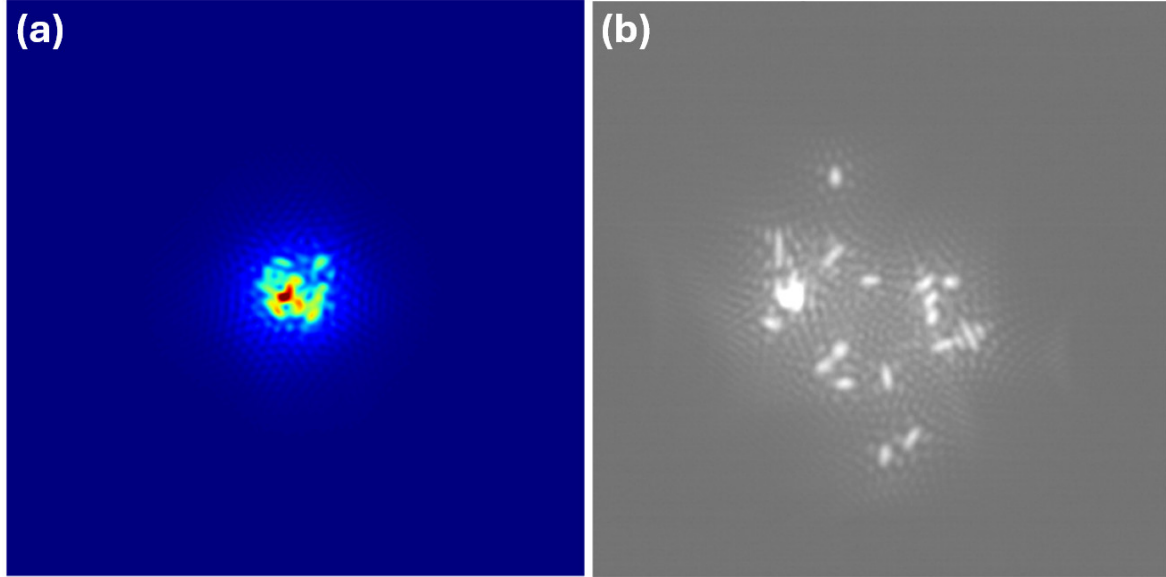


Figure 9.8: Comparison of the amplitude of: (a) the sum of 15 simulated low-convergence QC probes and (b) its experimental counterpart.

The difference between the simulation and the experiment was due to a combination of different factors. Just like for the DS1, small deviations of the probes from the simulations were expected, but in this case it became apparent how one of the eight pairs of electrodes was not working properly, either because of an undetermined fabrication issue or a problem with the biasing aperture holder. The issue was later solved by switching to a different device, confirming that one of the electrodes was defective, but the new device was instead subject to non-uniform charging, that slightly changed the shape of the structured probes each time the device was retracted and reinserted under the beam. The charging made it equally complicated to match simulations and experimental parameters. In the end, thanks to the presence of the high resolution camera, it was decided to take a slightly different approach and just record the structured probes on the camera instead of simulating them. The essence of the experiment remains the same, as long as the device is not moved between the probe image acquisition step and the HAADF signal acquisition step.

Figure 9.9 shows some examples of ways in which the charging would alter the convergent probe, both in the case of low-convergence and high-convergence. Panel (a) shows the standard STEM convergent probe. In the low-convergence configuration the device is not illuminated by the beam, but in practice it would often get exposed to the beam during insertion, because it was mounted on a manual aperture and its position had to be regulated manually. Thus, the probe would often be altered by the residual charging left on the electrodes. Panel (b) shows an example of such probe, obtained

by inserting the device and having all the electrodes biased at 0V. The effect can be partially compensated by biasing all the electrodes at a fixed offset negative value, but the charging is often non-uniform, and it can evolve with time. Panel (c) shows the probe observed in the same biasing conditions as in panel (b), measured four hours later. If the device is inserted with minimal charging, the probe often ends up looking like in panel (d), with a triangular shape. This is true for high-convergence measurements as well. In case of charging at high convergence the probe is like the one in panel (e), where the central spot is similar to panel (d), with four added caustics produced by the charge on the SPPs. Compensating for the charging in this case makes all the features converge at the center, but the probe is significantly broadened. Essentially, this means that in case of phase plate charging, when it is inserted it is not possible to measure a normal STEM image in focus, and this can cause issues in the choice of the area to image. A better explanation is found in Section 13.

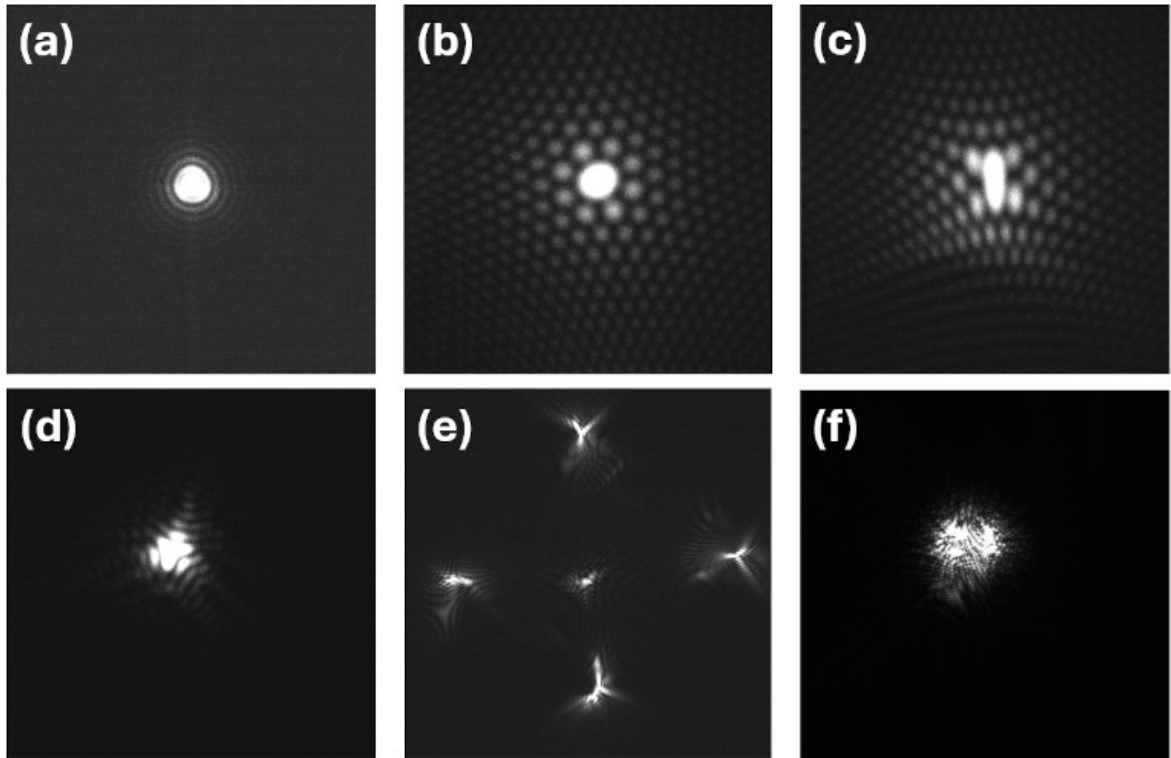


Figure 9.9: (a) Normal convergent STEM probe. (b) Effect of charging on the grounded device at low-convergence and (c) its evolution after four hours. (d) Probe seen with the phase plate inserted and minimal charging. (e) Effect of charging on the grounded device at high-convergence, with the four extra caustics produced by the SPPs, and (f) probe after compensation of the charging with an added uniform negative bias offset.

10 Electronics and Software for Measurements

This section describes briefly some of the hardware components mentioned previously in the two experimental setups for the measurements. Also, there is a subsection devoted to explaining the software that I wrote in Python to automatically control the structured probe generation process, the acquisition of the images of the probes, and the HAADF signal.

10.1 Electronics

For the electronics of the experimental setup, there were four components that were controlled by Python codes. They are:

- The power source connected to the Thermo Fisher Scientific biasing aperture holder, already shown in Subsection 1.6
- The microscope's scanning system for STEM
- The readout mechanism from the HAADF detector
- The microscope's high-resolution camera

In addition to these, there is the read-out mechanism of the microscope's Fischione HAADF detector, so the detector shall be briefly discussed as well. We start with the power sources. The two experimental setups for the DS1 and the QC phase plate used different power sources. The DS1 setup used a Texas Instruments digital-to-analog converter, model DAC81416 [101]. It is a 16-channel, buffered, high-voltage output Digital-to-Analog Converter (DAC) in 16-bit resolution with 2.5-V internal reference. The DAC provides bipolar output voltage up to ± 20 V and unipolar up to 40 V. It can be connected to a PC or laptop through USB. It does not come with a case, so it is quite susceptible to electrostatic discharge damage, and if programmed incorrectly it is easy to damage it by accidentally generating a large current.

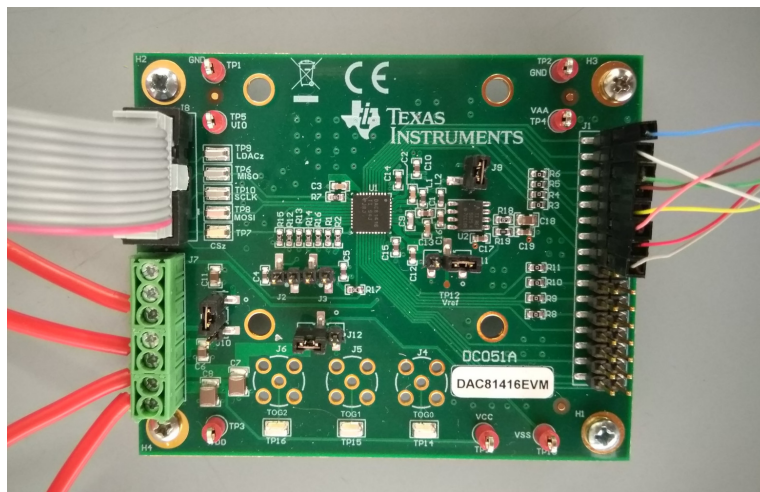


Figure 10.1: The Texas Instruments DAC. On the left side, the USB2ANY connector and the red cables to the power sources. On the right, the pins connected to the 8 cables going to the biasing holder: six for the biases and two for grounding.

The DAC was used in the DS1 setup. I programmed the control interface for the experiment in LabVIEW, starting from a preexisting interface made to control other MEMS devices. The DAC was replaced with a different power source for the experiments on the QC phase plate: a Stahl-Electronics 16-channel high-stability DC Source, seen in Figure 10.2. The Stahl power source was chosen because of its reliability and its operating range, as well as its easy control through the Python PyVISA package [102].



Figure 10.2: The Stahl-Electronics power source used in the experiments.

In terms of reliability, the Stahl power source provides bipolar outputs up to ± 40 V, with live feedback on the applied bias and current with uncertainty of a few mV and μ A respectively. The maximum output current is limited to 10 mA, which is a safe maximum current value to apply to the QC phase plate without risking damaging it inside the microscope. The power source has a response time of a few milliseconds, and it must be accounted for during the experiment, as it poses a time limit on the speed at which it is possible to switch between different probes. Only eight out of the sixteen channels were used, as the QC required only eight separate biases, and the holder only has eight biasing channels anyway.

The scanning system was the microscope’s native scanning coils. The scanning coils were only needed for the hybrid Scanning-ECGI method. The microscope used for that experiment was equipped with a Gatan DigiScan platform, that allows controlling the microscope’s scanning system via Digital Micrograph or Python scripting. In that way, the parameters of the scan could be customized, remotely selecting the pixel size, dwell time, scan rotation and raster X and Y sizes.

About the Annular Dark Field detector, our experimental setups always used a Fischione 3000 HAADF detector, on both microscopes used for microscale and nanoscale imaging. The Fischione detector is a relatively standard high-resolution STEM-HAADF detector. It is made with a single crystal yttrium aluminum perovskite (YAP) scintillator that physically detects the high-angle scattered electrons. The scintillator connects to a photomultiplier tube that amplifies the emitted signal, which is then processed by the microscope’s control system to display a digital STEM image. Figure 10.3 shows a schematic of the detector. The detector has two output ports: a digital output connection to the microscope, that can then directly read the signal and is controlled only by the microscope computer, and an analog BNC port. The BNC port is usually not connected to anything else, and is used for testing. In the earliest experiment, we read out the HAADF signal using a Keithley 2400 multimeter, connected to the BNC exit of the detector. The multimeter can be controlled through PyVISA, just like the Stahl power

source. Alternatively, the Gatan DigiScan system can read the signal from the BNC port and display the STEM HAADF images in Digital Micrograph. So, it is possible to write a Python script to automate the acquisition process and read out the HAADF signal as it is coming from the detector by saving small HAADF images. This method was implemented for the hybrid scanning-ECGI setup, that made use of the scanning system, and it has two main advantages: first, the signal is read out and acquired by the microscope's own electronics and software. So there is a lower risk of picking up electrical noise than when connecting an additional instrument, like a multimeter, to the BNC port on the detector. Second, if the readout process is controlled digitally, it can be easily synchronized with all the other experimental steps, and in particular with the scanning.

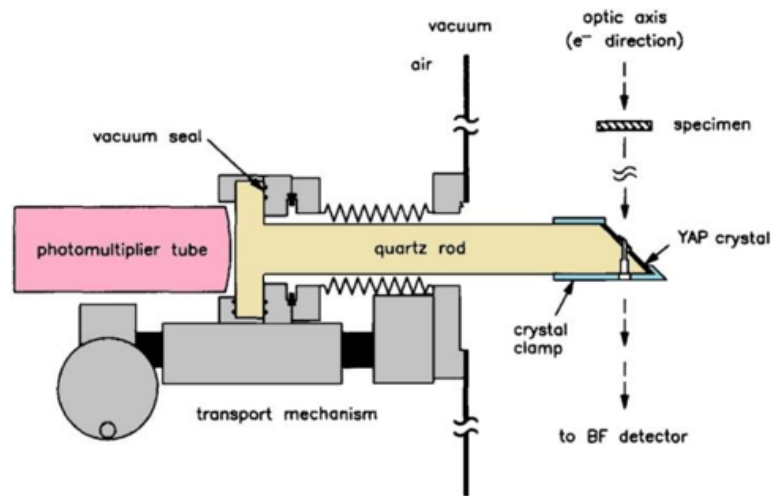


Figure 10.3: Schematic of the Fischione HAADF detector. The YAP crystal annulus is detecting the electrons scattered at high angle. The inner and outer angles of collection are set by the imaging system. The detecting annulus can be inserted and retracted from the column without breaking the vacuum, and when inserted it does not completely block the beam, so it is still possible to observe the low angle diffraction on the microscope's screen or TV camera. Adapted from [103].

Finally, the camera. The tests done to match the simulated structured probes to their experimentally measured counterparts necessitated actual images of the probes. Modern TEMs are equipped with CCD cameras to capture digital images of the samples. In the case of the DS1 phase plate, the probe images were acquired on a Thermo Fisher Scientific Talos TEM, equipped with a CETA CMOS camera, with a pixel resolution up to 4k x 4k. The QC modulator was used for experiments on the Thermo Fisher TITAN-HOLO TEM in Forschungszentrum Jülich, that is equipped with a Gatan GIF Quantum K2 camera. This second camera has a slightly lower pixel resolution of 3838x3710, but it has almost three times the DQE at the operating voltage of 300kV. It is more relevant to the project, as it was used to acquire the images of the probes for high-resolution imaging, that have very fine details. Also, the K2 camera was already interfaced with the Gatan Digital Micrograph software, that could run python scripts to set up an automated image acquisition process. Both those aspects proved vital in our case, as it was not possible to simulate the probes reliably and instead they had to be measured directly. So, the programmability significantly reduced the difficulty of the task.

10.2 Python

This Subsection is devoted to the explanation of the Python codes used for the measurements. The DS1 experimental setup was entirely controlled through a LabVIEW interface that I specifically modified for the task, but I will not spend time discussing it because the experimental steps are almost the same used for the more recent control codes written in Python. So, the following paragraphs provide pseudocode snippets that illustrate the function of the code used for the measurements with the QC setup in its latest stages, when the hybrid scanning-ECGI was fully implemented.

10.2.1 HTTP Server

The automated control of the microscope through python was managed by setting up an http server over the local network of the microscope. The code for the server was run in the Digital Micrograph of the Gatan support computer, that was already interfaced with the microscope. The server was managed through a Python class called Handler, that handled the requests to the microscope. The Handler class contains two main functions, POST and GET. POST commands control the acquisition step, while GET commands are for closing the server and saving .dm3 images (native format of Digital Micrograph). The most relevant python packages needed to set up the server are listed at the beginning. The *DigitalMicrograph* package contains all the functions that control the camera and the Digiscan system. The *http* packages are necessary to set up the server. The *numpy* package is needed for array manipulations.

```
1 import DigitalMicrograph as dm
2 import http
3 import http.server
4 import numpy as np
5
6 class Handler(http.server.BaseHTTPRequestHandler):
7
8     def do_POST(self):
9         try:
10             if function == 'camera':
11                 if command == 'acquire':
12                     return self.do_camera_acquisition_post(id, request)
13             elif function == 'digiscan':
14                 if command == 'acquire':
15                     return self.do_digiscan_acquisition_post(request)
16
17     def do_GET(self):
18         try:
19             if function == 'exit':
20                 self.do_exit()
21             elif function == 'camera':
22                 if command == 'acquire':
23                     return self.get_camera_acquisition(id)
24
25     def do_exit(self):
26         #close the camera
27         #delete temporary images
28         #exit from server loop
```

Listing 3: Pseudocode of the first part of the Handler class. Python package import and definition of the request tree and the server exit function.

Notice that the Digiscan has only a POST function, while the camera has both a POST and a GET function. That is because the STEM images made through the Digiscan system to acquire the HAADF signal on a raster scan of a structured probe were stored on the client side, on my laptop's SSD drive. The large images acquired by

the camera were instead stored on the server side, and the GET function executes that specific command. The Digiscan POST function calls another function that practically executes the acquisition using Digital Micrograph's internal functions. The pseudocode for both is listed below.

```

2  def do_digiscan_acquisition_post(self, request):
    # sets up the digiscan to do the acquisition and returns the image
    # to the client side as a numpy array
4  save_path = request.get('save-path')
    img = self._do_digiscan_acquisition(request['scan_params'])
6  np.save(save_path, img)

8  def _do_digiscan_acquisition(self, scan_params):
    # executes the image acquisition and returns the image as numpy array
10 param_id = DM.DS.CreateParameters(scan_params['width'],
    scan_params['height'], scan_params['rotation'], scan_params['dwelltime'])
12 try:
    array = np.zeros(shape=(scan_params['height'], scan_params['width']))
14 img = DM.CreateImage(array)
    img_id = img.GetID()
16 DM.DS.SetParametersSignal(param_id, img_id)
    DM.DS.StartAcquisition(param_id, continuous, synchronous)
18 finally:
    DM.DS.DeleteParameters(param_id)
20 return img.GetNumArray(), img

```

Listing 4: Pseudocode of the Digiscan acquisition functions of the Handler class.

The last part of the Handler class contains the camera POST and GET functions. The first sets up the camera with the parameters specified in the server request. The second function is the one actually running the acquisition, and the third is the function that handles the image saving.

```

2  def do_camera_acquisition_post(self, id, request):
    # starts the camera with the exposure and binning parameters, or stops it
    st = state[id]
4  if request['command'] == 'start':
    camera = dm.GetCameraByID(id)
    st['settings'] = {'exposure': float(request.get('exposure')),
6                      'xBin': int(request.get('xBin', 1)), 'yBin': int(request.get('yBin',
    1))}
    st['image'] = camera.CreateImageForAcquire(xBin, yBin)
    st['mode'] = 'fast'
10 else request['command'] == 'stop':
    self._try_camera_stop(id)

12 def _do_acquisition(self, id):
    # executes the image acquisition
    st = state[id]
14 camera = dm.GetCameraByID(id)
    image = st['image']
16 settings = st['settings']
    camera.GetFrameInContinuousMode(image)
18 return image.GetNumArray()

20 def get_camera_acquisition(self, id):
    # recovers the acquired image and saves it on disk as .dm3 file
22 save_path = query['save-path']
    st = state[id]
24 arr = self._do_acquisition(id)
    st['image'].SaveAsGatan(save_path)

26 httpd = http.server.HTTPServer(("0.0.0.0", 8888), Handler)
30 try:
    print("starting server")
32 httpd.serve_forever()
httpd.server_close()

```

Listing 5: Pseudocode of the camera acquisition function of the Handler class, and the few lines of code that activate the html server, keep it running, and close it.

The last five lines are not part of the Handler class, they start the http server and keep it running or close it depending on the commands received from the client side.

10.2.2 ECGI measurement

The code for the automated measurement was run on my laptop, in Jupyter Notebook, and it can be split in three main sections. The first one used two Python classes to control the power source, and consequently the QC phase plate. The first class is called Stahl, and contains a few functions to connect to the power source through the PyVISA package and control it. Listing 6 lists only the function used to set the voltage of one of the output channels to a certain bias value. This function is used extensively in the second class, that is used to manage the QC phase plate. This class also has more functions, defined for special circumstances, like a function that displays a scheme of the pairing between channels and electrodes of the phase plate. Like for the first class, the listing only shows the three most relevant functions. The first sets the device to a uniform $V_{bias} = 0V$ on all electrodes. The second applies an arbitrary configuration of biases to the device, and the third takes in input the desired values of topological charge (in ℓ units), and the tilts and astigmatism effects (in V units) for the horizontal and vertical directions of the phase plate, and calculates the V_{bias} configuration that applies those effects. Once calculated, the biases are applied and returned as an output of the function, so that they can be stored on disk. In lines 10 and 34, the power source and phase plate controllers are declared, and in line 35 the phase plate is set to a uniform $V_{bias} = 0V$.

```

1 import numpy as np
import pyvisa

3
class Stahl:
5     #uses the pyvisa.ResourceManager() fuction to USB connect to the power source
    def set_voltage(V, ch)
7         # sets the voltage of channel "ch" to "V"

9 Pwr_src = Stahl()

11 class quadrichopsticks:
    #controls the phase plate through the power source control class
13     def allzero():
        "Puts all the channels to zero"
15         for ch in range(8):
            Pwr_src.set_voltage(0, ch)

17
    def arb_conf(biases):
19         for ch in range(8):
            Pwr_src.set_voltage(biases[ch], ch)

21
    def Lta_set_and_save(OAMs = np.ones(4), Tilt = np.zeros(2), Ast = np.zeros(2),
vs_charge=0):
23         "Sets device configuration by converting the OAM values, tilts and
astigmatisms to bias voltages"
        Vs = np.multiply(OAMs, OAM_to_V_factors)
25         Ofst = np.array((-Tilt[1]+Ast[1], -Tilt[0]+Ast[0], Tilt[0]+Ast[0], Tilt[1]+Ast
[1]))
        combine = np.array([Vs[0]+Ofst[0], -Vs[0]+Ofst[0], Vs[1]+Ofst[1], -Vs[1]+Ofst[1],
27         -Vs[2]+Ofst[2], Vs[2]+Ofst[2], Vs[3]+Ofst[3], -Vs[3]+Ofst[3]])+ vs_charge
        for ch in range(8):
29             Pwr_src.set_voltage(combine[ch], ch)
        return combine

31
QC = quadrichopsticks() #Make a quadrichopsticks instance
33 QC.allzero()          #Set all biases to 0V

```

Listing 6: Pseudocode of the Python classes that control the Stahl power source and the QC phase plate.

At the end of Subsection 3.2 it was mentioned how the biasing of the electrodes can also be interpreted as applying a combination of topological charge, tilt and two-fold astigmatism to the phase plate. To understand each effect better, Figure 10.4 shows a small set of examples that illustrate the principle driving the generation of the structured probes according to the code in lines 24 through 29 of Listing 6.

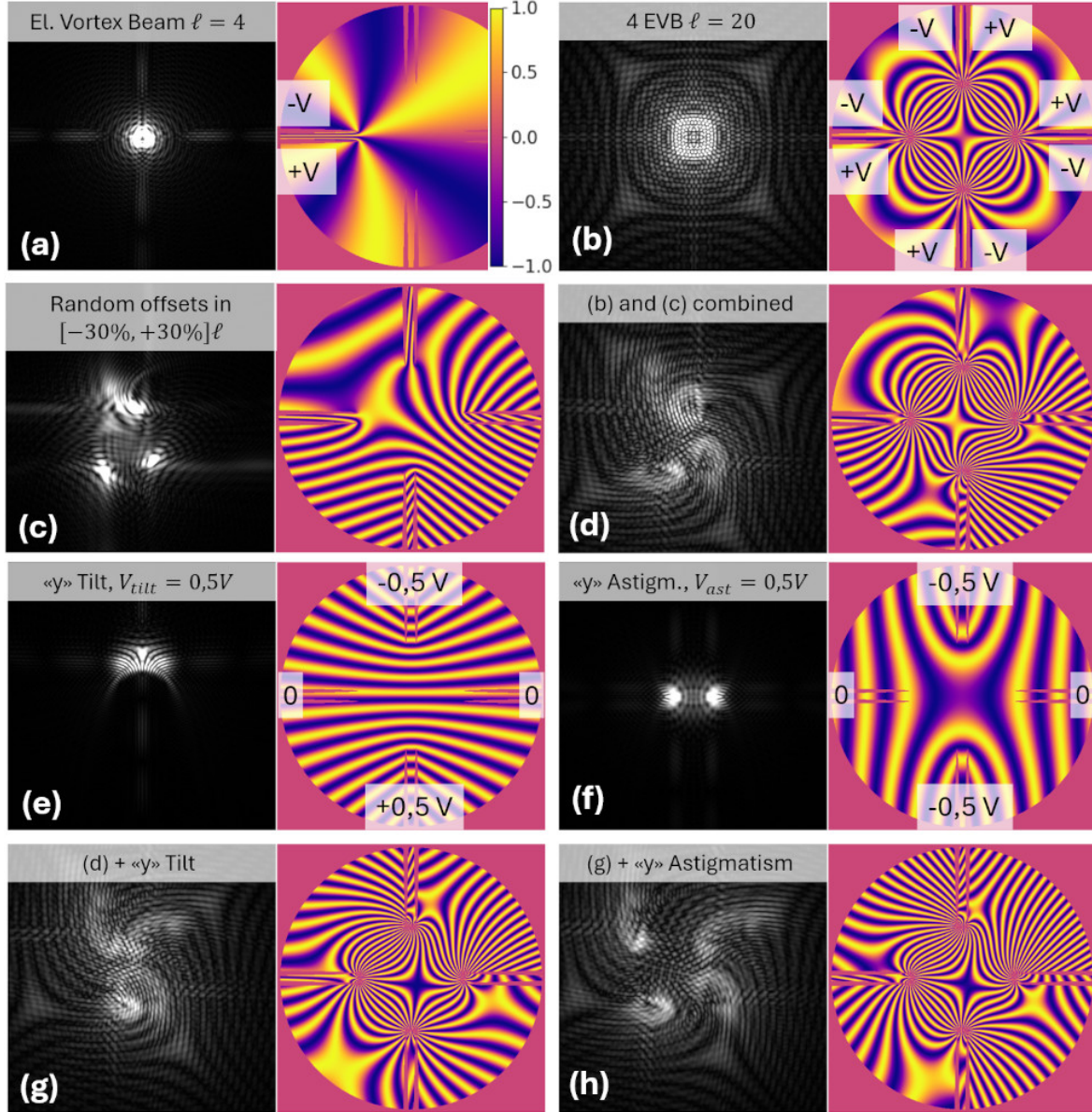


Figure 10.4: Simulated example of generation of a structured probe. All the panels show both the probe's amplitude and sine of the phase (in color). (a) Vortex beam with $\ell = 4$. (b) Combination of four vortices with $\ell = 20$ and alternating sign around the phase plate, the base configuration. (c) Random perturbation of maximum 30% of ℓ . (d) Combination of the base configuration and perturbation. (e) Effect of QC tilt on a simple convergent probe. (f) Effect of QC astigmatism on a convergent probe. (g) alteration of the perturbed base configuration through the added tilt and, (h) tilt + astigmatism.

The second section of the code contains the client-side functions that post commands to the http server controlling the camera and the Digiscan system. These functions required the *request* Python package to send requests to the server. The server reads

json (JavaScript Object Notation) files containing the instructions, so we also needed the *json* package to compile those files. The four main functions are listed below, and they carry out the main tasks related to automated acquisition from the camera and the Digiscan system. The first function sets up the camera to begin the acquisition, specifying the exposure time (in *s*) and the binning of the sensor (if required). The second stops the camera. The third launches the acquisition of a single image with the previously set exposure, that is then saved in the specified path. The fourth function controls the acquisition of a STEM image from the Digiscan system. Since the command for each Digiscan acquisition is compiled right before execution, there is no need for start and stop functions for that system.

```

1 import requests
2 import json
3 server_url = "http://192.168.0.11:8888/"
4
5 def start_camera(exposure, binning, xbin, ybin):
6     #starts the camera and sets the exposure time (in seconds) and the binning of the
7     #sensor
8     if binning:
9         resp = requests.post(server_url+"camera/0/acquire/",\
10                             json={'command': 'start', 'exposure': exposure, 'xBin': xbin, 'yBin': ybin})
11     else:
12         resp = requests.post(server_url+"camera/0/acquire/",\
13                             json={'command': 'start', 'exposure': exposure})
14
15 def stop_camera():
16     #stops the camera
17     resp = requests.post(server_url+"camera/0/acquire/", json={'command': 'stop'})
18
19 def image_acquisition(save_path):
20     #acquires an image with the camera, using the parameters set while starting it
21     data = load_as_array(requests.get(server_url+"camera/0/acquire/"))
22     np.save(savepath, data)
23
24 def stem_acquisition(rotation, width, height, dwelltime, save_path):
25     #acquires a STEM image of the desired size at the magnification set on the TEM
26     args = {'scan_parameters':
27             {'rotation': rotation, 'width': width, 'height': height, 'dwelltime': dwelltime, }}
28     data = load_as_array(requests.post(server_url+"digiscan/acquire/", json = args))
29     np.save(savepath, data)

```

Listing 7: Pseudocode of the client-side functions used to interact with the http server of the microscope.

The last part of the code is the one used to run the actual experiment in the hybrid scanning-ECGI configuration. The pseudocode is found in Listing 8. The first lines specify the number of probes to use and the parameters of the raster scanning and the generation of the structured probes as demonstrated previously. Then, the acquisition begins. The first cycle iterates on the number of probes, generating them and storing both the OAM, tilt and astigmatism values, and the configurations of nominally applied biases (lines 17 through 21). Then, for each probe, the bias configuration is applied to the phase plate (line 21) and the software waits for a stabilization time for the probe of the order of $100 \mu s$. The change of the biases takes only a few milliseconds, but at large convergence angles, when large portions of the tips are illuminated, there might be some residual charging that introduces a slower transient between different stable configurations of the device.

The software proceeds to measure the HAADF signal, and then saves the STEM image (small, 32x32 or 64x64 pixels) on the local drive. When the HAADF signal acquisition is done, the program saves the configuration lists on disk and the first step is done.

The sample can be retracted from under the beam, and the software can then proceed to start the camera and measure the images of the probes in vacuum (lines 33 through 41). After the process, the phase plate is biased uniformly to zero and the camera is stopped.

```

#Number of images to acquire (= No. probes) and dwell time (us)
2 N_probes = # of probes to use
dw_t = #in us >= 1
4 side = # in px, >= 16

6 #make arrays to store applied OAMs, tilts and astigmatisms and the corresponding
  biases
OAM_tilt_astig = nominal_biases = np.zeros((N_acq,8))

8
symm = np.array([1,-1,1,-1]) #base array of the applied OAMs (alternating)
10 nom_OAM = #maximum nominal applied OAM on each SPP
OAM_var = #maximum perturbation of the OAM on each SPP
12 tilt_mod = #maximum value of applied tilt on each axis
astig_mod = #maximum value of applied astigmatism on each axis
14
for i in range(N_probes):
16
    oams = nom_OAM * symm * np.random.choice((1,-1))+np.random.uniform(-1,1,4)*OAM_var
18    tilt = tilt_mod*(np.random.rand(2)-0.5)
    astig = astig_mod*(np.random.rand(2)-0.5)
20    OAM_tilt_astig[i,:] = np.concatenate((oams,np.concatenate((tilt, astig))))
    nominal_biases[i,:] = QC.Lta_OAM_set_and_save(OAMs=oams, Tilt=tilt, Astig=astig,
    vs_charge=q)
22
    time.sleep(t_probe_stab) #wait for probe stabilization (s)
24    #acquire stem image using the Digiscan
    stem_img = stem_acquisition(rotation=a, width = side, height = side, dwelltime =
    dw_t)
26    np.save(savepath, stem_img) #save stem image

28 np.save(savepath, OAM_tilt_astig)
np.save(savepath, nominal_biases)
30
#remove the sample from under the beam, without changing anything else
32 #load the nominal applied biases of the previously used structured probes
nominal_biases = np.load(savepath/nominal_biases)
34
start_camera(exposure=exposure_time, binning=False)
36
for i in range(N_probes):
38
    QC.arb_conf(nominal_biases[i]) #apply biases
40    time.sleep(t_probe_stab) #wait for probe stabilization
    image_acquisition(save_path)
42
QC.allzero()
44 stop_camera()

```

Listing 8: Pseudocode of the measurement process, done in two steps: first, the acquisition of the HAADF signal in the form of STEM images from the Digiscan system, then the acquisition of the images of the probes. The entire measurement process can also be carried out "in reverse" measuring the probes first and then the HAADF signal.

11 Samples

Experimental runs were made on four different samples. The first two were made for microscale imaging with the Double Sorter 1 setup. The second pair was selected to attempt nanoscale- and atomic resolution imaging with the Quadrichopsticks.

11.1 Microscale binary masks

For the microscale imaging, the two samples were simple amplitude masks mimicking the Siemens star targets used in the numerical simulations of the experiments of Subsection 7.3. The masks were fabricated by Focused Ion Beam milling on a commercially available 200 nm thick silicon nitride Si_3N_4 membrane, covered with ~ 150 nm of Au. The Si_3N_4 membrane is mostly transparent to electrons, while the thick Au layer is basically opaque. Then, these kinds of sample are an excellent approximation of the binary Siemens star target. Figure 11.1 shows the two targets. The one in (a) is a five-spoke Siemens star about 40 μm in diameter, the one in (b) is about one order of magnitude smaller, with a diameter of ~ 3 μm and ten spokes instead of five (to increase the spatial frequency of the finest features).

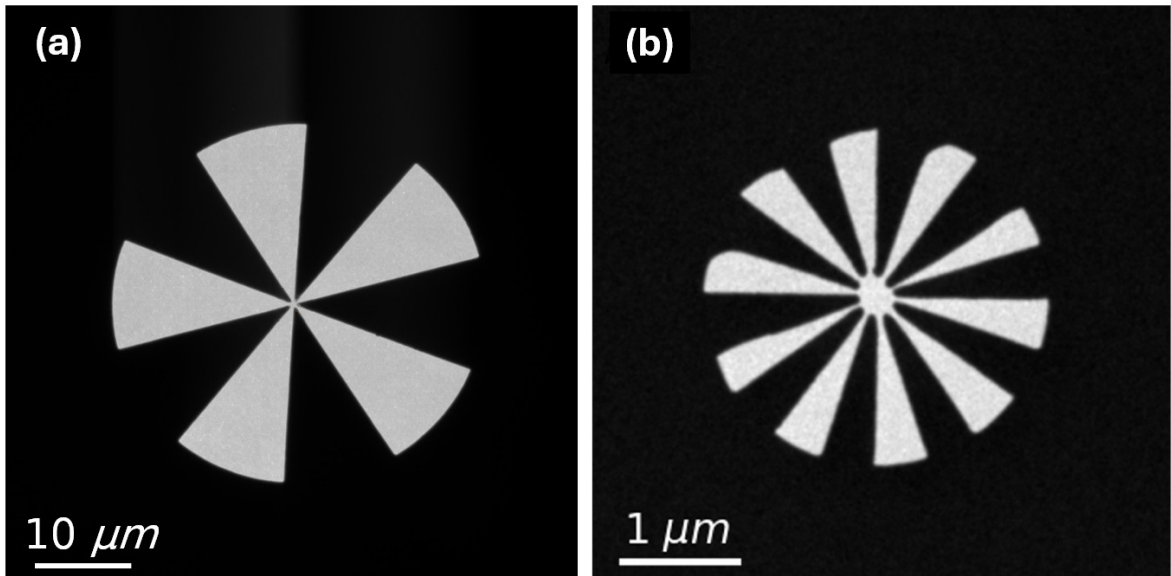


Figure 11.1: TEM images of the two microscale imaging samples. They were produced to match the simulation work that used a binary Siemens star target. Notice that the size of the two objects is quite different, because while designing them we did not know exactly the extent of the region explored by the probes as a function of the applied biases.

11.2 Au nanoparticles and twisted MoS_2

For the nanoscale- and atomic resolution imaging, we selected two samples. The first sample was a standard microscope grid with Au nanoparticles (NPs) on amorphous carbon, the second was a sample of twisted molybdenum disulfide (MoS_2). The choice of the first was natural, as it is the standard sample used in TEMs to prepare STEM alignments, so it was readily available. The contrast of Au NPs on the carbon in STEM

HAADF is quite large, so it is a good sample to use for preliminary measurements, because it is easy to determine whether the NPs appear in the reconstructions. It is also possible to locate Au NPs that have visible lattice fringes, to select a target that has features at atomic resolution. Imaging the lattice fringes on those nanoparticles would already prove that the technique could reach atomic resolution, so the standard sample is good enough for preliminary measurements. Figure 11.2 shows 300 kV STEM HAADF images of the Au NPs sample taken with a nominal illumination convergence angle of 7.23 mrad , including some lattice fringes and the corresponding diffractogram.

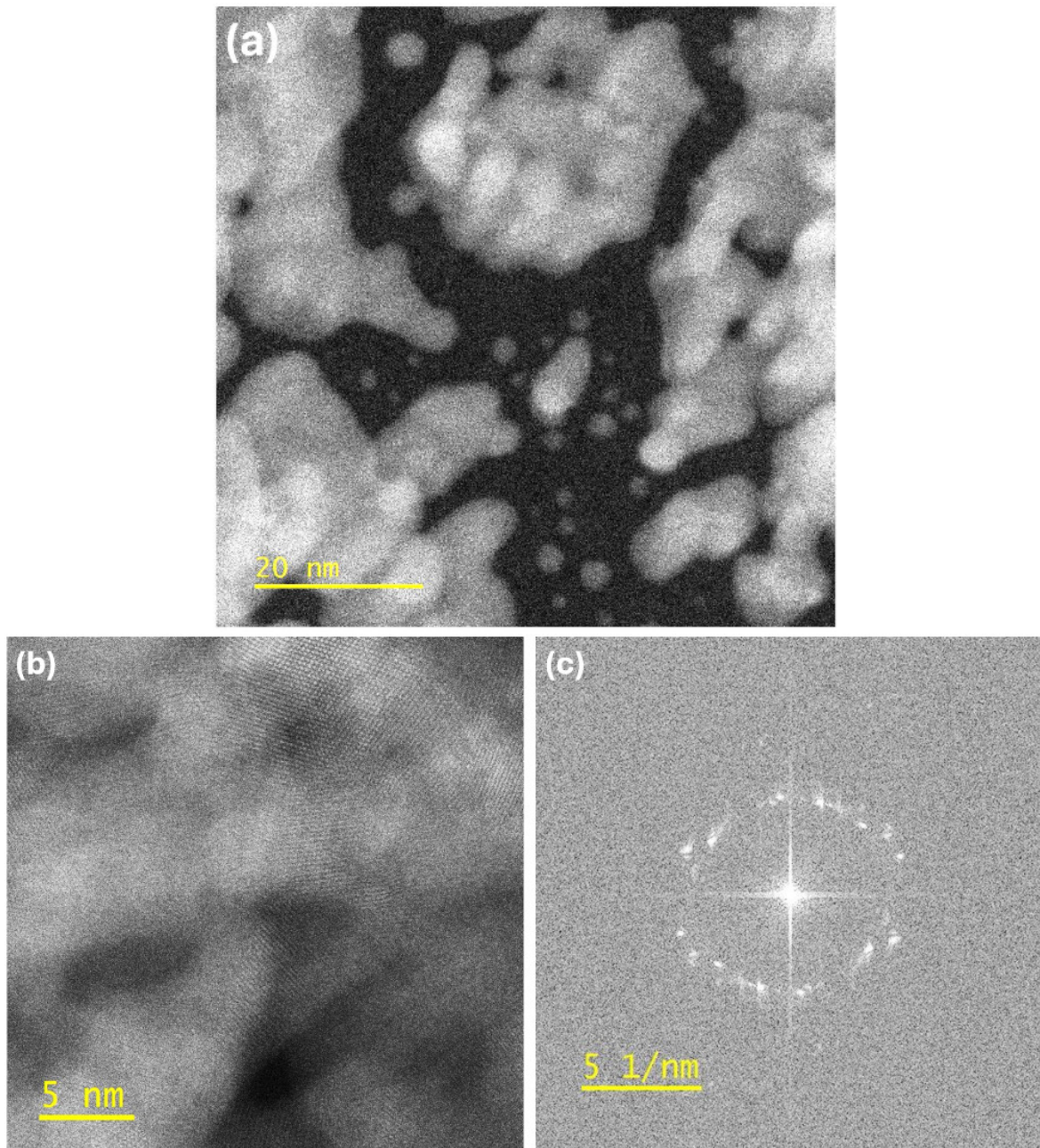


Figure 11.2: The standard TEM sample. (a) Example of a small region of the standard sample, with a group of small Au nanoparticles at the center. (b) Detail of some lattice fringes and (c) corresponding diffractogram.

The other sample we used was a twisted MoS_2 sample, made by a colleague with a tape exfoliation technique. The sample is created by stacking two flakes of molybde-

num disulfide on top of each other, at an angle of $\sim 10^\circ$. The sample had different regions where the two stacked flakes had different thickness. So, there is a region with multilayer-multilayer stacking, one with multilayer-monolayer stacking and another with monolayer-monolayer stacking. Figure 11.3a shows an optical microscope image of the MoS_2 flakes stacked on a hexagonal grating of holes on a Si membrane. The sample is free-standing on the holes. I acquired a few STEM-HAADF images from it on a microscope that was dedicated to only STEM and was equipped with a probe corrector. One of the images is shown in Figure 11.3b, taken in the monolayer-monolayer region. The image and its FFT show that the sample was not very clean, and it ended up contaminating too much when inserted in the microscope where the Ghost Imaging setup was prepared. Thus, in the end it was not possible to use if for the actual experiment, because it would build up a thick layer of contaminants in seconds, accentuating charging effects as well.

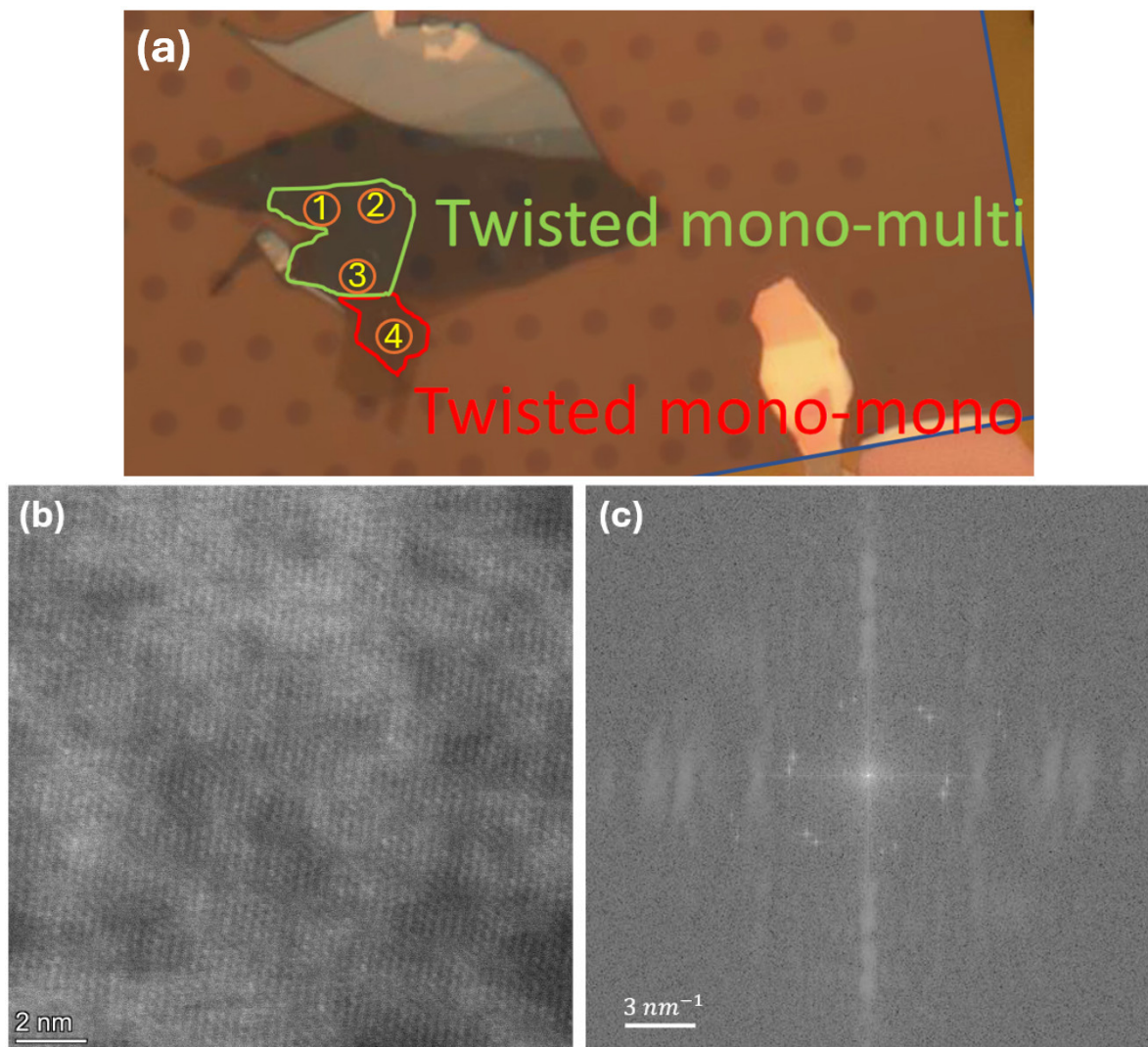


Figure 11.3: The twisted MoS_2 sample. (a) Optical microscope image of the stacked MoS_2 flakes. (b) HRSTEM image of the monolayer-monolayer region and (c) corresponding FFT showing the two twisted diffraction patterns.

Part III

Experimental Results

This last Part presents the experimental results obtained with the two experimental setups, making comparisons with the simulation work. Moreover, the final Sections discuss the current limitations to the execution of an experiment, and give some perspective on some possible future developments of ECGI.

12 Double Sorter I Setup

The setup that used the DS1 phase plate, described in Section 8, had the phase plate positioned in the sample holder, because it was not possible to introduce it in the condenser system while maintaining full control over the illumination condition. Thus, the samples had to be positioned in a different plane along the optical axis, so in the end they were put in the Selected Area Diffraction plane. As it was already mentioned before, the samples for microscale imaging were simple binary amplitude masks realized by patterning a silicon nitride membrane with a thick layer of gold. For this setup there was no interest in the effects of scattering from single atoms, as the illuminating probes were at micron size, so the measurement of the signal corresponded with the measurement of the total transmitted intensity. The only detector available for such measurement on the microscope was the HAADF detector, that had to be used in an off-axis configuration. The detector only has two available positions, inserted and retracted from the column. Then, the transmitted beam had to be deflected onto the HAADF detector using the lenses and beam shifts of the projection system.

12.1 Acquisition of the Probes

Regarding the acquisition of the probes, it was difficult to observe the probes generated by the DS1 because of the unconventional microscope alignment necessary to complete the measurement. The correspondence between simulated and experimental probes was already confirmed during the device characterization stage, and even though it was not perfect, as showed in Section 9, it was deemed acceptable for a proof-of-concept measurement. So, the probe images used for the numerical reconstructions were simulated using the lists of biasing configurations that generated them, and that were recorded at the time of measurement. So, all the probe images were numerically simulated as images with a resolution of 512x512 pixels.

12.2 Signal Acquisition

The measurement of the signal was carried out as specified in Section 10, using a Keithley 2400 multimeter to read out the voltage output from the HAADF detector's BNC exit. The measurements were performed without scanning the probes, so there was no need to synchronize the acquisition with the scanning. However, it was necessary to synchronize the change of illumination pattern with the signal acquisition. The control code for the experimental setup was written in LabVIEW, and it automatically took care of the synchronization problem. However, it was necessary to run some prelim-

inary tests to determine the optimal acquisition time and delay between measurements.

The multimeter measured a single reading by averaging over a number of acquisitions of equal duration. So, the measurement time for a single point could be regulated both in terms of single acquisition duration and number of averages, with the total time per measurement point $t_{tot} = t_{single} \cdot N_{avg}$. We tested twelve combinations of single acquisition time ($10ms$, $2ms$ and $200\mu s$) and averaging (over 1, 10, 20, or 30 single acquisitions) on a constant signal obtained by keeping a probe fixed on the detector, without any sample in between. Except for the run with no averaging and acquisition duration of $200\mu s$, all the conditions had a similar mean signal of $250 mV$ and a standard deviation of $2-4mV$. Thus, the single acquisition duration was selected to be $2ms$ and each acquisition was averaged over 10 single acquisitions, for a total of $\sim 20ms$ per measurement point. The signal also sometimes showed some erratic variations not related directly to beam instability or other obvious causes, probably because the system was picking up some external noise. We assume that the effect was not related to the beam because it persisted even upon disconnection of the multimeter from the detector's BNC port. An example is shown in panel (a) of Figure 12.1.

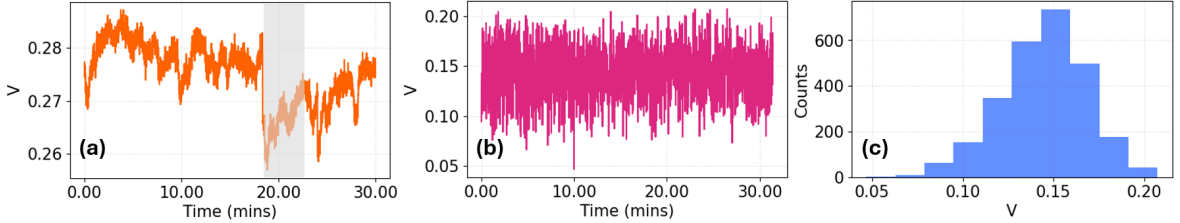


Figure 12.1: (a) An example of instability of the signal measured from the HAADF detector over a long time. The region in gray marks the interval over which the detector was disconnected from the detector. (b) Plot of the HAADF signal acquired for the first successful experimental image reconstruction and (c) corresponding histogram. Notice that the time scale of the measurement is very large, 30 minutes of acquisition for 2500 different probes.

Panels (b) and (c) instead show how the signal from a measurement on a binary target usually looks like, with most values at an average intensity and fewer at the extremes. This is because if the probe distribution covers a region roughly 50% transparent, then the probes should be on average blocked only partially. The total measurement time in panels (a) and (b) is quite long, because of an added $0.7s$ of wait time to wait for complete stabilization of the probes. The transient between two different biasing configurations is not so long, and in absence of charging is of the order of the ms at most, but for these initial measurements we opted for a safer and slower acquisition.

A final remark about the signal normalization. The signal used for the reconstruction algorithms should be normalized between 0 and 1, with zero representing no transmission and 1 representing total transmission. The corresponding normalization values were estimated by averaging a hundred measurements in conditions of blanked beam and probe entirely on detector, and later used for renormalization according to the formula:

$$sig_{norm} = \frac{sig_{exp} - sig_{dark}}{sig_{max} - sig_{dark}} \quad (12.1)$$

12.3 Microscale Imaging

The experimental measurements for microscale imaging used the aforementioned methods of acquisition. Here we show the images made by using the two Siemens star samples presented in Section 11. The first sample used was the largest one. The samples were produced with very different sizes, because the unusual alignment meant that it was not immediately clear how large the probes would actually turn out to be in the SAD plane. The need to integrate all the transmitted intensity on the detector defined an upper limit for the extent of the probes, that had to be confined within the inner and outer radii of the detector. Correctly positioning the large sample also proved to be a challenging task, and the first few attempts at imaging only showed edges between an opaque and a bright region. The first successful imaging attempt made apparent what the problem was, and the result is displayed in Figure 12.2.

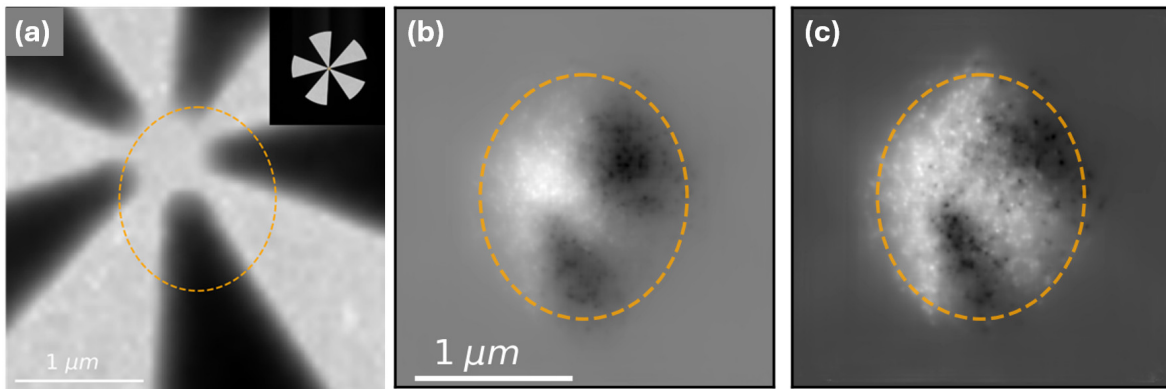


Figure 12.2: (a) Detail of the large Siemens star binary mask (full TEM image in the inset). The orange circle marks the imaged region. (b) Image of the sample reconstructed using the TGI algorithm and (c) the OGI algorithms, integrating 2500 different probes.

The issue lied in the fact that the probes were all concentrated in a very small area, of about $1\mu\text{m}$ in radius, and the only features of that size in the sample were the tips of the spokes in its center. Panels (b) and (c) show the results obtained by applying the two variants of the non-iterative reconstruction algorithms to the measured dataset. The Orthogonalized GI algorithm is slightly better at reconstructing the sample in the regions explored by fewer probes, like the bottom-right area inside the orange ellipses, but the two results are otherwise fairly similar. The probes had been concentrated in a small region in an attempt to see a good reconstruction even with only 2500 measurements, and this result already worked as a proof of concept. However, we decided to try with the smaller sample to see if the device could reconstruct finer features. The spokes of the smaller Siemens star were only a few hundred nm in the parts close to its center, so it was interesting to see if the setup could resolve those as well.

The measurement on the small Siemens star was made by increasing a little the range of the applied biases, pushing the probes to explore a larger region. This meant increasing the number of measurement points as well, to make sure that the probes would sample the larger area in a more uniform way. The long acquisition time was not a problem in this case, as the sample was basically not drifting at all during the measurement. The acquisition was made using 11600 different probes, distributed over an area of about 2

μm in radius. The reconstructions are presented in Figure 12.3.

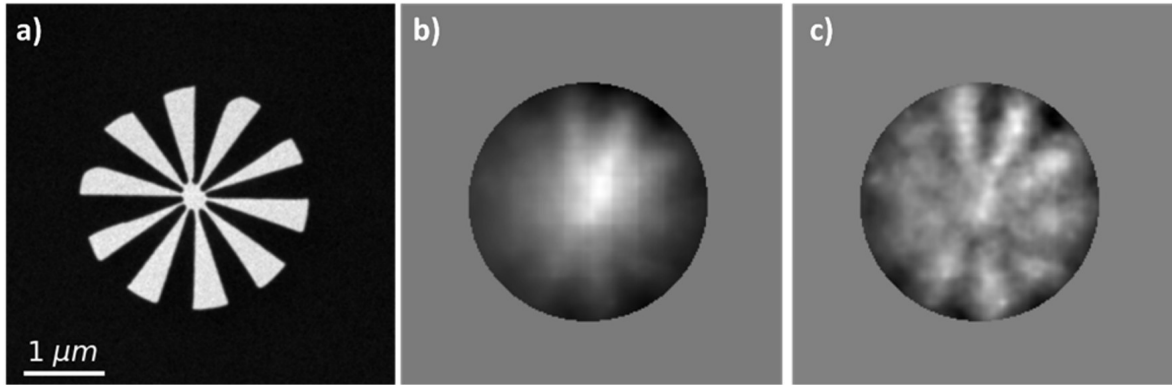


Figure 12.3: (a) TEM image of the small Siemens star binary mask. (b) Image of the sample reconstructed using the TGI algorithm and (c) the OGI algorithms, integrating 11600 different probes. The circular mask on top helps to identify the outer edge of the transparent part of the sample.

This second result better illustrated the ability of the Orthogonalized GI algorithm to recover a stronger contrast. Both reconstructions clearly show that one region was illuminated better than the right part of the sample was illuminated more frequently than the left one, because the recovered contrast is better on the right. This second experiment was repeated at a later date, about a month after the first one, so the microscope alignment was possibly slightly different and it might have caused this asymmetric illumination condition. The device and all the equipment were not altered in any way between measurements, so it is unlikely that the defect is related to it. With this result, we decide that the proof-of-concept stage was completed, because the method had technically worked, albeit poorly. At this point, rather than focusing the efforts on optimizing this setup, we started moving on to the next stage, towards atomic resolution imaging. Some issues of the experimental setup had already manifested during these initial experiments, and they were accounted for while planning future developments.

In particular, reading the HAADF signal from an external multimeter appeared to expose the system to a lot of noise and unforeseen perturbations, so one goal was to eliminate the multimeter from the setup. This issue was later solved with the introduction of the hybrid scanning-ECGI approach. Similarly, the acquisition time was too long for nanoscale systems, where the use of measurement times similar to the ones used at the microscale would imply subjecting the sample to a lot of damage and cause thermal drifts that would partially or completely compromise the measurements. Another issue was the imperfect matching of the simulated probes with the ones used experimentally, that could be an additional source of error during the reconstruction process.

13 Quadrichopstick Setup

The setup based on the QC phase plate, described in Section 8, had it positioned in the last condenser aperture of the microscope before the sample position. For the microscope used during the experiments, that was the C3 position, so there were two lenses and aperture positions before the device, that were used to select the illumination condition. That meant selecting simultaneously the area of the device that was illuminated and the diffraction-limited resolution. The two conditions used experimentally have already been presented in Section 9: the low-convergence illumination was used in a series of attempts made to image small groups of nanoparticles with sizes of a few nanometers, while the high-convergence probes were used in pursuit of atomic resolution imaging, because they exhibited features with spatial frequencies that are well-suited to sample atomic lattice fringes.

The rest of the experimental setup used the usual microscope optics. The signal needed for the reconstructions was the high-angle scattering signal collected at angles larger than 55mrad s, as tested in the multislice simulations of Subsection 7.2. The signal was collected with the HAADF detector, aligned in the proper on-axis geometry used in any STEM measurement, with the diffractogram positioned concentrically to the detector's annulus.

13.1 Acquisition of the Probes and Optimization

As explained in Section 9, the characterization of the QC phase plate showed that there was some effect that perturbed the ideal behavior of the device. In the case of the DS1 phase plate, the experimentally measured probes and the simulations mostly had minor differences, but it was still possible to identify a relationship between them. In the case of the QC phase plate, it was determined that the differences were mostly due to an uneven charging of the device that would change even after briefly retracting the phase plate from under the beam. Thus, it was extremely challenging to successfully model the experimental probes by means of numerical simulation. Fortunately, the microscope was equipped with a high-resolution Gatan K2 camera, that was used to measure the images of the probes used as illumination for each acquisition. When the microscope is in STEM mode, the imaging system is set to diffraction mode, to project the sample's diffraction onto the HAADF detector. Switching from diffraction to image when the phase plate was inserted under the beam would show the structured probe, that could then be imaged onto the camera. As long as the phase plate was not moved between the measurement of the signal and the acquisition of the probes, the charging would not evolve significantly, allowing to record probe images identical to those used on the sample, as demanded by the method.

The probe images were too large at native resolution to be used in numerical calculations, so they had to be rebinned to lower resolutions by means of cubic polynomial interpolation. The rebinning was done after calibrating the probe pixel size as described in Section 9, to avoid loss of information by remaining above the Nyquist sampling frequency. Given the different resolution limits of the low- and high-convergence probes, that meant being able to reduce the probe size from 3838×3710 pixels to about 256×256 or 512×512 pixels respectively, significantly reducing the time required for calculations

using the SPI algorithms. The experimental images were filtered with a top-hat function low-pass frequency filter, set at the probes' aperture limited resolution frequency, to remove any unwanted high-frequency noise.

The issue of phase plate charging produced a double inconvenience: if the measurement was made by acquiring the signal first and then the images of the probes, there would be no way to use the lateral overlap optimization strategy discussed in Subsection 7.3, but it would be possible to have a better control of the area to image, because the microscope could function in normal STEM mode. In the opposite case, when optimizing for probe acquisition order, it would be necessary to leave the device inserted while choosing the area to image. Thus, the STEM images would appear defocused due to the broadening of the probe caused by charging, seen in Figure 9.9. An alternative was to record the position of the area to image before inserting the device, and then use the navigation tool of the stage to automatically go back to it, but the limited precision of the stage made it possible to use this feature consistently only while imaging large areas and not at atomic resolution. The measurement signal-first was used for the high-convergence images, where the scan grid was so fine that the lateral overlap criterion was not applicable, and the convergence angle of illumination was actually large enough to see the lattice fringes. The measurement probes-first was used in the case of the low-convergence for the opposite reasons, since with very low convergence it would be much harder to determine whether the sample was precisely in focus in any case. In both cases, the acquisition time for the probes had to be set to a relatively low value, between 200 and 500 *ms*, because the acquisition function of the camera would intrinsically add about 700*ms* for an unknown reason, either to write on disc or for internal file management issues. Thus, the acquisition time per probe was about 1*s*, meaning that acquiring too many probes could also subject the phase plate to an evolution of the charging if it took too long. Although uncommon, this effect compromised a few measurements.

13.2 Signal Acquisition

Regarding signal measurements, the first experiments at low convergence still used the acquisition of the HAADF signal from the physical BNC exit of the detector, read by a Keithley 2400 multimeter, with 1*ms* of measurement time per experimental point, and a reduced wait time in between the measurement of signals, that was set to 100*ms* to wait for the transient between phase plate configurations. This meant that the first experiments with the QC phase plate still required to measure a few thousand probes over the span of ~ 10 minutes. Once the scanning was fully implemented in the code, the acquisition was switched to digital, and the signal was stored in the form of STEM HAADF images, one for each probe, containing the measurements for all the raster scan points for that probe. The dwell time per scan point was fixed to 200 μ *s* for 32x32 scans and 50 μ *s* for 64x64 scans, to maintain a low total measurement time while keeping a relatively low noise. For a few hundred different probes this meant acquiring hundreds of thousands of data points over about five minutes, which was a remarkable improvement in the amount of acquired information. Note that for these measurements at the nanoscale and at atomic resolution the total measurement time becomes an issue both in terms of delivered electron dose and thermal drift. The normalization values for the signal were still measured with blanked beam (reference zero) and beam fully

on the detector (reference max).

13.3 Low Convergence - Nanoscale Imaging

The early measurements at low convergence did not include the scanning, and were made basically with the same approach as the microscale images, with the extra step of probe acquisition at the beginning. Once the sample position was chosen and the device was centered, the list of biases used to generate the probe data set was re-applied in sequence to the phase plate to measure the values of HAADF signal from the detector. The first reconstructions showed here were obtained with a set of 7000 low-convergence probes, measured on the standard TEM sample of Au nanoparticles on amorphous carbon. These images were also the first ones on which we tested both the AP algorithm and the TGI one, before transitioning fully to the first one in later experiments. Figure 13.1 displays the reconstructions of two large areas.

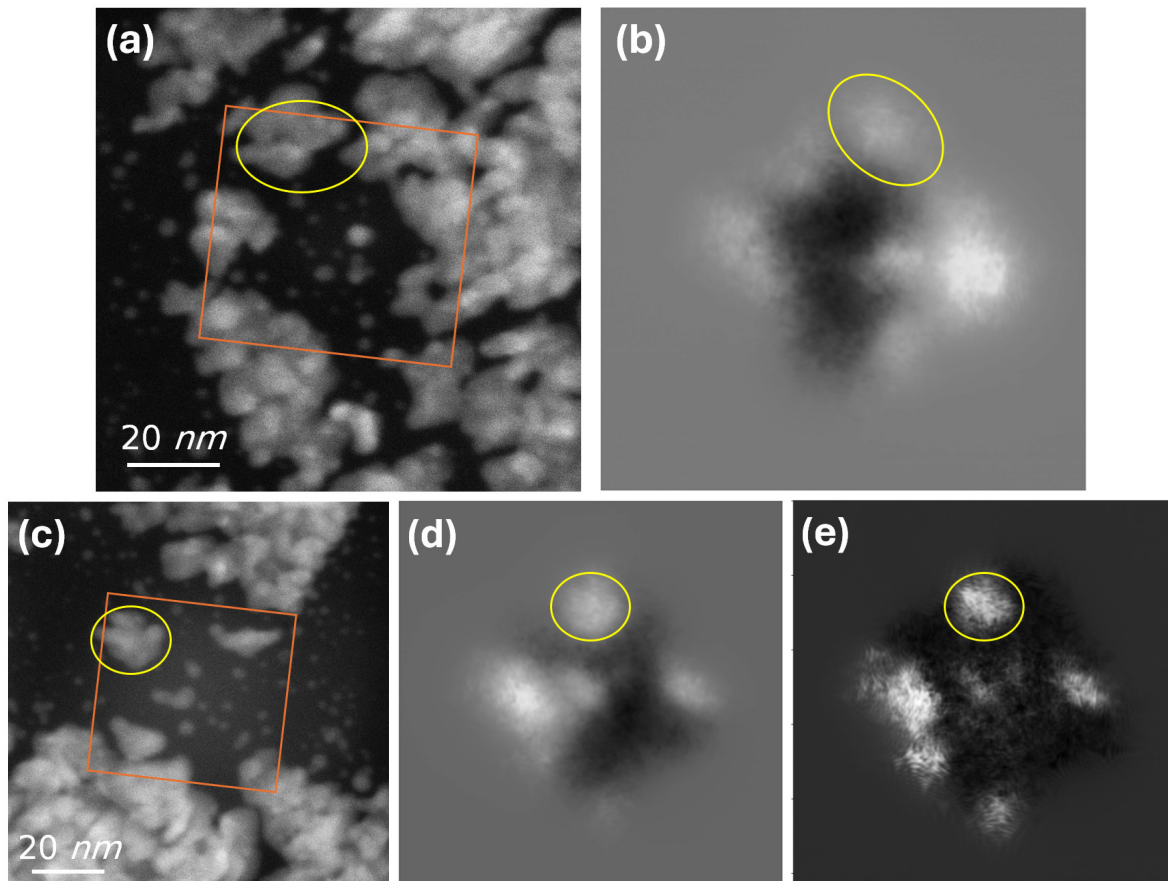


Figure 13.1: (a) STEM image of the first imaged region, acquired before the measurement. (b) Image of the sample in (a) reconstructed using the TGI algorithm. Only the very lowest frequencies were visible, all the fine details were lost. (c) STEM image of the second imaged region, acquired after all the measurements. Notice that the imaged area is brighter due to contamination. (d) TGI reconstruction and (e) AP reconstruction of (c). The AP algorithm performed better, as expected, and produced better defined edges, but also accentuated some features of the single probes. The rotation of the reconstructed images is due to the rotation angle between the plane of the phase plate and the image plane on the camera that acquired the probes.

These results are very similar to the ones at the microscale, where the low-frequency features are visible, but the algorithms struggle to recover finer features, like the smaller nanoparticles. At this stage, the main drawback of the method was the long measurement time, that prevented the acquisition of a large number of probes. Using more probes would have facilitated the reconstruction. The two images to be reconstructed were 512×512 pixels, so the sampling ratio with 7000 probes was $N_{meas}/N_{pixels} = \frac{7000}{512^2} = 2.67\%$, extremely low. As mentioned in Section 2, a perfect reconstruction of an image of N pixels requires at least N *orthogonal* illuminations, so a sampling ratio of at least 1 (orthogonal as in having zero overlap). Any lower value of sampling ratio decreases the quality of the image, and if the illuminations are not orthogonal the proper reconstruction is even more difficult. So, it was clear that we needed to increase the sampling ratio to improve the quality of the results. After this experimental session, a number of other tests were made to add the scanning of the probes to increase the sampling ratio.

The next batch of results was produced using fewer different probes, that were scanned on small 32×32 or 64×64 grids, bringing the number of acquired probes from a few thousands to a few hundred thousands. The acquisition of the signal was also changed, because the HAADF signal was measured through the electronics of the Gatan Digiscan system instead of the external multimeter. Figures 13.2 through 13.4, show a few different images obtained by scanning 150 different probes on grids 32×32 , with a step of 0.9808 nm , and HAADF signal measured with a dwell time of $200 \mu\text{s}$ for each scan point. Thus, the sampling ratio increased to values between 50% and 200%.

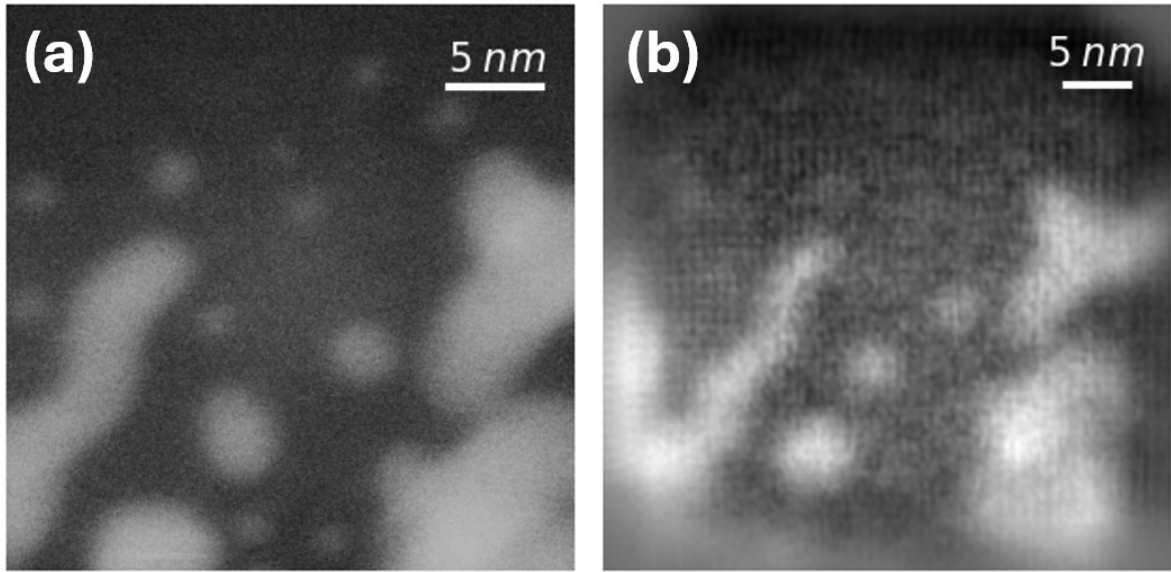


Figure 13.2: (a) STEM image of the region selected for imaging, acquired before the measurement. The blurriness is caused by the broadening of the STEM probe induced by charging. (b) Image of the sample reconstructed using the AP algorithm. The extended field of view is accessed thanks to the structured probes, that cover a larger area compared to the normal STEM probe at the same magnification.

From this image it is immediately clear that the smallest particles are not easily imaged, but also that the method can recover differences in contrast inside the Au islands (see the bottom left corner of panel (b), with a bright spot surrounded by grayer parts).

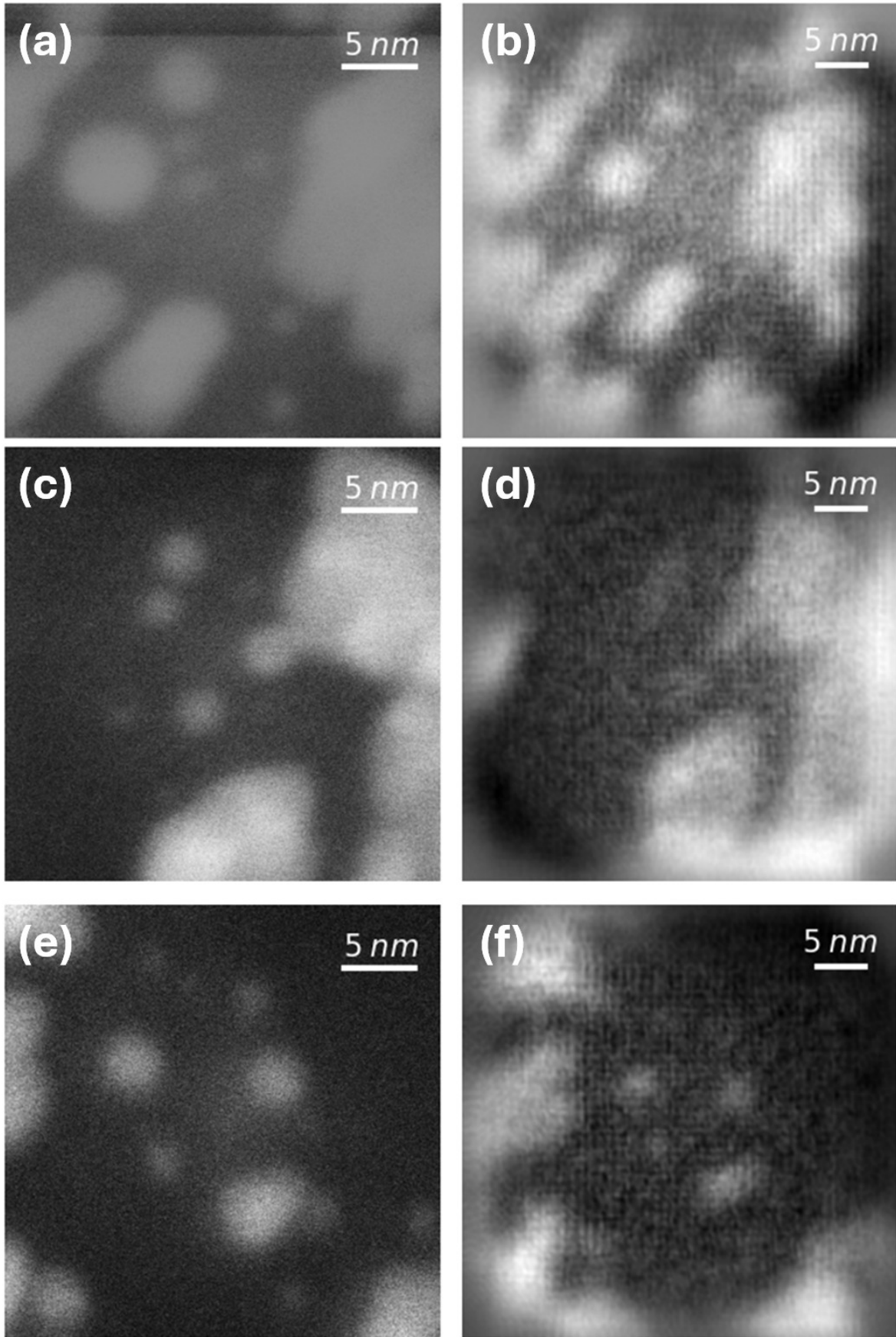


Figure 13.3: (a)(c)(e) STEM images of other imaged regions, acquired before the measurement. (b)(d)(f) Images of the samples reconstructed using the AP algorithm.

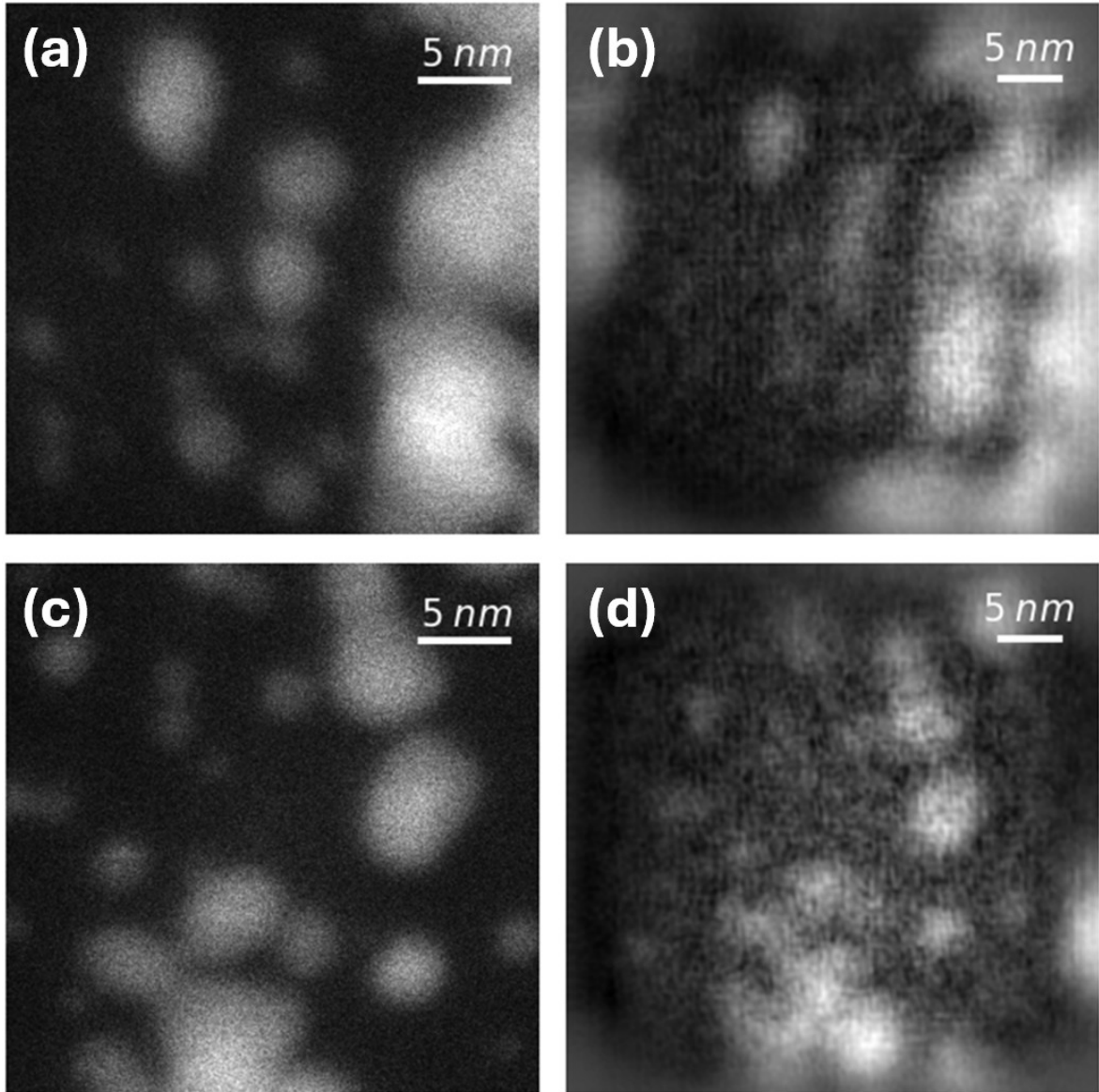


Figure 13.4: (a)(c) STEM images of other imaged regions, acquired before the measurement. (b)(d) Images of the samples reconstructed using the AP algorithm.

Notice how these reconstructed images are actually square as expected from the simulations (if one ignores the blurry areas close to the edges and corners). In all these images it is possible to see a fine grid of points, that are the shifting artifacts already observed in the simulations of Section 7. Their frequency matches exactly that of the STEM scan grid used for the acquisition of the signal. In all the images, the smaller particles that are well distinct in the normal STEM images can not be resolved easily. This might be caused by a combination of multiple effects:

- low signal coming from these tiny particles, that causes them to stand out less on the background;
- sample thermal/mechanical drift;
- beam-induced contamination and damage;

- artifacts of the AP algorithm that make it harder to see the dimmest variations of contrast in the recovered image, and become more visible after many iterations.

The same issue is present also in images recovered using a finer scan grid of 64×64 points, that increased the number of measurement points by a factor 4. Figure 13.5 shows the two images recovered in these conditions, with 150 probes scanned over 64×64 points at the same magnification as the previous images, using half of the previous scan step, so 0.4904 nm . The scanning artifacts are now harder to see, but the quality of the reconstructions has not improved dramatically.

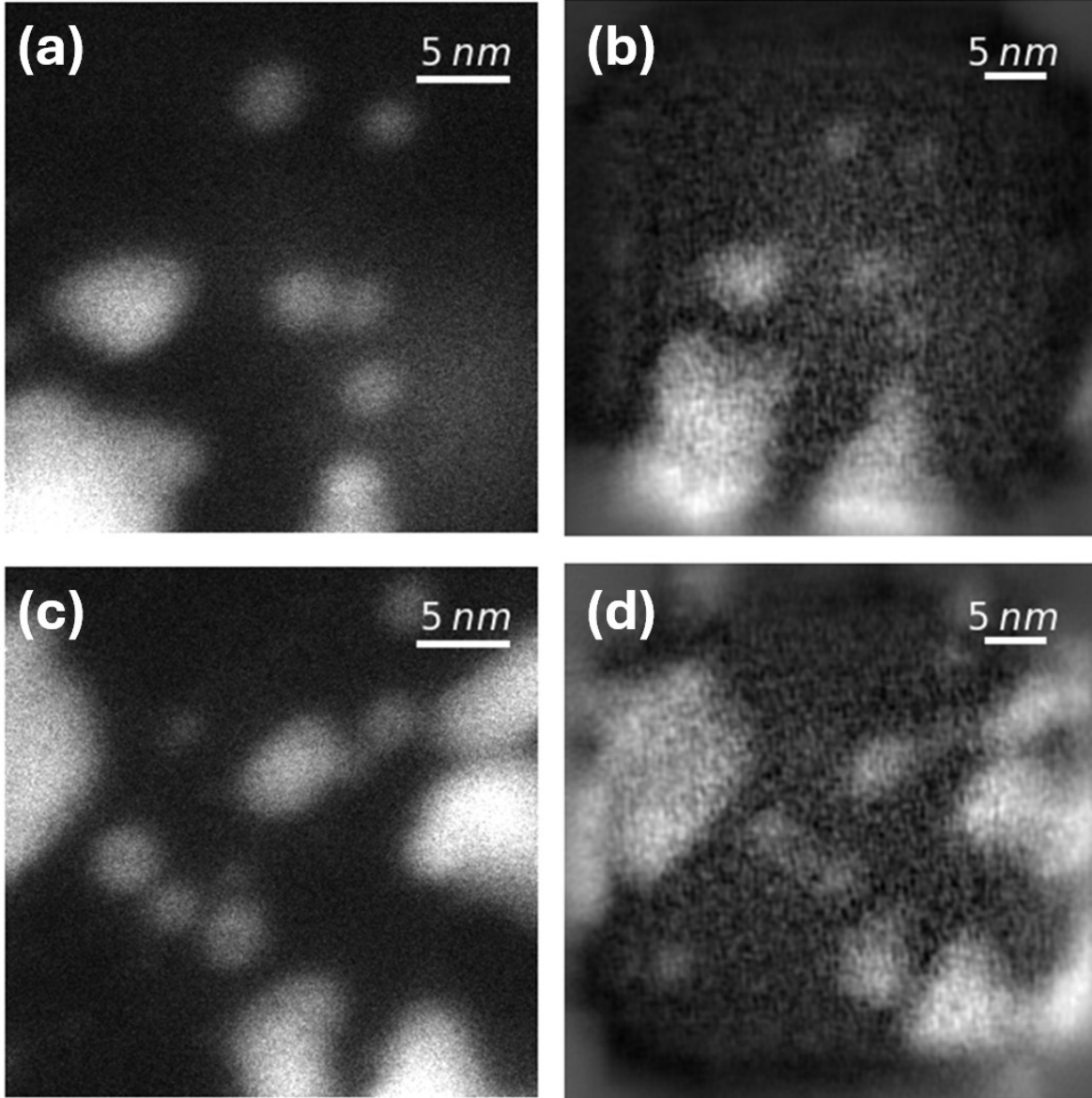


Figure 13.5: (a)(c) STEM images of the first imaged region, acquired before the measurement. (b)(d) Image of the samples reconstructed using the AP algorithm, with a halved scan step compared to the previous examples.

The drift and the beam-induced damage and contamination can be seen in Figure 13.6. The *before* and *after* images were measured at the beginning and end of acquisitions of 1000 probes, for a total measurement time of ~ 15 minutes. The signals from the scan of the 150 probes used for the reconstructions were acquired in the first 2.5 minutes,

so it is perfectly possible that these effects alone have altered the sample enough to perturb the reconstruction process.

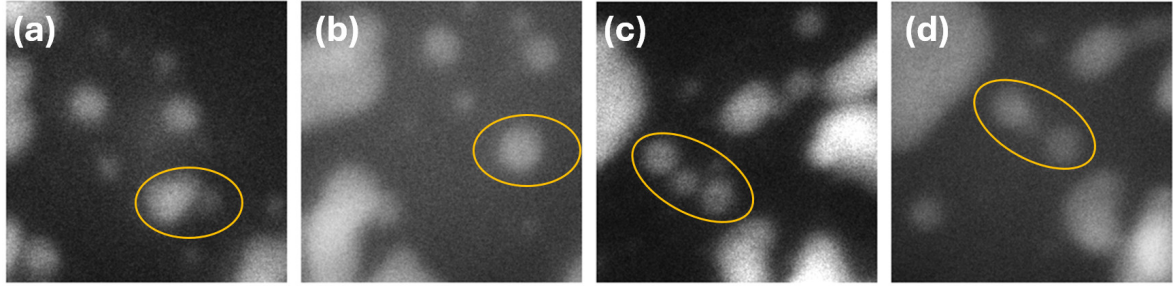


Figure 13.6: STEM images from two previously showed positions, (a)(c) before a measurement and (b)(d) after. The circles show how the smaller particles can merge together during the course of an acquisition, making it harder to resolve them because their signal keeps changing over time. The contamination is visible in the increased brightness of the carbon substrate. Notice also the large drift between the *before* and *after* conditions.

About the artifacts produced by the AP algorithm, if we look at the Fourier transform of one of the images made on the 32x32 scan grid, the shifting artifacts are clearly visible, and can be filtered out selectively. The result is displayed in Figure 13.7.

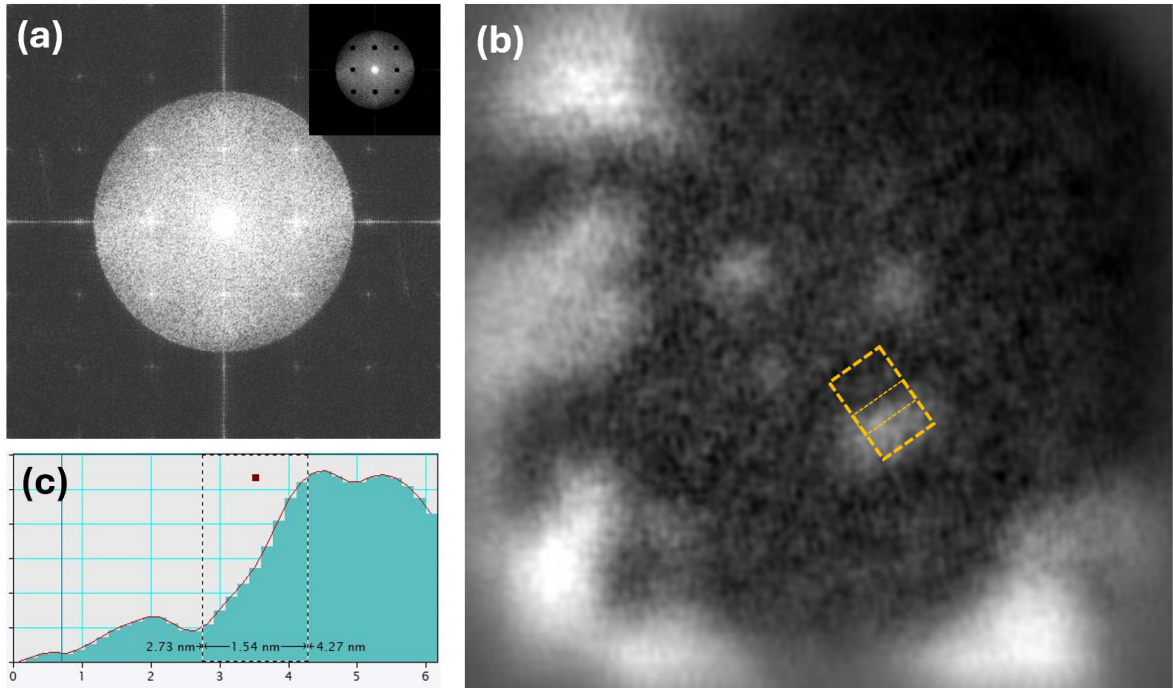


Figure 13.7: (a) FFT of the reconstruction from Figure 13.3(f), showing the bright spots associated to the shift artifacts. If they are filtered out (inset) and the image is inverse Fourier transformed, it reveals (b) a network of non-periodic oscillating features similar to random noise, that adds artificial fine features to the recovered image. This makes it hard to estimate the resolution of the image, but by looking at (c) an average profile taken across and edge of a particle, the contrast seems to vary over a distance of $\sim 1.5\text{nm}$, so the resolution can be estimated at that value.

From the profile in panel (c), the estimated resolution would be around $1.5nm$, but it is difficult to find a criterion that can provide a strong quantitative evaluation of the actual resolution accessible with the low-convergence probes. That is because at low convergence the probes themselves have an aperture limited resolution larger than the characteristic spatial frequencies of the sample, so it is impossible to see lattice fringes, and the smallest nanoparticles yield a very low signal, so their contrast is not particularly enhanced. However, by looking at the Fourier transform of the reconstructions, we can get some more insight on the results. Figure 13.8 displays the radial averages of the FT of each image, calculated as:

$$radial_average(r) = \frac{\sum_{r-0.5}^{r+0.5} \ln |\mathcal{F}(image)|^2(r)}{2\pi r} \quad (13.1)$$

with r being the distance in pixels from the center of the FT.

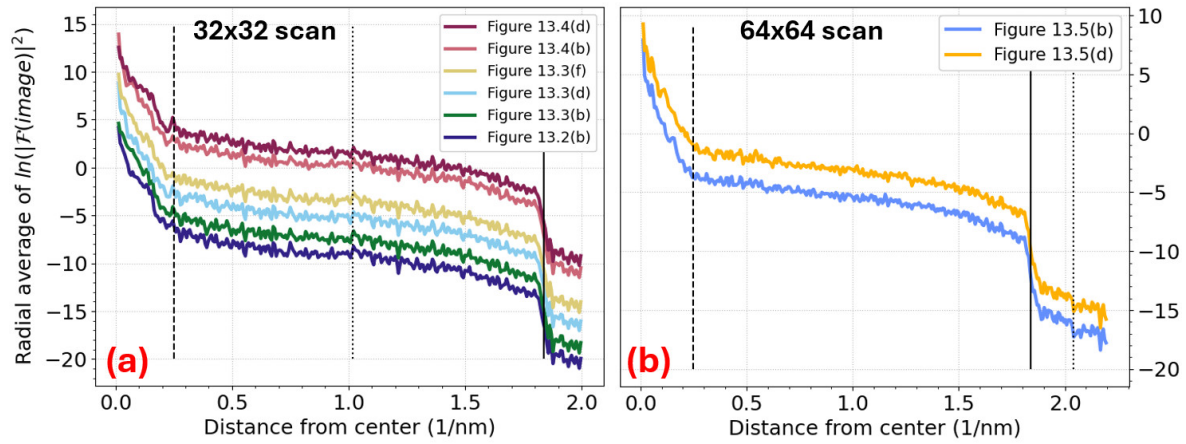


Figure 13.8: (a) Radial averages of the FT of the experimental reconstruction made with scans of a 32x32 grid and (b) 64x64 grid. The dotted lines mark the frequencies of the scan step of the grid, that as suggested earlier are below and above the resolution limit for the two different scan steps. The dashed lines mark the frequency corresponding to $4nm$, where there is a first change in the trend of the radial average. The solid line marks the cutoff frequency of the low-pass filter used for the denoising of the probe images, set at the aperture-limited resolution value. Each curve is shifted by two units vertically, to improve legibility.

The dashed lines in both panels was put as a reference of the spatial frequency $\frac{1}{4nm}$, that more or less separates the low-frequency information resolved simply because of the size of the probes from the high-frequency information made accessible because of their fine details. Next, notice how all the profiles appear to have an inflection point around the frequency of $1nm^{-1}$. In panel (a) the change in curvature between the two parts is more apparent because of the scan artifacts, that sit very close to that frequency, but the same consideration holds for panel (b), where the scan artifacts are at a higher frequency. The radial average decreases by several units past the cutoff frequency used to filter the high-frequency noise from the probes. It was set at the value of the aperture-limited resolution, and everything above that value is basically just high-frequency noise. The interesting aspect of these profiles is that they show that there is information all the way up to the resolution limit for the probes, marked by the solid lines. At the same time, since the profiles decrease by several units past the

frequency of $1nm$ up to the aperture cutoff, the high-frequency content of the image gets progressively lower as the frequency increases. It is apparent from all the experimental images that particles with a radius of $5nm$ or slightly smaller are still visible, and even the smaller particles appear with very low contrast and can be identified when comparing the STEM images with the reconstructions. Considering both the analysis of the variation of contrast at the edge between carbon and gold, and the radial profiles of the FT of the images, we can give a realistic estimate of the resolution to be between 1.0 and $1.5nm$. Despite that, we want to remark that the information extends up to the frequency of the aperture limited resolution, that for the low-convergence probes is 1.85 nm^{-1} .

All the images showed previously were obtained after 10 iterations over the respective data set using the AP algorithm. After the 10th iteration, the artifacts start to noticeably change the contrast of the reconstructions. An example is presented in Figure 13.9.

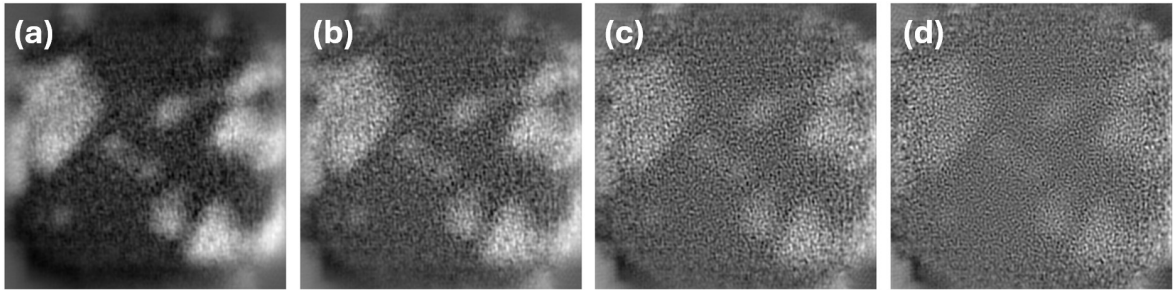


Figure 13.9: Example of artifacts emerging in the reconstruction of experimental data with the AP algorithm. Reconstruction at (a) iteration 10, (b) 30, (c) 100 and (d) 300. The contrast is progressively flattened as the "algorithmic noise" increases with each iteration.

Note that this effect was not observed in simulations, despite running the AP algorithm for hundreds of iterations. A plausible explanation might be that the simulations are noiseless, while our experimental setup constrained us to use both the signal from the HAADF (that would statistically be noisy, to some extent) and the noisy images of the probes acquired on the camera. The probes were processed and filtered to reduce their noise, but part of it could not be removed without altering the details of the structured probes as well. The applications of Computational Ghost Imaging seen in Section 2 typically use noiseless simulated images of the illumination, so our constraint to use experimental images of the probes is definitely an additional source of error.

Because the code of the reconstruction algorithm required to match the scan step with the pixel size of the probes, we decided to simulate the effect of using a mismatched scan step in the code, to be sure that the kind of artifact presented in Figure 13.9 was not arising from a simple miscalibration of the distances used to combine the images in the code. The simulation was made with the same parameters of the atomic resolution simulated image at the end of Section 7. Both the sample and the data set of probes were calculated with a pixel size that was 5% larger, going from $0.42\text{\AA}$ to $0.441\text{\AA}$. The images at larger pixel size were used to simulate the signal obtained from a 64×64 pixel scan with scan step equal to the pixel size. Then, the AP algorithm was used to

reconstruct the image of the sample by combining the signal calculated with the larger pixel size and the old probes with smaller pixel size. The result is in Figure 13.10.

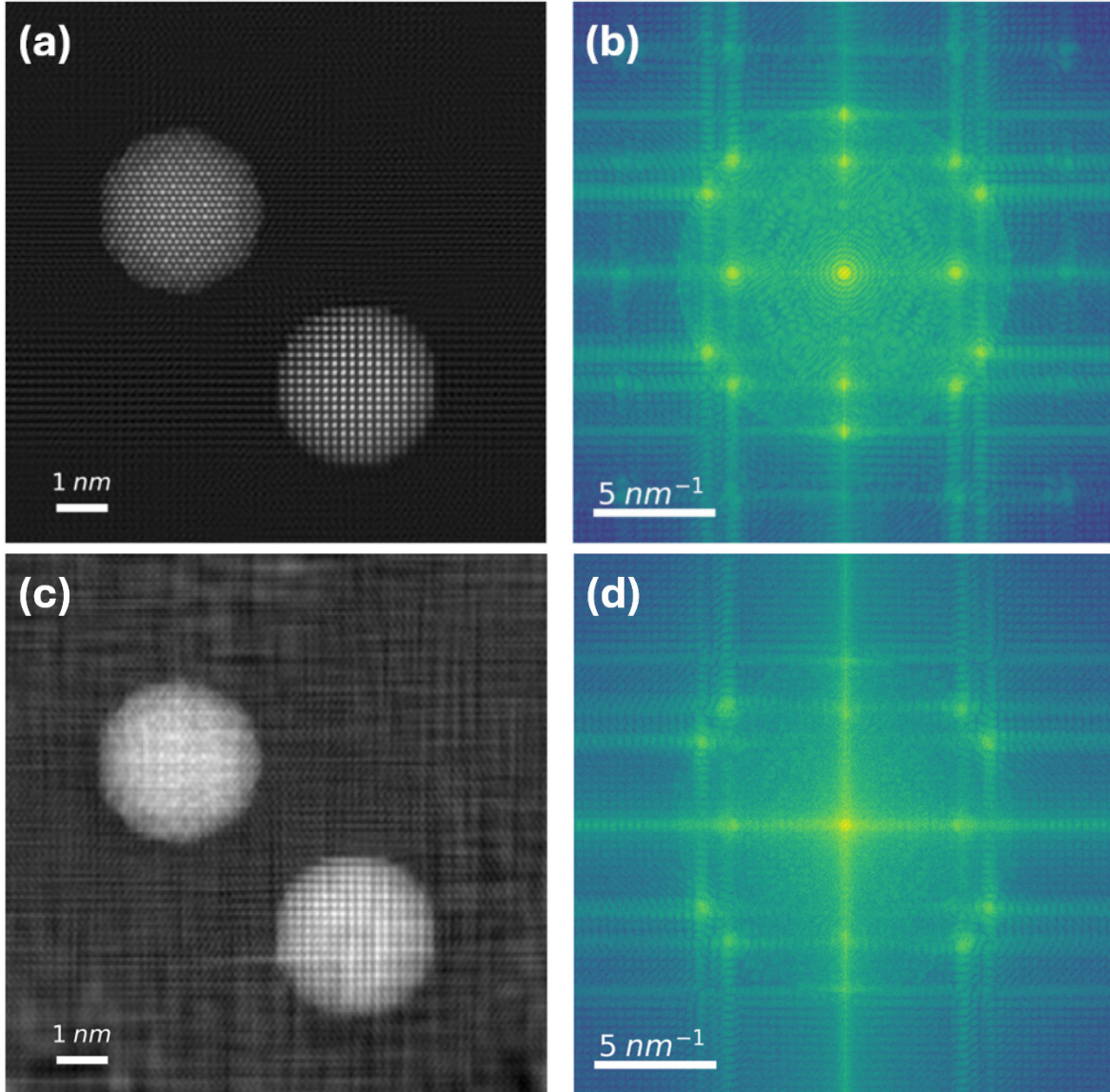


Figure 13.10: (a) AP simulated reconstruction of the Au nanoparticles showed in Sub-section 7.3, and (b) its Fourier transform. (c) Result obtained with the "mismatched scan" signal, and (d) its Fourier transform, recovered in the same conditions of (iterations)*(N probes)*(scan points).

Panel (c) shows a surprisingly good reconstruction, similar to the one in panel (a), with correctly recovered lattice fringes on the particles and the same lattice fringe artifacts on vacuum. The "mismatched scan" image also has an additional set of artifacts in the form of a network of vertical and horizontal stripes that alter the contrast. This result is interesting, but it does not resemble the effect seen in Figure 13.9, where there are not stripes, but rather some wavy shapes. So, probably the unidentified artifact is not due to mismatched distance calibrations in the code. Identifying its origin is quite important, because it prevents the execution of the image reconstruction code for large numbers of iterations over the dataset, that are normally used to reach more accurate results.

13.4 High Convergence - Atomic Resolution Imaging

After obtaining the results at low convergence, we moved on to the high convergence angle illumination condition to attempt to see some images at higher resolution. The measurements at high-convergence were made using two grids with different scan step: a broad scan, with similar parameters to the low-convergence measurements, on a 64×64 scan grid with step size of 0.4778 nm , and a fine scan, 64×64 with step size of $0.42 \text{ \AA}$, so about 10 times smaller. The purpose was to see if imaging with a total field of view analogous to that of the previous low-convergence measurements but the high-convergence probes would display an increased resolution. The finer scans were instead made with the objective of reconstructing information at atomic resolution. In order to be sure that the imaged region exhibited some features at high resolution, the acquisition of the signal and probe images was switched. By measuring the signal first, and then the corresponding probe images, it was possible to center the imaging region on the sample using the ordinary STEM mode of the microscope, unperturbed by the device. The phase plate was only inserted at the time of signal measurement, and afterwards the sample was moved to a broken window of the carbon grating to acquire the images of the probes in vacuum.

The target area of the sample for the broad scan and the corresponding reconstructed image are displayed in Figure 13.11.

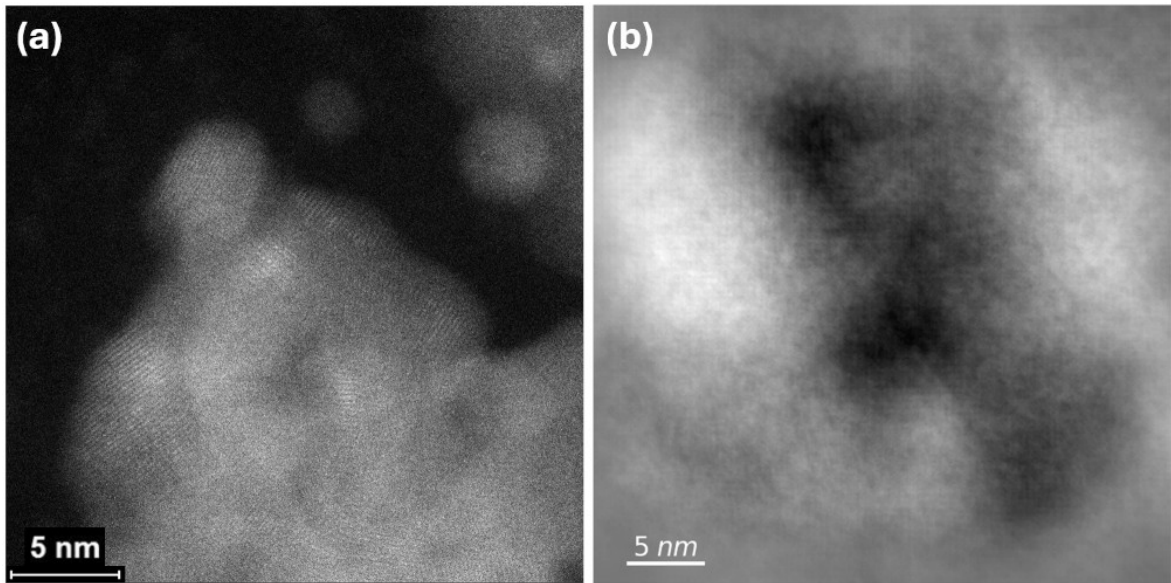


Figure 13.11: (a) STEM image of the area chosen for imaging, note that now the lattice fringes are visible. (b) AP reconstruction, made with only one iteration over the data set acquired on the broad scan grid. The two regions appear vastly different, but it is unclear why. Reconstructing over only a few more iterations already creates the same kind of artifact of Figure 13.9.

Differently from the results at low convergence, in this case the recovered image in panel (b) is not matching at all the corresponding STEM image. The cause of this issue is unclear, as the signal was measured immediately after acquiring the STEM image and inserting the device. It is unlikely that the sample could have drifted so far from the chosen position, but it is possible that the phase plate introduced some tilt

along with the beam shaping effect. The tilt would have moved the illuminated area to a different position of the sample, effectively creating this unforeseen issue. In the low-convergence measurements that was never a problem, because the reference STEM images were acquired when the device was *already inserted*. Unfortunately, this issue did not become apparent until a few days later, during the analysis of the experimental data. Also, since the sample was drifting significantly and speed was of the essence to make a measurement of a fixed area, I tried spending as little time as possible between the moment in which I selected the region to image and the actual measurement of the HAADF signal. Ideally, I should have invested some more time in using a very concentrated probe to take a blurry image of the sample *after* inserting the phase plate. That image would have worked perfectly well as a secondary reference for the results, and it could also have been processed to be used as prior information for the reconstruction, to initialize the reconstructed image to a non-zero array. The need for regularization of the reconstruction algorithm with prior information was also a problem in the early days of Ptychography, so it is not too surprising to see that a similar need arises here as well, given the similarities between the two methods pointed out in Section 5.

Another notable observation is that the reconstruction displayed in Figure 13.11 is calculated over a *single AP iteration* of the data set, because of a stronger version of the "algorithmic noise" artifact. When the AP algorithm is run on the high-convergence data, it is far more prone to create the unusual artifact seen before. It appears about ten times faster, because on the low-convergence data it became visible between the 10th and 20th iteration, and got progressively more severe over hundreds of iterations, while in this case it appears already at the second one. Figure 13.12 shows what happens to this data set in the first 20 iterations.

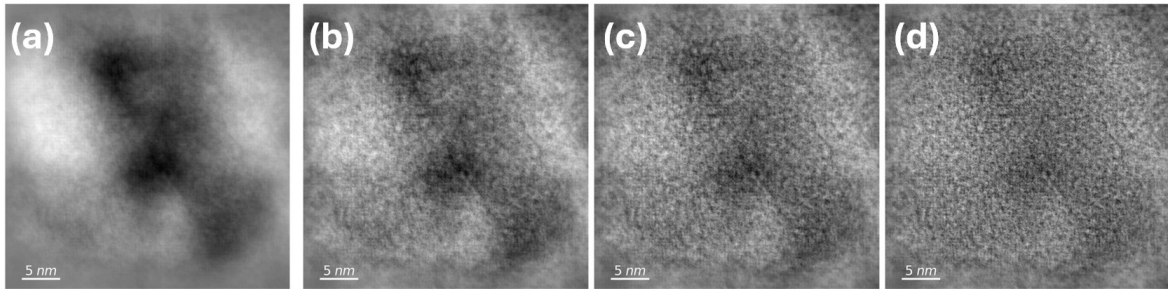


Figure 13.12: Progressive onset of the "algorithmic noise" artifact in the high-convergence results. The four images show the result from Figure 13.11 at (a) 1 iteration, (b) at 5 iterations, (c) at 10 and (d) at 20.

When the fine scan grid is used, the effect becomes even more pronounced, because in those conditions all the probes are almost confined in the same position, as the side of the scanning grid is $0.42\text{\AA} \cdot 64 \text{ points} = 2.69\text{nm}$. If this artifact is caused, as I suspect, from the combined effect of the noise on the HAADF signal and the noise from the experimental images, then it would make sense to see a quicker onset when the probes are all concentrated almost in the same position. The difference between the low and high-convergence probes is the exposure time, that was increased from 200ms to 500ms to capture the finest details of the high-convergence probes, so the image noise was also increased. Figures 13.13 and 13.14 show the results obtained using the fine scan grid.

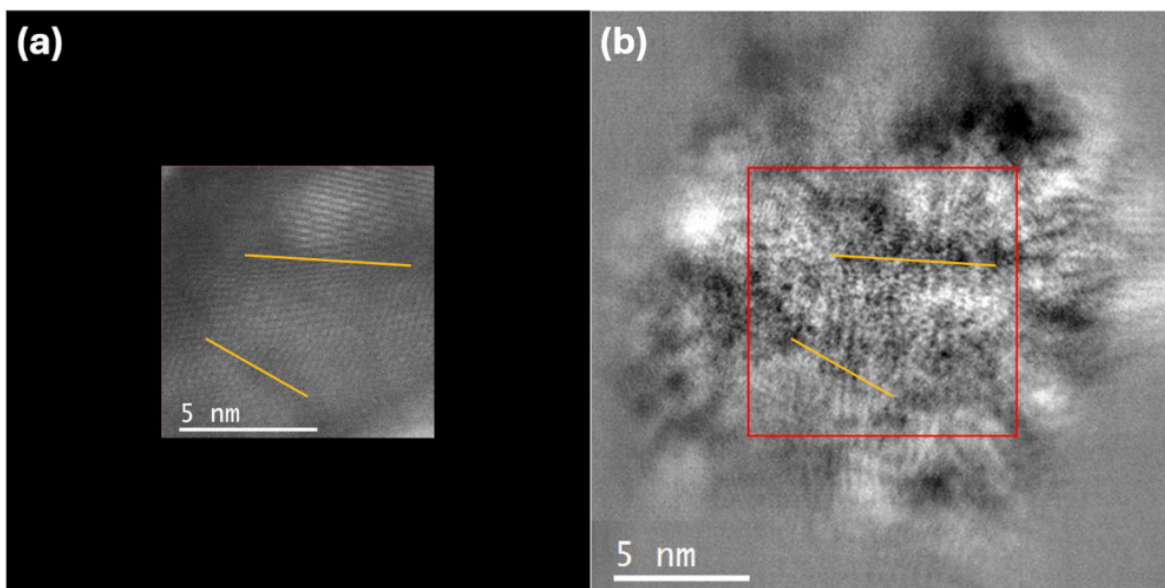


Figure 13.13: (a) STEM image of the area chosen for imaging and (b) AP reconstruction, made with only one iteration over the data set acquired on the fine scan grid. The square and lines help to compare the two images by separating three most prominent features.

The first reconstruction on the fine scan grid shows contrast that is not immediately associated to that of the expected imaging area. The frame and guide lines help to identify three regions separated by two slightly darker ridges, but still, a result obtained after a single iteration of the AP algorithm is not to be trusted entirely. Considering this, one might wonder if it could be worth it to fall back to the TGI algorithm but the results of the two algorithms after a single iteration are very similar. Figure 13.14 shows the results for the second region.

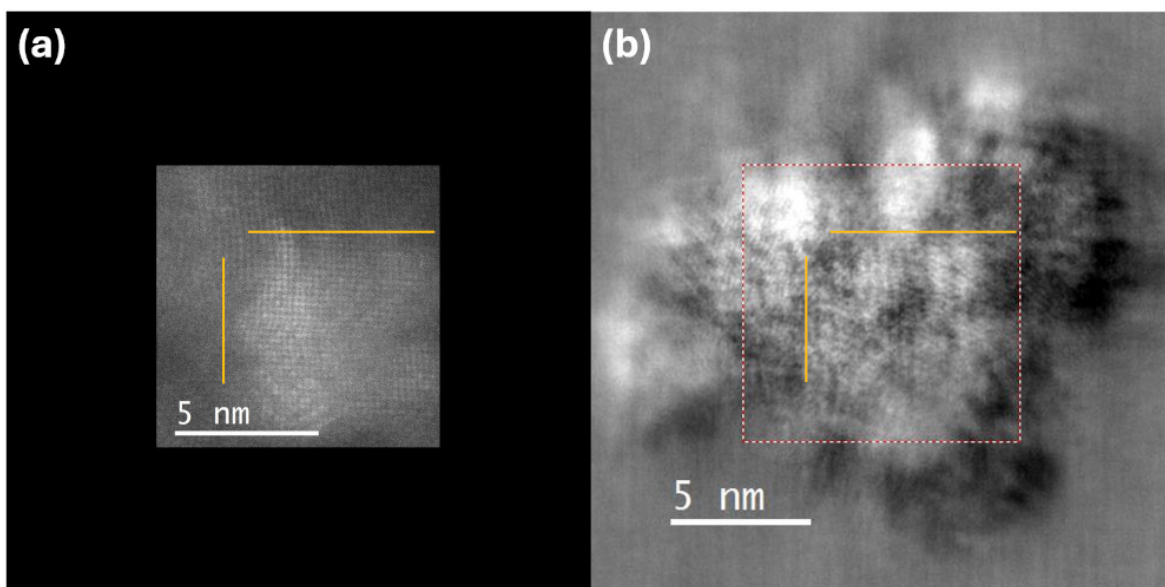


Figure 13.14: (a) STEM image of the area chosen for imaging and (b) AP reconstruction, made with only one iteration over the data set acquired on the fine scan grid. Notice how this reconstruction, like the one in the previous Figure, is no longer square with sides parallel to the frame, this happens when the scan grid is much smaller than the area explored by the probes. The square and lines help to compare the two images.

For what concerns these images, similar considerations apply for both. I tried adding some lines in both images to mark some features that seem to appear in the reconstructed images. For the high-convergence probes there seems to be a problem in matching both the high and low frequencies at the same time. The two recovered images show variations of contrast that vaguely match the low frequency part of the STEM images, while the variations of contrast at high frequency are not matching the lattice fringes that we see clearly in both STEM images. A possible explanation for this problem is that these seemingly erratic variations in contrast are actually due to channeling, because the high-convergence probes have finer features and can potentially produce a strong signal by interacting with a particle oriented favorably in the vicinity of the imaging area.

A careful inspection of the two reconstructed images also reveals that there are some high-frequency fringes appearing in both images. However, they are most likely just another kind of reconstruction artifact, as they appear in the same positions in both images. They seem to emerge from the combination of the high-convergence probes in a small raster, because they are not visible in the reconstruction on the broad scan grid. Since the set of illumination probes was different for each image, but these fringes appear in both fine-grid images, they are most likely just an emergent effect. Figure 13.15 points out the matching fringe-like features.

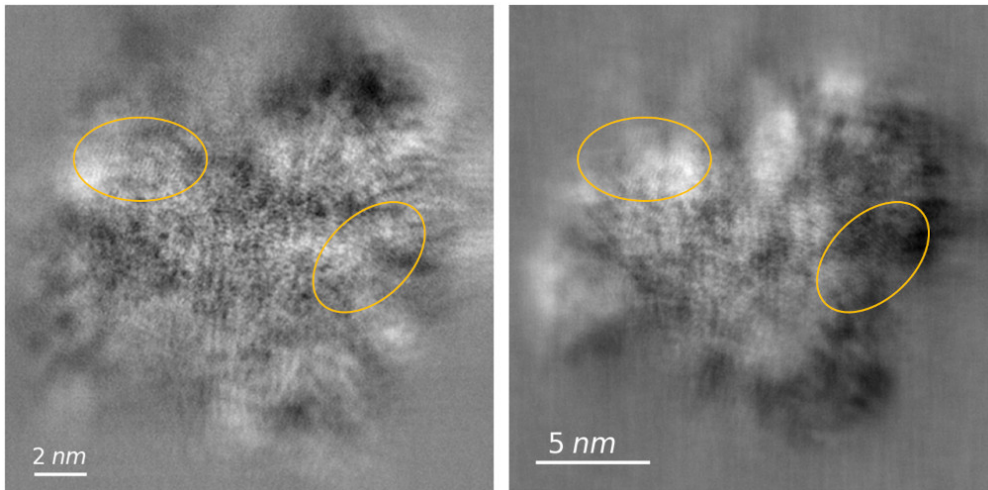


Figure 13.15: Corresponding positions of the lattice-fringe-like artifacts.

One possible reason for the manifestation of these artifacts in both reconstructions could be that the symmetry of the device, combined with the biasing condition used for these measurements, favored the generation of fringes in the patterns along certain preferential directions. Then, as they appear more frequently in the database of probes, it is more likely for them to probe their corresponding spatial frequency in the sample and provide a larger contribution to the reconstruction. As an example, let us imagine that the sample is a square lattice with vectors parallel to two planar x and y axes. If the symmetry of the phase plate favors the generation of fringes in the x or the y direction, rather than at an angle, it is more likely that those frequencies will be pushed up in the reconstructions, as they are measured more often and have a greater change to find a condition of large overlap. This is also another aspect that pushes forward the issue of matching both the low and the high frequencies of a sample in the reconstructions.

13.5 Some considerations about the experimental protocol

After seeing all the results obtained so far, it is instructive to spend a few words about the issues of the experimental protocol and its aspects that could be improved. The first issue that comes to mind is undoubtedly the current level of control and characterization of the electrostatic phase plates used to generate the structured illumination. The production of the devices is realized in a research facility, and consequently the phase plates are basically just prototypes, from a technical point of view. Thus, they are more exposed to issues like fabrication defects, contamination, and charging than what serially produced devices would be. Those issues forced us to adopt a less conventional implementation of the technique. In fact, they made it very complex to numerically reproduce the structured illuminations, as it is customarily done in light-optics, forcing us to resort to the use of noisy experimental images of the illuminating probes. Future works should employ strategies that either avoid putting the device directly under the beam, or do so in a way that is consistently reproducible through simulations. Nonetheless, it was encouraging to see that the apparatus worked despite this technical difficulty, and the quality of the results can only improve when it is solved.

The second limitation concerns the measurement of the signal from the sample. While presenting the results, it was discussed how the signal from the HAADF detector was fundamentally left unprocessed in most cases, except for its normalization within the range $[0,1]$, and it was not filtered to reduce noise. Another way to push the research forward would be to characterize the output signal from the HAADF detector, and devise strategies to minimize its noise. This pairs with the previous indication to make it possible to simulate the images instead of measuring them. Solving both these problems would ideally remove or minimize the unidentified artifact noise that was limiting the application of the AP algorithm in most cases. Another possible study would be to try to determine the effect of the change in collection angle on the experimental results. All the measurements showed here were made with an inner collection angle of 55 mrad , close to the formal criterion specified in Subsection 1.7 for the onset of the HAADF regime (convergence angle + 40 mrad). However, the results of the multislice simulations of Section 7.2 showed a dependency of the linearity of the measurement on the collection angle, so it would be interesting to investigate it further from an experimental point of view. Of course, our choice for the collection angle was oriented towards the maximization of the HAADF signal, and higher inner collection angles would likely worsen the S/N ratio, but it is a possible path to explore for further works.

Another remark about the HAADF signal that was not stressed previously in the discussion of the results with the QC phase plate is that the HAADF detector of the microscope used for the experiments frequently exhibited some kind of malfunction, with discontinuities in the measured signal. An example is already visible in Figure 13.3(a), where at the top there is a clear discontinuity in the contrast. This kind of effect appeared in completely erratic ways, and sometimes it even completely prevented the acquisition of normal STEM images. Two examples are shown in Figure 13.16. In panel (a), an image at low magnification of the Au nanoarticles. There are horizontal stripes that correspond to times in which the detector would temporarily change its gain and contrast values. In panel (b) there is an Au nanoparticle imaged at atomic resolution, where towards the end of the scan the detector went out completely.

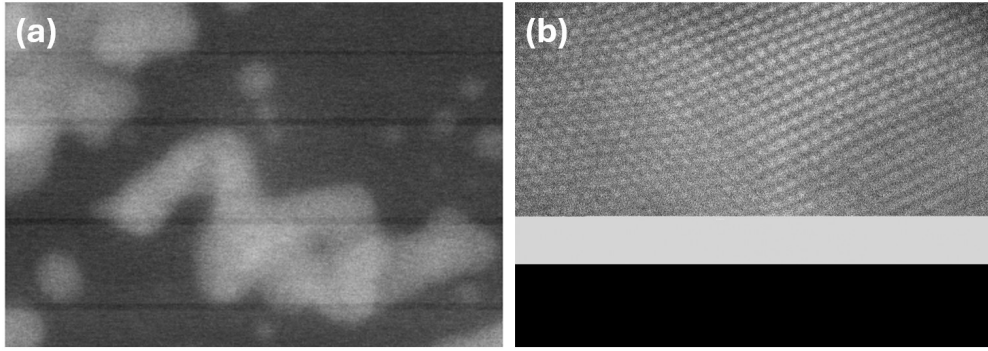


Figure 13.16: Two images from the Au nanoparticles sample. (a) Stripe artifacts generated as the gain and contrast of the detector changed erratically. (b) An example of how the detector could suddenly stop working during a measurement.

The example of panel (b) is easy to spot, because the detector suddenly stops working and when the measurements of the HAADF signal for an ECGI acquisition are plotted the curve suddenly drops to negative values (after normalization, so below the minimum). Such events were uncommon, and they never directly affected the results of a measurement. The same can not be said for the "stripes" appearing in panel (a). We were unable to identify the exact cause of this behavior, but given the fact that it is altering the signal only by a little bit, it is also very difficult to say if it ever happened during measurements, because the acquisitions were made on scans of 32×32 or 64×64 points where there is already a modulation given by the changes in probe-sample overlap and noise. So, this is an additional effect that could have worsened the results. The only solution in this case would probably be to change the detector entirely, and it was not feasible within the time frame of my experiments.

Some other observations can be made about the software side of the method. The acquisition of the experimental measurements took a long time, both for the images of the probes and most importantly for the HAADF signal. The sensitivity of the samples to electron beam damage and contamination requires to adopt strategies that minimize the acquisition time, and this also means optimizing the software to remove all the unnecessary dead times. Also, concerning specifically the measurements that used the periodic raster scan to maximize the amount of data collected, it would be best to draw inspiration from other techniques that employ probe scanning, like Ptychography. In cases where the scanning is necessary, the acquisition is made on a non-periodic grid, where the positions of the scan points are not equally spaced but rather distributed in other geometries, or with tiny perturbations of a periodic raster to remove translation artifacts in the reconstructed images. In our results, the use of periodic scanning created additional artifacts that complicated the estimation of the resolution. Future experimental setups that use scanning should be adapted to account for this effect.

14 Conclusion

The research presented in this thesis aimed to create a fully functional experimental protocol that adapted a light-optics technique, Computational Ghost Imaging, to Transmission Electron Microscopy, with the goal of imaging amplitude samples at atomic resolution, and possibly also beyond the aberration limited resolution. At the current stage, our version of the Electron Computational Ghost Imaging technique is realized with a proposed apparatus that relies on two key elements: the electrostatic phase plate that produces structured illumination, and the HAADF detector that measures a single-pixel signal carrying the information about the sample.

The results indicate that the Computational Ghost Imaging method has been successfully adapted for TEM, as it was showed that our experimental protocol is capable of nanoscale imaging, with an estimated resolution $r \gtrsim 1nm$. However, the analysis of the experimental setup showed that the method can probe information for frequencies up to the aperture-limited resolution. So, despite the fact that atomic resolution imaging is currently inaccessible because of a series of technical limitations, the results hint to the fact that such imaging condition can be reached once those issues will be solved. Any future work on this topic should then tackle the problems pointed out in this thesis.

In particular, it was pointed out that the phase plate technology employed to generate the structured probes needs to be refined to make the process easier to simulate. Also, there are other issues bound to the measurement of the HAADF signal, in terms of noise, acquisition time and periodic scan artifacts, that must be addressed to further improve the method. Some possible strategies were outlined after the discussion of the experimental results.

In conclusion, the findings briefly recollected above show how my efforts brought forth the world's first functioning Electron Computational Ghost Imaging experimental setup realized in a TEM, proving that the core of the method has legitimate foundations in physics. This work has also exposed how at the present state the resolution of ECGI can be estimated around $1nm$, with strong hints towards the feasibility of imaging at atomic resolution, overcoming the limitations of uncorrected microscopes. Keeping in mind that the technique is still facing all the challenges of early development, such results are quite encouraging and the hope is that once the current issues will be solved, ECGI can become a sought-after technique.

Acknowledgements

First and foremost, my sincere thanks go to the members of the Electron Optics and Microscopy group at CNR NANO-S3 and UniMoRe, particularly: my PhD supervisors, Prof. Stefano Frabboni and Dr. Vincenzo Grillo, that encouraged me and were always ready to provide guidance in times of need; Dr. Enzo Rotunno, for his constant support and his pragmatic, rational, and calm approach to problem-solving, that I admire and can only hope to match in the future; and Dr. Paolo Rosi, for being always available to clear up doubts about the collection of experimental data and the subtleties of Python coding.

Next, I must mention all the staff at the Ernst-Ruska Centrum for Microscopy and Spectroscopy with Electrons of the Forschungszentrum Jülich, that have welcomed me for over a year for my period abroad and without whom this project would not have been possible. Particularly, I have to thank:

Prof. Dr. Rafal Dunin-Borkowski, that has decided to keep investing in the long-standing collaboration with our group in Modena;

Dr. Amir Hossein Tavabi, with whom I shared countless hours, for his genuine interest in supporting my career, and for passing on to me some of his incredible mastery over many of the microscopes at the institute;

Marie Göcking and Alexander Müller, for their help with bureaucracy and digital services;

Dr. Thibaud Denneulin, René Borowski, Alexander Clausen, and Dieter Weber, for the time they dedicated to me whenever I needed some help with the software side of the microscopes;

and finally, Dr. Govind Ummethala and my dear friend Přemysl Páchl for their warmth and kindness. It has been delightful to get to know them, and spend time together both at work and outside.

I also have to remember all my friends, old and new, for the times shared together. As we all grow older and get busier with the many aspects of adult life, they chose to save some of their precious time for me, and I am ever grateful for it. They believe in me more than I do myself, and it is also because of their encouragement that I found the strength to overcome the hardships that emerged in this project. Particularly, I want to thank two groups. First, my long-standing group of friends from the bachelor years of University: Adamo, Alessia, Chiara, Edoardo, and Gaia. I wish them tenfold the happiness they brought to me over these years together. Second, the youth group of my hometown, particularly my "colleague" Luca, and all the young men and women for whom I have now been an educator for many years: it is a privilege and one of my greatest joys to see you grow up and find your way in the world, knowing that I might have helped you even just a little bit.

Finally, my biggest thank you goes to my family. Over many years of study they shared the joy for my successes, alleviated my concerns and offered me unconditional love and understanding. You're simply the best.

Innanzitutto, i miei sinceri ringraziamenti vanno ai membri del gruppo di Ottica Elettronica e Microscopia del CNR NANO-S3 e UniMoRe, in particolare: il miei supervisori di PhD, il prof. Stefano Frabboni e il Dr. Vincenzo Grillo, che mi hanno incoraggiato e sono sempre stati pronti a guidarmi quando ne ho avuto bisogno; il Dr. Enzo Rotunno, per il suo supporto costante e per il suo approccio pragmatico, calmo e razionale alla soluzione dei problemi, che io ammiro e che posso solo sperare di eguagliare in futuro; e il Dr. Paolo Rosi, per essere stato sempre disponibile a chiarire i dubbi sulla raccolta dei dati sperimentali e sulle sottigliezze della programmazione in Python.

Dopodiché, devo menzionare tutto lo staff del Centro Ernst-Ruska per la Microscopia e la Spettroscopia con Elettroni del Forschungszentrum Jülich, che mi hanno accolto per più di un anno per il mio periodo all'estero e senza i quali questo progetto non sarebbe stato possibile. In particolare devo ringraziare: il Prof. Dr. Rafal Dunin-Borkowski, che ha deciso di continuare a investire nella collaborazione di lungo corso con il nostro gruppo di Modena; il Dr. Amir Hossein Tavabi, con il quale ho trascorso innumerevoli ore, per il suo genuino interesse nel supportare la mia carriera e per avermi transesso un po' della sua incredibile padronanza di tanti dei microscopi dell'istituto; Marie Göcking e Alexander Müller, per il loro aiuto con la burocrazia e i servizi digitali; il Dr. Thibaud Denneulin, René Borowski, il Dr. Benjamin Zingsem, Alexander Clausen e Dieter Weber, per il tempo dedicatomi quando avevo bisogno d'aiuto con la parte software dei microscopi; e infine, il Dr. Govind Ummethala e il mio caro amico Přemysl Páchl, per il loro calore e la gentilezza. È stato splendido poterli conoscere e passare del tempo con loro sia al lavoro che fuori.

Devo anche menzionare tutti i miei amici, vecchi e nuovi, per i momenti condivisi insieme. Mentre tutti noi invecchiamo e siamo sempre più impegnati con i molti aspetti della vita adulta, loro hanno scelto di riservare un po' del loro prezioso tempo per me, e io gliene sono sempre grato. Credono in me più di quanto io creda in me stesso, ed è anche grazie al loro incoraggiamento che ho trovato la forza di superare le difficoltà emerse in questo progetto. In particolare, voglio ringraziare due gruppi. In primo luogo, il mio gruppo di amici di lunga data dagli anni della laurea triennale: Adamo, Alessia, Chiara, Edoardo e Gaia. Auguro loro di decuplicare la felicità che mi hanno regalato in questi anni insieme. In secondo luogo, il gruppo dei giovani della mio paese, in particolare il mio "collega" Luca, e tutti i ragazzi e le ragazze per i quali sono ormai educatore da molti anni: è un privilegio e una delle mie gioie più grandi vedervi crescere e trovare la vostra strada nel mondo, sapendo che forse vi ho aiutato anche solo un po'.

Infine, il mio più grande ringraziamento va alla mia famiglia. In tanti anni di studio hanno condiviso la gioia per i miei successi, hanno alleviato le mie preoccupazioni e mi hanno offerto amore e comprensione incondizionati. Siete semplicemente il meglio.

References

- [1] B. H. Bransden, and C. J. Joachain, *Quantum Mechanics*, Pearson, 2nd ed. (2000)
- [2] David B. Williams, C. Barry Carter, *Transmission electron microscopy - a textbook for materials science*, Springer, (2009)
- [3] A. Tonomura, *Electron Holography*, Springer Series in Optical Sciences: Vol. 70, Springer, (1999)
- [4] M. Knoll and E. Ruska, *Das Elektronenmikroskop*, Zeitschrift Für Phys. 78, 318 (1932).
- [5] E. Ruska, *The Development of the Electron Microscope and of Electron Microscopy*, Biosci. Rep. 7, 607 (1987).
- [6] H. Busch, *The Mode of Operation of the Concentrating Coil in a Braun Tube*, Arch. Fur Elektrotechnik 18, 583 (1927).
- [7] M. De Graef, *Introduction to Conventional Transmission Electron Microscopy*, Cambridge university press, (2003).
- [8] H. Kohl and L. Reimer, *Transmission Electron Microscopy*, Vol. 36, Springer New York, (2008).
- [9] Ray F. Egerton *Physical Principles of Electron Microscopy*, Springer, (2005)
- [10] A. Lipson, S.G. Lipson, H Lipson, *Optical Physics, 4th edition*, Cambridge University Press, (2011)
- [11] O. Scherzer, *Über Einige Fehler von Elektronenlinsen*, Z. Phys. 101, 593 (1936).
- [12] J. M. Zuo, and J. C. H. Spence, *Advanced transmission Electron Microscopy - Imaging and Diffraction in Nanoscience*, Springer (2017)
- [13] H. E. Keller, K. R. Spring, J. C. Long, and M. W. Davidson, *Optical Aberrations Interactive Tutorials - Astigmatism*, <https://micro.magnet.fsu.edu/primer/java/aberrations/astigmatism/index.html>
- [14] A. Lubk, and F. Röder, *Semiclassical TEM image formation in phase space*, Ultramicroscopy 51 (2015)
- [15] D. Typke, and K. Dierksen, *Determination of image aberrations in high-resolution electron microscopy using diffractogram and cross correlation methods* Optik 99, 155 (1995)
- [16] *JEOL Glossary of TEM Terms*, <https://www.jeol.com/words/emterms/>
- [17] K.Y. Bliokh, I.P. Ivanov, G. Guzzinati, L. Clark, R. Van Boxem, A. Béch  , R. Juchtmans, M.A. Alonso, P. Schattschneider, F. Nori, J. Verbeeck, *Theory and applications of free-electron vortex states*, Physics Reports 690 (2017) pp. 1–70
- [18] J. Thomas, T. Gemming, *Analytical Transmission Electron Microscopy*, Springer, (2014)

- [19] S. J. Pennycook, D. E. Jesson, *High-resolution Z-contrast imaging of crystals*, Ultramicroscopy 37, (1991)
- [20] J. Broeckx, M. Op de Beeck, D. Van Dyck, *A useful approximation of the exit wave function in coherent STEM*, Ultramicroscopy 60, (1995)
- [21] D. Van Dyck, M. Op de Beeck, *A simple intuitive theory for electron diffraction*, Ultramicroscopy 64, (1996)
- [22] P. Nellist, S. J. Pennycook, *The Principles and Interpretation of Annular Dark-Field Z-Contrast Imaging*, Advances in Imaging and Electron Physics, Elsevier, Volume 113, (2000)
- [23] P. Geuens, D. Van Dyck, *The S-State model: a work horse for HRTEM*, Ultramicroscopy 93, (2002)
- [24] G. R. Antis, D.Q. Cai, D. H. J. Cockayne, *Limitations on the s-state approach to the interpretation of sub-angstrom resolution electron microscope images and microanalysis*, Ultramicroscopy 94 (2003)
- [25] D. Gabor, *A new Microscopic Principle*, Nature 161 (1948)
- [26] D. Gabor, *Microscopy by reconstructed wave-fronts*, Proc. Royal Soc. A 197 (1949)
- [27] J. M. Cowley, *Twenty forms of electron holography*, Ultramicroscopy 41 (1992)
- [28] P. W. Hawkes, and J. C. H. Spence, *Springer Handbook of Microscopy*, Ch. 16-17 (2019)
- [29] J.M. Rodenburg, and H.M. Faulkner, *A phase retrieval algorithm for shifting illumination*, Appl. Phys. Lett. 85 (2004)
- [30] A.M. Maiden, and J.M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, Ultramicroscopy 109 (2009)
- [31] T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, *Optical imaging by means of two-photon quantum entanglement*, Physical Review A volume 52, number 5, (1995)
- [32] A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato *Correlated imaging, quantum and classical* Phys. Rev. A. 70 (2004)
- [33] A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato *Ghost Imaging with Thermal Light: Comparing Entanglement and Classical Correlation* Phys. Rev. Lett. 93 (2004)
- [34] A. Valencia, G. Scarcelli, M. D'Angelo, and Y. Shih *Two-Photon Imaging with Thermal Light* Phys. Rev. Lett. 94 (2005)
- [35] F. Ferri, D. Magatti, A. Gatti, M. Bache, E. Brambilla, and L. A. Lugiato *High-Resolution Ghost Image and Ghost Diffraction Experiments with Thermal Light* Phys. Rev. Lett. 94 (2005)
- [36] Jeffrey H. Shapiro, *Computational Ghost Imaging*, Phys. Rev. A 78 (2008)

- [37] V. Arrizón, G. Méndez, and D. Sánchez de La Llave, *Accurate encoding of arbitrary complex fields with amplitude-only liquid crystal spatial light modulators*, Opt. Express 13, (2005)
- [38] Miles Padgett, Reuben Aspden, Graham Gibson, Matthew Edgar, and Gabe Spalding, *Ghost Imaging*, Optics & Photonics News 27 (2016)
- [39] B. Sephton, I. Nape, C. Moodley, J. Francis, and A. Forbes, *Revealing the embedded phase in single pixel quantum ghost imaging*, Optica 10 (2023)
- [40] K. Komuro, Y. Yamazaki, and T. Nomura, *Transport-of-intensity computational ghost imaging*, Appl. Opt. 57 (2018)
- [41] G. M. Gibson, S. D. Johnson, and M. J. Padgett, *Single-pixel imaging 12 years on: a review*, Optics Express vol. 28, No.19 (2020)
- [42] B. I. Erkmen, and J. H. Shapiro, *Signal-to-noise ratio of Gaussian-state ghost imaging*, Phys. Rev. A 79 (2009)
- [43] E. J. Candes, and M. B. Wakin, *An introduction to compressive sampling*, IEEE Signal Process. Mag. 25, 21 (2008).
- [44] J. Romberg, *Imaging via Compressive Sampling*, IEEE Signal Process. Mag. 25, 14 (2008).
- [45] O. Katz, Y. Bromberg, and Y. Silberberg, *Compressive ghost imaging*, App. Phys. Lett. 95 (2009)
- [46] M. Lyu, W. Wang, Hao Wang, Haichao Wang, G. Li1, N. Chen, and G. Situ, *Deep-learning-based ghost imaging*, Sci. Rep. 7, 17865 (2017)
- [47] Y. He, G. Wang, G. Dong, S. Zhu, H. Chen, A. Zhang, and Z. Xu, *Ghost Imaging Based on Deep Learning*, Sci. Rep. 8, 6469 (2018)
- [48] T. Shimobaba, Y. Endo, T. Nishitsuji, T. Takahashi, Y. Nagahama, S. Hasegawa, M. Sano, R. Hirayama, T. Kakue, A. Shiraki, and T. Ito, *Computational ghost imaging using deep learning*, Opt. Comm. 413 (2018)
- [49] F. Wang, C. Wang, C. Deng, S. Han, and G. Situ, *Single-pixel imaging using physics enhanced deep learning*, Photonics Research, Vol 10, No. 1 (2022)
- [50] C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, *Deep learning for real-time single-pixel video*, Sci. Rep. 8, 2369 (2018)
- [51] X. Yu, R. Ivanov Stantchev, F. Yang, and E. Pickwell-MacPherson, *Super Sub-Nyquist Single-Pixel Imaging by Total Variation Ascending Ordering of the Hadamard Basis*, Sci. Rep. 10, 9338 (2020)
- [52] S. Georgiou, C. Koukouvinos, and J. Seberry, *Hadamard matrices, orthogonal designs and construction algorithms*, chapter in *DESIGNS 2002: Further computational and constructive design theory*, Kluwer (2003)
- [53] E. Toninelli, D. Stellinga, B. Sephton, A. Forbes, and M. J. Padgett, *Single-pixel imaging using caustic patterns*, Sci. Rep. 10, 2281 (2020)

- [54] S. Li, F. Cropp, K. Kabra, T. J. Lane, G. Wetzstein, P. Musumeci, and D. Ratner, *Electron Ghost Imaging*, Phys. Rev. Lett. 121 (2018)
- [55] P. Rosi, L. Viani, E. Rotunno, S. Frabboni, A. H. Tavabi, R. E. Dunin-Borkowski, A. Roncaglia, and V. Grillo, *Increasing the Resolution of Transmission Electron Microscopy by Computational Ghost Imaging*, Phys. Rev. Lett. 133 (2024)
- [56] W. Van den Broek, B. W. Reed, A. Béché, A. Velazco, J. Verbeeck, and C. T. Koch *Various Compressed Sensing Setups Evaluated Against Shannon Sampling Under Constraint of Constant Illumination*, IEEE Transactions on Computational Imaging, vol. 5, no. 3 (2019)
- [57] Karlsruhe Institute of Technology, IMT thesis proposal, *Generation of focused light patterns at fiber tips using Digital Micromirror Device (DMD)*, https://www.imt.kit.edu/open-student-projects_2098.php
- [58] G. Pozzi, *Advances in Imaging and Electron Physics - Particles and Waves in Electron Optics and Microscopy*, Academic Press (2016)
- [59] R. Danev, and K. Nagayama, *Chapter Fourteen - Phase Plates for Transmission Electron Microscopy*, Methods in Enzymology, Academic Press, Vol. 481 (2010)
- [60] M. Malac, S. Hettler, M. Hayashida, E. Kano, R. F. Egerton, and M. Beleggia *Phase plates in the transmission electron microscope: operating principles and applications*, Microscopy, Vol. 70 (2021)
- [61] R. Danev, and K. Nagayama, *Transmission electron microscopy with Zernike phase plate*, Ultramicroscopy 88 (2001)
- [62] C. Yu, F. Vega Ibañez, A. Béché, and J. Verbeeck *Quantum wavefront shaping with a 48-element programmable phase plate for electrons*, SciPost Phys. 15, 223 (2023)
- [63] A. H. Tavabi, P. Rosi, E. Rotunno, A. Roncaglia, L. Belsito, S. Frabboni, G. Pozzi, G. C. Gazzadi, P. Lu, R. Nijland, M. Ghosh, P. Tiemeijer, E. Karimi, R. E. Dunin-Borkowski, and V. Grillo, *Experimental Demonstration of an Electrostatic Orbital Angular Momentum Sorter for Electron Beams*, Phys. Rev. Lett. 126 (2021)
- [64] Y. Aharonov, and D. Bohm, *Significance of electromagnetic potentials in quantum theory*, Physical Review. 115 (1959)
- [65] C.J. Edgcombe, A. Ionescu, J.C. Loudon, A.M. Blackburn, H. Kurebayashi, and C.H.W. Barnes, *Characterisation of ferromagnetic rings for Zernike phase plates using the Aharonov–Bohm effect*, Ultramicroscopy, Vol. 120 (2012)
- [66] A. Béché, R. Van Boxem, G. Van Tendeloo, and J. Verbeeck, *Magnetic monopole field exposed by electrons*, Nature Physics volume 10 (2014)
- [67] A. H. Tavabi, M. Beleggia, V. Migunov, A. Savenko, O. Öktem, R. E. Dunin-Borkowski, and G. Pozzi, *Tunable Ampere phase plate for low dose imaging of biomolecular complexes*, Sci. Rep. vol. 8 (2018)

- [68] H. Müller, J. Jin, R. Danev, J. Spence, H. Padmore, and R. M. Glaeser, *Design of an electron microscope phase plate using a focused continuous-wave laser*, New J. Phys. 12 (2010)
- [69] A. Konečná, E. Rotunno, V. Grillo, F. J. García de Abajo, and G. M. Vanacore, *Single-Pixel Imaging in Space and Time with Optically Modulated Free Electrons*, ACS Photonics, Vol. 10 (2023)
- [70] B. Barwick, D. J. Flannigan, and A. H. Zewail *Photon-induced near-field electron microscopy*, Nature Vol. 462 (2009)
- [71] Amir Hossein Tavabi, Vadim Migunov, Christian Dwyer, Rafal E. Dunin-Borkowski, Giulio Pozzi, *Tunable caustic phenomena in electron wavefields*, Ultramicroscopy, Vol. 157 (2015)
- [72] G. Ruffato, E. Rotunno, L.M.C. Giberti, and V. Grillo, *Arbitrary Conformal Transformations of Wave Functions*, Phys. Rev. Applied 15 (2021)
- [73] M.V. Berry, *Waves and Thom's theorem*, Advances in Physics, Vol. 25 (1976)
- [74] G. Matteucci, G.F. Missiroli, M. Muccini, and G. Pozzi *Electron holography in the study of the electrostatic fields: the case of charged microtips*, Ultramicroscopy, Vol. 45 (1992)
- [75] Z. Zhang, X. Wang, G. Zheng, and J. Zhong, *Hadamard single-pixel imaging versus Fourier single-pixel imaging*, Optics Express Vol. 25, Issue 16, (2017)
- [76] L.D. Landau, E.M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, Butterworth-Heinemann, Burlington (1981)
- [77] G. Pozzi, P. Lu, A. H. Tavabi, M. Duchamp, and R. E. Dunin-Borkowski, *Generation of electron vortex beams using line charges via the electrostatic Aharonov-Bohm effect*, Ultramicroscopy, Vol. 181 (2017)
- [78] P. Rosi, *Experimental Orbital Angular Momentum manipulation and measurement in a Transmission Electron Microscope*, PhD thesis in Physics and Nano Sciences of the University of Modena and Reggio Emilia (2021)
- [79] L. Bian, J. Suo, Q. Dai, and F. Chen, *Experimental comparison of single-pixel imaging algorithms*, J. Opt. Soc. Am. A 35, 78-87 (2018)
- [80] A. Kallepalli, L. Viani, D. Stellinga, E. Rotunno, R. Bowman, G. M. Gibson, M. Sun, P. Rosi, S. Frabboni, R. Balboni, A. Migliori, V. Grillo, and M. J. Padgett, *Challenging Point Scanning across Electron Microscopy and Optical Imaging using Computational Imaging*, Intell. Comput. (2022)
- [81] C. Papa, *Constrained Kaczmarz extended algorithm for image reconstruction*, Linear Algebra and its Applications, Volume 429, Issues 8–9 (2008)
- [82] T. Strohmer, and R. Vershynin, *A Randomized Kaczmarz Algorithm with Exponential Convergence*, J. Fourier Anal. Appl. (2009)

- [83] M. Sun, C. Gu, and P. Tang, *On Randomized Sampling Kaczmarz Method with Application in Compressed Sensing*, Mathematical Problems in Engineering, Volume 2020, Issue 1
- [84] K. Guo, S. Jiang, and G. Zheng, *Multilayer fluorescence imaging on a single-pixel detector*, Biomed. Opt. Express 7 (2016)
- [85] J. R. Fienup, *Phase retrieval algorithms: a comparison*, Appl. Opt. 21 (1982)
- [86] G. Zheng, R. Horstmeyer, and C. Yang, *Wide-field, high-resolution Fourier ptychographic microscopy*, Nature Photon. 7 (2013)
- [87] R. W. Gerchberg, and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures*, Optik. 35 (1972)
- [88] J. W. Cooley, and J. W. Tukey, *An algorithm for the machine calculation of complex Fourier series*, Math. Comput. 19 (1965)
- [89] A. M. Maiden, D. Johnson, and P. Li, *Further improvements to the ptychographical iterative engine*, Optica 4 (2017)
- [90] E. J. Kirkland, *Advanced Computing in Electron Microscopy*, 2nd edition, Springer (2010)
- [91] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach*, J. Opt. Soc. Am. A Opt. Image Sci. Vis. (2012)
- [92] Z. Chen, Y. Jiang, Y. Shaol, M. E. Holtz, M. Odstrčil, M. Guizar-Sicairos, I. Hanke, S. Ganschow, D. G. Schlom, and D. A. Muller, *Electron ptychography achieves atomic-resolution limits set by lattice vibrations*, Science 372 (2021)
- [93] C. Lee, and D. A. Muller, *Enhancing Depth Resolution of Multislice Ptychography with Data-Driven Prior and Regularization*, Microscopy and Microanalysis, Volume 30 (2024)
- [94] J. Lee, M. Lee, Y. Park, C. Ophus, and Y. Yang, *Multislice Electron Tomography Using Four-Dimensional Scanning Transmission Electron Microscopy*, Phys. Rev. Applied 19 (2023)
- [95] G. Van Rossum, and F. L. Drake, *Python 3 Reference Manual.*, Scotts Valley, CA: CreateSpace (2009)
- [96] *COMSOL Multiphysics®*, www.comsol.com, COMSOL AB, Stockholm, Sweden
- [97] T. Kluyver et al, *Jupyter Notebooks – a publishing format for reproducible computational workflows*, Positioning and Power in Academic Publishing: Players, Agents and Agendas (2016)
- [98] N. W. Ashcroft, and N. D. Mermin, *Solid State Physics*, Harcourt: New York (1976)
- [99] V. Grillo, and E. Rotunno, *STEM_CELL: a software tool for electron microscopy. Part I: Simulations*, Ultramicroscopy (2012)

- [100] V. Grillo, and F. Rossi, *STEM_CELL: A software tool for electron microscopy. Part 2: Analysis of crystalline materials*, Ultramicroscopy (2012)
- [101] *16-channel 16-bit high-voltage output DAC with integrated internal reference*, <https://www.ti.com/product/DAC81416>
- [102] *PyVISA library, Version 1.13*, <https://pyvisa.readthedocs.io/en/latest/>
- [103] E. J. Kirkland, and M. G. Thomas, *A high efficiency annular dark field detector for STEM*, Ultramicroscopy 62 (1996)