UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Thesis for the Doctor of Philosophy degree

in Physics and Nano Sciences

Cycle XXXIII

Experimental Orbital Angular Momentum manipulation and measurement in a Transmission Electron Microscope

Tutor: Dr. Vincenzo Grillo

Co-Tutor: Prof. Stefano Frabboni

Candidate: Paolo Rosi

PhD School Coordinator: Prof. Stefano Frabboni

Academic Year 2020-2021

Index of Contents

Abstract

In optics structuring light or electron beams is the coherent control of the wavefunction aimed to impart any arbitrary amplitude and phase structure.

A smart choice of the wavefunction can be used to enhance or inhibit parts of the beam sample interaction or to optimize the measurement. For this reason, beam shaping is nowadays used in many fields such as information encoding, communication, material science, life science and fundamental physics. Furthermore, this idea can allow to overcome the limits of classical light and electron optics techniques and it is at the forefront of the research.

This thesis will focus on the more recently developed field of electron beam shaping with particular reference to the Orbital Angular Momentum (OAM) degree of freedom. The eigenstates of OAM are the electron vortex beams (EVBs) whose early experimental demonstration in a transmission electron microscope (TEM) was in 2010-2011. The striking property of EVBs is that they open new ways to study materials with a chiral structure such as magnetic materials, plasmonic materials and proteins.

Given the natural cylindrical symmetry of the microscope and of the electron-atom interaction at high energies, the OAM is a natural observable in the scattering event and the selection rules of the scattering can be typically directly measured projecting on OAM states. However, both the measurement and the flexible creation of arbitrary OAM states is complicated with the present electron optics. My thesis is therefore a contribution in the creation of innovative devices and ideas for the generation and the measurement of OAM states in an electron microscope. I will mainly focus on the OAM measuring device called “quantum sorter” or “OAM sorter”.

In the first chapter I will describe the instrumentation. First the functioning of a transmission electron microscope (TEM) and why it is among the best and most quantitative instruments to study fundamental material properties at the nanoscale, but also a very interesting optical bench for electrons. Then I will explain the working principle of SEM (Scanning Electron Microscope) and FIB (Focused Ion Beam), as these machines are the main instruments I used to create the new electron optics devices.

In the following chapter I will describe the theory of the electron microscopy techniques, mainly phase related ones, that I and the research team that I am part of used during the testing and characterization steps of our devices. Furthermore, I will provide the theoretical framework with some practical examples for electron beam shaping, both holographic and electrostatic, focusing mainly on vortex beams.

Lastly, I will show the original research results of my PhD, which comprise mainly of experimental works regarding both the generation and measurement of EVBs by means of synthetic holographic phase plates and novel electrostatic devices based on nanofabricated MEMS, and even the theoretical design study for a functioning OAM-sorter based on magnetic fields.

Outside of the main topic of this PhD thesis I delved in other topics, whose results will be shown at the end of the thesis.

Abstract Italiano

In ottica, possiamo strutturare fasci di luce o elettroni con un controllo coerente della funzione d'onda per impartire l'ampiezza e/o la fase che desideriamo.

Una scelta intelligente della funzione d'onda, infatti, può essere usata per stimolare o inibire alcuni fenomeni in un campione oppure ottimizzare la misura di una sua proprietà. Proprio per questo, i fasci strutturati vengono sempre più utilizzati in scienze dei materiali e biologiche. Mentre avevano già assunto importanza in ottica, ad esempio, nel campo della codifica delle informazioni, nella comunicazione.

In questa tesi mi focalizzerò sul campo di recente sviluppo dei fasci di elettroni strutturati all'interno di un microscopio elettronico a trasmissione (TEM), e in particolare sui fasci dotati di momento angolare orbitale (in inglese Orbital Angular Momentum (OAM)) detti fasci a vortice. I fasci a vortice di elettroni, in inglese electron vortex beams (EVBs), sono autofunzioni dell'OAM e sono stati artificialmente generati per la prima volta nel 2010. Una fondamentale caratteristica che ha rapidamente reso famosi gli EVB è che proprio grazie al loro OAM offrono la possibilità di studiare in maggior dettaglio materiali con struttura chirale come proteine, materiali plasmonici o magnetici.

Data la simmetria cilindrica sia del microscopio che delle interazioni tra elettrone del fascio e atomo, l'OAM è un osservabile sempre presente negli eventi di scattering, e le regole di selezione dello scattering possono esser misurate direttamente proiettando sui relativi autostati. Purtroppo, sia creare un EVB con OAM arbitrario che misurare l'OAM è abbastanza complicato con gli elementi ottici presenti nei TEM moderni. La mia tesi mostrerà il mio contributo nel lavoro di introdurre nuovi dispositivi, compatibili con i microscopi già esistenti, che permettono di generare o analizzare EVB con facilità. Il dispositivo su cui mi focalizzerò principalmente è quello che permette la misura dell'OAM ed è chiamato quantum sorter o partitore quantistico.

Nel primo capitolo della mia tesi descriverò la strumentazione: il TEM è il principale strumento che ho usato durante la fabbricazione dei dispositivi, il "dual-beam". Per il TEM spiegherò i principi di funzionamento e perché è uno dei migliori strumenti per studiare, soprattutto dal punto di vista quantitativo, le proprietà della materia su scala nanometrica. Mi soffermerò poi sul "dual-beam", microscopio dotato di due sorgenti, una di elettroni (microscopia SEM) e una di ioni (microscopia FIB) Il FIB è qui utilizzato nella nano-fabbricazione dei dispositivi.

Il secondo capitolo è dedicato alla teoria, sia delle tecniche di microscopia elettronica che son state utilizzate per la caratterizzazione dei dispositivi, sia dei metodi olografici o elettrostatici di controllo della funzione d'onda degli elettroni, in particolare nel campo degli EVB. Spiegherò poi in dettaglio anche quali sono le possibili applicazioni dei fasci a vortice.

Infine, mostrerò i risultati ottenuti durante il mio dottorato riportando gli articoli di cui sono coautore e il mio ruolo in essi. Tra essi la maggior parte sono lavori sperimentali che mostrano i nostri progressi nel campo della generazione e misura di EVB. Partendo inizialmente da piatti di fase olografici siamo giunti a dispositivi

elettrostatici costruiti con la tecnologia dei MEMS (Microelectromechanical systems). Vi è infine uno sguardo al futuro con un lavoro sulla teoria e il progetto di un partitore quantistico basato su campi magnetici.
Al di fuori dell'argomento principale della mia tesi ho potuto lavorare in ambiti diversi. Riporto quindi i risultati ottenuti in un capitolo a parte.

Introduction

The last decades of the Nineteenth Century and the first few ones of the Twentieth Century represent one of the most flourishing and important periods for modern physics, paving the way to the technological advances which have drastically changed the society. The five seminal papers from Albert Einstein from his nowadays known *annus mirabilis* (1905) [1–5] overturned a series of physical principles that were commonly accepted in the physics community. The two most known papers of that year [1,4] founded special relativity, he also studied the Brownian motion of small particles [2,5] and was one of the first to introduce the concept of wave-particle duality for light [3](in it he also introduced the theory of the photoelectric effect and that a photon energy is $E = h\nu$, h being Planck's constant and ν being the photon frequency). In fact, Einstein was one of the first supporters of Planck's concept of energy quantisation of light particles [6]. In 1907 he published another paper dwelling on the quantization of energy for lattice vibrations [7], which founded the quantum theory of solids. Another important contribution from Einstein in the field of optics came in 1916-1917 when he proposed and demonstrated that photons (the quanta of light) carried a linear momentum that was proportional to their wavelength: $p = \frac{h}{\lambda}$ [8]. De Broglie extended this relation to massive particles as well [9] (particles which have a mass) such as electrons or atoms, thus claiming that particles with a mass when in motion have a wave-like nature. The experimental demonstration of this hypothesis came a few years later (in 1923 De Broglie published his paper and in 1927 the experimental realisation was observed) in two independent experiments on electron diffraction. George Paget Thomson and Alex Reid obtained a diffraction pattern similar to those obtained from X-Rays by sending an electron beam through a thin metal film [10], a similar result was obtained by Davidson and Germer by looking at the diffraction pattern created by electrons that had been backscattered by the surface of a crystal of nickel [11]. In both experiments the electrons were generated and moved in a cathode ray tube. Almost in the same year early works were being carried out which would later lead to the realisation of a rudimental (for today standards at least) electron microscope [12,13]. In a transmission electron microscope (TEM) the wave-particle nature of free electrons naturally manifest itself in almost every experiment, making it one of the most suitable instruments to study quantum properties of both matter and electrons themselves. In fact, depending on the problem, either the particle picture of electrons (described in its simplest form by classical electrodynamics) or the wave picture (described by Schrödinger equation [14]) may be more suitable. In the wave picture the wavefunction describes how an electron wave behaves as a function of space and time. In particular, the phase of the wavefunction contains information on how the wave propagates and can reveal how a sample has perturbed the wave-motion. Thus, the ability to measure the phase of an electron's wavefunction became of paramount importance not long after the (transmission) electron microscope was invented. One method to obtain information on the phase of the wave was proposed by Dennis Gabor in 1948 [15], who at the time was working with electron microscopes and wanted to improve the microscope's resolution, and is nowadays the technique known as holography. Even though the technique was born in electron microscopy it quickly grew

and it became widely known in the light optics community (mostly in in the laser community) since it could not directly be performed in the electron microscope mainly due to limitations in the coherence of the source. The first realisation of an holography experiment in a TEM came few years after Gabor's proposal by Haine and Mulvey [16], where they compared Gabor's original proposal to a modified version (both versions can be defined as On-Axis holography, a concept that will be explained in Chapter 2). The results obtained in both cases where not completely satisfactory. An improvement came after the invention of the electron biprim by Möllenstedt and Düker [17,18] (a metallic wire kept at constant potential). The biprism can be used to obtain the interference pattern of two electron wavefronts (one perturbed by the specimen and one not), a technique that is now commonly known as Off-Axis electron holography (the basics of which are explained in Chapter 2 of this thesis). An early comprehensive review on the topic of electron interferometry and interference which describes the results obtained in the 30 years after Gabor's proposal has been written by Missiroli, Pozzi and Valdrè [19]. However, since techniques such as Off-Axis holography are established, nowadays the focus has shifted on finding ways to modify the phase of a beam to improve measurements. An object that can modify the phase of a beam is referred to as phase plate. Phase plates have been widely used in electron microscopy (and even in light optics) to enhance contrast, for example (i.e., phase contrast microscopy). The ability to freely modify the phase of a beam is known as beam shaping. This thesis reports my contribution to the field of electron beam shaping, in particular in the field of Electron Vortex Beams (EVBs). Electron Vortex Beams are beams with an helicoidal phase term (the wave function is $\propto e^{im\theta}$ where θ is the azimuthal angle and m an integer number) and thank to this they carry an Orbital Angular Momentum (OAM) and a magnetic momentum. These two properties make them ideal in studies on magnetic materials. Free space EVBs were first generated in a TEM between 2010 and 2011 and reported in the seminal papers of Uchida and Tonomura [20], Verbeeck *et al.* [21] and McMorran *et al.* [22]. Since then, the research field on EVBs has substantially grown and the necessity of measuring the OAM of an Electron Vortex Beam as arisen. Stemming from a previous realisation of a method to measure the OAM spectrum of a EVB by Grillo *et al.* [23] (the device that they developed is known as OAM sorter), in this thesis I report the results of our efforts to improve this method by employing novel electrostatic phase plates.

My contribution to the original research works reported from Chapter 5 and forward is mostly on the experimental part and naturally in the drafting phase of the scientific articles. In all the experiments I was responsible for the sample preparation and characterisation. For the sample fabrication and to amend to some of the flaws in the OAM sorter elements I used a dual beam instrument. In particular, I used fib milling to fabricate the phase masks used in the experiments with an OAM sorter and to 3-D shape the needles in the OAM sorter device itself, while I used FIB deposition to produce the hollow cylindrical structures that are reported in Chapter 8. The sample characterisation was carried out in the TEM using technique such as Low Angle Diffraction and Off-Axis Electron Holography (the theory of these techniques is reported in Chapter 2). For the main experiment with the OAM sorter, I was not the main TEM user, but I actively participated in it.

This thesis is structured to provide a basic understanding of the theory of the instrumentation (Chapter 1 and Chapter 2), of the techniques that I used to characterize the samples (Chapter 2) and of vortex beams, particularly Electron Vortex Beams. Chapter 1 is dedicated to the electron microscope, Chapter 2 in its first part I describe the main components of a Transmission Electron Microscope and the relevant techniques used in this thesis, while in its second part I outline what a dual-beam machine (SEM-FIB) dwelling on how both columns work. In Chapter 3, it's introduced the Orbital Angular Momentum (OAM) observable, the properties of Vortex Beams are shown, the two most common vortex beams modes are reported, and you will find an explanation on how Electron Vortex Beams (EVBs) can be generated or how one can measure their OAM.

Chapter 4 is an "hybrid" chapter which links the three previous theoretical chapters to the next four chapters reporting original research work. In fact, Chapter 4 provides an overview on the topic of beam shaping through Synthetic-Computer Generated Holograms (S-CGHs), starting from the theoretical aspects of S-CGHs, it then provides a guide on how to fabricate them and lastly in its third part it reports a series of examples that are related to EVBs.

Chapter 5 is the main chapter of this thesis. In it are reported the results on the experimental realisation of an electrostatic OAM sorter, which represents the bulk of my research work during my PhD. The realisation of the electrostatic OAM sorter is a major mile stone in the field of EVBs as it allows to measure the OAM spectrum with greater precision and increase in the signal-to-noise ratio with our previous realisation through phase S-CGHs [23]. Here, I fabricated the phase S-CGHs used to generate the EVBs, I 3-D shaped and cleaned from the imperfections the electrodes, acquired the phase images of the Sorter elements via Off-Axis holography and actively participated in the experiments used to "benchmark" the device.

In Chapter 6 a method to analyse and quantify detrimental effects on the OAM spectrum is proposed. This method features a Convolution Neural Network (CNN) to recognize detrimental effects such as defocus, misalignments in the system or improper excitation of the electrodes. In this study, I helped in the early phases of the initialisation of the CNN and I participated in the experimental session during which the experimental patterns were acquired.

A new approach for the realisation of an OAM sorter is proposed in Chapter 7. There we report the theoretical model and relative calculations that support the feasibility of an OAM sorter that uses magnetic elements. Here I helped with the theoretical calculations and in writing the article draft.

In Chapter 8 I report the findings on a side project parallel to the main thesis topic (EVBs and the OAM sorter); it is closely connected to it since it is a proof of concept with early experimental results on a new approach for the realisation of an electrostatic phase plate for beam shaping. This new approach uses bimetallic hollow cylinder to impart a phase shift to the electron beam. In this work I fabricated the sample through FIB Induced

Deposition and FIB milling, characterized them in the TEM using Low Angle Diffraction and lastly, I analysed the data and performed all the fits. This work is a proof of concept that will kick-off a series of more in-depth and systematic experiments for the realisation of a complete and functioning device.

Lastly, Chapter 9 contains the concluding remarks of this thesis and summarizes the results obtained during my PhD.

Chapter 1 - The Electron Microscope

The electron microscope has evolved and become much more complex from the first exemplar that Ruska built in 1933 and later commercialized in 1939. However, the working principles have not changed much since then. We can divide the types of electron microscopes that are used worldwide in two macro categories: transmission electron microscopes, commonly addressed as TEMs, and scanning electron microscopes (SEM). The main difference between the two is actually where the detector is positioned with respect to the electron source and the sample. In TEMs, as the name might suggest, the detector is positioned after the sample, so that it receives the electrons after they passed through the sample. On the other hand, in a SEM the detector (actually detectors) is used to collect either secondary electrons (SE) or backscattered electrons (BSE), this means that geometrically the detector is actually between the source and the sample. Similarly to electrons, ions are also charged particles and can be used to study samples, the FIB (Focused Ion Beam) is components wise very similar to an SEM. A transmission ion microscope is also realizable and small number of research groups around the world have started working on Helium based Transmission microscopes, but they are facing many challenges mainly due to the destructive nature of the interaction between accelerated ions and the sample and charging under illumination [24].

In the next sections I will first describe the basics of an electron microscope, with a description more inclined towards the TEM, and then in the next chapter I will describe some of the technical details behind the TEM and the SEM and FIB since they comprise the “dual-beam” machine.

1.1 Basics of an Electron Microscope

Historically, what led Ruska and his collaborators to construct the Electron Microscope (EM) is their interest to overcome the resolution of light microscopes. In fact, in light microscopes the resolution, defined as the minimum distance between two points on a specimen that can still be distinguished, is diffraction limited at it is given by the Rayleigh criterion:

$$d_{min} = 0.61\frac{\lambda}{n\,sin\,\theta} \tag{1}$$

where θ is the half acceptance angle of the lens, $n \sin\theta$ is know as the numerical aperture (NA) and λ is the wavelength. For an optical microscope, where $\lambda \in [450{,}750]nm$, if we consider NA=1 the resolution ranges from $274.5 \div 457.5\ nm$. Even for electrons, Eq. 1 can be used in most cases to estimate the actual resolution. The electron wavelength in vacuum can be calculated using the relation proposed by De Broglie:

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} \tag{2}$$

where h is the Planck's constant, p is the electron momentum, m is the electron relativistic mass and E is the electron kinetic energy. Inside EMs, electrons are accelerated to tens or hundreds of keV (due to energy conservation $E = e\mathrm{V}_a$, e is the absolute value of the electron charge and V_a is the accelerating voltage expressed in Volts), so the electrons velocity approaches the speed of light. After the proper relativistic correction Eq. 2 can be rewritten as:

$$\lambda = \frac{h}{\sqrt{2m_0 eV_a\left(1 + \frac{e}{2m_0 c^2}V_a\right)}} \tag{3}$$

So that by substituting the values of the physical constants we obtain

$$\lambda = \frac{12.2643}{\sqrt{\mathrm{V}_a(1 + 0.978476 \times 10^{-6}\mathrm{V}_a)}} \tag{4}$$

Where λ is given in $\mathrm{\AA}$ and $1\mathrm{\AA} = 0.1\mathrm{nm}$. Common values of accelerating potentials used in EMs and the corresponding wavelength are listed in *Table 1* below:

V_a	30 kV	60kV	100kV	200kV	300kV
λ [Å]	0.0698	0.0487	0.0370	0.0251	0.0197

Table 1: Values of the relativistically corrected wavelength at different accelerating potentials.

For example, if we now consider $200keV$ electrons moving in vacuum, so that $\lambda = 0.0251$ and $n = 1$, and a lens acceptance angle of $10mrad$, then $d_{min} = 1.53\mathrm{\AA}$. And even though a $2\mathrm{\AA}$ resolution was reached only 40 years after the first TEM was built, it is clear the advantage that an electron microscope has over an optical one.

In 1932, Ruska and Knoll built the first magnetic lens for electron, and this allowed them to build a year later a working Electron microscope [13,25,26]. As we all know, electrons as charged particles are subject to a force when passing through an electric field or a magnetic field or both. The total force is given by the sum of the Coulomb and Lorentz forces

$$\vec{F} = -e\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{5}$$

and this force can be used to focus electrons. In fact, in 1926 Busch demonstrated theoretically that axially symmetric electric and magnetic fields could act as a lens for charged particles. In the particular case of a magnetic field, he demonstrated that the focal length f can be calculated using the formula [12,27,28]

$$\frac{1}{f} = \frac{\eta^2}{4\mathrm{V}_{a-r}{}^2}\int_{z_0}^{z_N} B^2(z)dz \tag{6}$$

where $\eta = \frac{|e|}{\sqrt{2em_0}}$ and $\mathrm{V}_{a-r} = \mathrm{V}_a\left(1 + \frac{e}{2m_0c^2}\mathrm{V}_a\right)$ is the relativistic acceleration potential. As it can be observed from Eq. 6 a short focal length can be obtained by concentrating the magnetic field, this was done by enclosing a coil in an iron shield, apart from a small aperture known as pole piece. In this configuration the magnetic field is naturally rotationally and axially symmetric.

The incident electrons actually interact with the sample, what this means is that in reality the electron microscope is not an instrument limited to "imaging", but can be seen as an ideal experimental bench to study fundamental physical and chemical properties of organic and inorganic matter. The interaction between electrons and the sample is through the Coulomb forces and it results in various scattering effects. The main distinction between the scattering events is if the interaction involves electron energy loss or no. In case of inelastic scattering, apart from losing energy, and consequently a change in momentum, the electron beam also loses coherence (a concept which will be explained in greater detail in Section 1.2). Both losses are measurable and can tell information on the sample. While coherence gives us information on the state of neighbouring illuminated atoms, the energy loss is either element-specific or related to material properties. Low energy loss events (<50eV) are usually related to phonon or plasmon scattering events, while higher energy ones are element-specific and can be used to know the sample's elements concentration and hybridization. This technique is known as Electron Energy Loss Spectroscopy (EELS), and one of the advantages of performing it inside a TEM is that it can be done locally, even in regions few nanometers wide or at the atomic scale. Elastic scattering is coherent and the scattering angle changes if the electrons interact with the atomic electron and the nuclei or just the nuclei, low scattering angles in the first case, high scattering angles or even backscattering in the second. Direct transmission and reflection are also a possibility. Furthermore, a multitude of complementary signals are generated from the electron-sample interaction such as Auger electrons, secondary electron, characteristic and continuum X-Rays, photons in the spectral range of UV, visible and IR (commonly known as catholuminescence), vibration phenomena such as phonons and plasmons. [29]

So, after taking everything into account, in a TEM the electron beam intensity after passing through a thin sample can be written as

$$I_0 - I_B = I_t + I_e + I_i = I_{zero} + I_i = I_{tot} \tag{7}$$

where I_0, I_B, I_t, I_e and I_i are the incident beam, backscattered, transmitted, elastic and inelastic scattering intensities, respectively. I_{zero}, the zero-loss intensity, is defined as $I_{zero} = I_t + I_e$. Through EELS it is possible to measure both I_{zero} and I_i with respect to the total intensity I_{tot} and experimentally it has been found that

$$I_i(\beta) = I_{tot}(\beta)\left[1 - e^{-t/\lambda_i(\beta)}\right] \tag{8}$$

where β is the spectrometer collection angle and λ_i is the electron inelastic mean free path (MFP).

1.2 Electron Wave Function and Propagation

Inside electron microscopes it is possible to observe both the wave-like and the particle-like behaviour of electrons, diffraction patterns are a clear demonstration of the first, while electron-magnetic lenses demonstrate the second. The duality give rise to a series of properties that here we summarize.

First, an electron beam has associated to it a wavefunction, ψ. It is not possible to measure the wave function, but what can be measured is the wave intensity (that also describes the electrons' distribution)

$$I(\vec{r}) = |\psi(\vec{r})|^2 = \psi(\vec{r}) \cdot \psi^*(\vec{r}) \tag{9}$$

The wave function should always be normalized so that the probability of finding an electron is always 1, and

$$P = \int_{-\infty}^{+\infty} \psi(\vec{r}) \cdot \psi^*(\vec{r}) d^3\vec{r} = 1 \tag{10}$$

The intrinsic properties of the electron beam can be studied by measuring the expectation value of an appropriate operator:

$$\langle \hat{A} \rangle = \langle \psi^*(\vec{r}) \big| \hat{A} \big| \psi(\vec{r}) \rangle = \int_{-\infty}^{+\infty} \psi(\vec{r}) \hat{A} \psi^*(\vec{r}) d^3\vec{r} \tag{11}$$

so that, for example, the operators $\hat{A} = -ih\,{}^{\partial}/_{\partial x}$ and $\hat{A} = ih\,{}^{\partial}/_{\partial t}$ give the momentum and energy of the electrons.

For an electron freely moving in vacuum (free-space) its wave function must satisfy the Schrödinger equation [14]:

$$\left(-\frac{\hbar^2 \nabla^2}{2m} + i\hbar \frac{\partial}{\partial t} \right) \psi(x,t) = 0 \tag{12}$$

where $-\frac{\hbar^2\nabla^2}{2m} = \mathrm{H}$ is the energy operator and it is commonly referred as Hamiltonian. It is possible to notice from Eq.12 that the resulting wave-function for electrons is complex and it is linear, meaning that there are no higher order terms including products of ψ in the wave equation. So that if two waves, ψ_1 and ψ_2, are both solution of the wave equation, then $\psi' = \psi_1 + \psi_2$ is also a solution. This can be extended to any number of such waves, so we have that

$$\psi = \sum_i \psi_i \tag{13}$$

is still a solution. We refer to this property as superposition principle and this allows us to write the wavefunction as a sum of any solutions. Superposition gives rise to the concept of wave coherence, both spatial and temporal. Coherence, in general, is the measure of correlation of the phases of different point on

a wave, temporal coherence for the correlation along the propagation axis, while spatial for the correlation transverse to the optical axis. As previously said, this will be discussed more thoroughly later.

It is worth remembering that the complex wave function ψ can generically be written as the product between the amplitude, a position dependent phase and a time dependent phase, as such:

$$\psi = A\, exp\big(i2\pi\vec{k}\cdot\vec{r} + i\delta\big)\exp\,(-i\omega t) \tag{14}$$

where $\vec{k}$ is the wavevector, with $\left|\vec{k}\right| = \frac{1}{\lambda} = \sqrt{k_x^2 + k_y^2 + k_z^2}$, δ is the initial phase and ω is the frequency. If we now try to examine the result of superposition between two waves, ψ_1 and ψ_2, and for sake of simplicity we limit to the one-dimensional wave case, then the resulting wavefunction is

$$\psi'(x,t) = \psi_1(x,t) + \psi_2(x,t)$$

$$= A_1\, exp(i2\pi k_1 x + i\delta_1)\exp(-i\omega_1 t) + A_2\, exp(i2\pi k_2 x + i\delta_2)\exp(-i\omega_2 t) \tag{15}$$

The intensity of the superimposed wave is then

$$I(x,t) = {A_1}^2 + {A_2}^2 + 2A_1A_2 cos\,[2\pi(k_1 - k_2)x - (\omega_1 - \omega_2)t + \delta_1 - \delta_2] \tag{16}$$

The typical exposure time T of an experiment is usually in the order of seconds for each acquisition, this means that we actually observe the average intensity, so

$$I_{obs}(x,t) = {A_1}^2 + {A_2}^2 + 2A_1A_2\langle cos\,[2\pi(k_1 - k_2)x - (\omega_1 - \omega_2)t + \delta_1 - \delta_2]\rangle_T \tag{17}$$

here, the first two terms of, the intensity are the intensities of the originating waves, while the third is an interference term. Depending on the relation between ω_1 and ω_2 we have several possibilities:

First Case: if $\omega_1 \neq \omega_2$ and $T \gg 2\pi/|\omega_1 - \omega_2|$. In this case, the average value of the cosine is zero, so that the observed intensity is just

$$I_{obs}(x,t) = {A_1}^2 + {A_2}^2 \tag{18}$$

Second Case: $\omega_1 = \omega_2$ and both δ_1 and δ_2 are constant, so

$$I_{obs}(x,t) = {A_1}^2 + {A_2}^2 + 2A_1A_2 cos\,[2\pi(k_1 - k_2)x + \delta_1 - \delta_2] \tag{19}$$

and as it is possible to observe, the intensity only varies with x so it is only position dependent and time independent.

Third Case: $\omega_1 = \omega_2$ and both δ_1 and δ_2 vary randomly with time. In this case, in the cosine while the time dependent term cancels out, the term depending on the initial phase is random, so the average value of the cosine is zero even in this case so that

$$I_{obs}(x,t) = {A_1}^2 + {A_2}^2 \tag{20}$$

Fourth Case: $\omega_1 \neq \omega_2$ and $T \ll 2\pi/|\omega_1 - \omega_2|$. In this case, the intensity is also time-dependent, so the most generic case. However, if again both δ_1 and δ_2 vary randomly with time, then again, the average value of the cosine will be zero.

The First and third case, where the observed intensity is just the sum of the intensity of the individual waves, are denoted as incoherent superpositions, while the second is said coherent. These are the two extreme cases, while in practice the total wave is usually a partially coherent summation of waves.

It is then useful to define and measure the degree of partially coherence. Let us consider the general case of interference between two points, P_1 and P_2, on a propagating wavefront at a distance $\vec{r}$ one to the other. The two points can be assumed to be source points of two waves, that will arrive through two different paths at the detector. If the first wave arrives at time t_0 and the second at $t_0 + \tau$, then the averaged intensity collected by the detector over a typical exposure time is

$$\langle I \rangle = \langle |\psi_1(t_0) + \psi_2(t_0 + \tau)|^2 \rangle$$

$$= \langle |\psi_1(t_0)|^2 \rangle + \langle |\psi_2(t_0 + \tau)|^2 \rangle + 2\mathrm{Re}\{\langle \psi_1(t_0){\psi_2}^*(t_0 + \tau) \rangle\}$$

$$= I_1 + I_2 + 2\mathrm{Re}\{\Gamma_{12}(\tau)\} \tag{21}$$

Where

$$\Gamma_{12}(\tau) = \langle \psi_1(t_0){\psi_2}^*(t_0 + \tau) \rangle \tag{22}$$

is known as the correlation function of the two waves and it measures the degree of similarity between the two functions. It is possible to define the autocorrelation function $\Gamma_{11}(\tau) = \langle \psi_1(t_0){\psi_1}^*(t_0 + \tau) \rangle$, i.e., the function that defines the self-correlation, so that the averaged intensity is rewritten as

$$\langle I \rangle = I_1 + I_2 + 2\sqrt{I_1 I_2}\mathrm{Re}\{\gamma_{12}(\tau)\} \tag{23}$$

Where

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} \tag{24}$$

is the complex degree of coherence. By recording the interference pattern of the two waves it is possible to measure the degree of coherence from the maximum and minimum intensities recorded. In fact, if we define the visibility, or Michelson contrast, as

$$v = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{25}$$

then, if $I_1 = I_2$, $v = |\gamma_{12}(\tau)|$ and I_{max} and I_{min} are obtained when $\gamma_{12}(\tau) = |\gamma_{12}(\tau)|$ and $\gamma_{12}(\tau) = -|\gamma_{12}(\tau)|$, respectively. Furthermore, the normalization with the autocorrelation functions in $\gamma_{12}(\tau)$ ensures that its maximum will not exceed 1, and in particular we have that the fully coherent case corresponds to $|\gamma_{12}(\tau)| = 1$, while the fully incoherent to 0 [30].

Experimentally, with interferometry it is possible to measure the longitudinal and transverse components of $\gamma_{12}(\tau)$. The longitudinal coherence, also known as temporal coherence, can be measured with a Michelson interferometer, while the transverse coherence, or spatial coherence, can be measured by the Young's two slits experiment.

1.2.1 Wave packets

Before we can delve into the concept of coherence, it is essential that we go back to discuss about the electrons wavefunction. Up to now we considered our electrons to be monochromatic, with a single frequency and that they propagated with a sinusoidal wave, which is continuous and infinite. In reality, the wavefunction needs to be normalizable and the emitted electrons have a finite energy spread, as well as a finite angular distribution and a range of electron momenta ($\propto k$). However, since the sinusoidal wavefunction is still a solution of the wave equation, the wave function of free electrons in the non-relativistic case can be expressed as the superposition of sinusoidal waves

$$\psi(\vec{r}, t) = \int \psi(\vec{k}) exp\,(i2\pi\vec{k}\cdot\vec{r}) exp\left(-i\pi\frac{hk^2}{m}t\right) d^3\vec{k} \tag{26}$$

where $\psi(\vec{k})$ is the complex amplitude of the sinusoidal wave with wave vector $\vec{k}$, and ω was substituted with $\pi\frac{hk^2}{m}$ as one can derive from the Schrödinger equation (Eq.12) for an electron travelling in vacuum with zero potential. It is necessary to point out that, as previously mentioned, the electrons inside the Electron microscope have a relativistic behavior, however here the choice of the non-relativistic description was for sake of simplicity in the following calculations (since the relativistic-corrected case is far more complicated). The wave function in Eq.26 is called a wave packet and it can be normalized as follow

$$\begin{aligned}\int \psi(\vec{r}, 0)\psi^*(\vec{r}, 0) d^3\vec{r} &= \int\int \psi(\vec{k})\psi^*(\vec{k}')\, d^3\vec{k} d^3\vec{k}' \int \exp\left(i2\pi(\vec{k}-\vec{k}')\cdot\vec{r}\right) d^3\vec{r} \\ &= \int\int \psi(\vec{k})\psi^*(\vec{k}')\,\delta(\vec{k}-\vec{k}') d^3\vec{k} d^3\vec{k}' \\ &= \int \psi(\vec{k})\psi^*(\vec{k}) d^3\vec{k} = 1\end{aligned} \tag{27}$$

It should be pointed out that both $\psi(\vec{k})$ and $\psi(\vec{r}, t)$ are the same electron wave function, one in momentum or k space, while the other in the real space. Furthermore, inside the electron microscope, the electrons have

a small divergence angle of few tens of milliradians (mrad) and an energy spread of ~0.3 to ~2 eV. A useful approximation of the wavefunction in k space is given by

$$\psi(\vec{k}) = A(k_x, k_y)\frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(k_z - k_0)^2}{2\sigma_k^2}\right] \tag{28}$$

where $A(k_x, k_y) = 1/(\pi k_{max}^2)$ for $\sqrt{{k_x}^2 + {k_y}^2} \leq k_{max}$ and 0 otherwise. This model assumes that momentum is uniformly distributed within a disk along the x and y directions and has a Gaussian distribution along the z direction with mean value $k_0 = 1/\lambda$.

1.2.2. Temporal coherence

The electron wave packet has a finite length, and as a consequence it has also a limited temporal coherence $\gamma_{12}(\tau)$. Borrowing a formalism that is similar to the one used In light optics to define temporal coherence, if we now consider the case where the electron momenta spread along the x and y directions are very small (i.e. we are in the commonly known paraxial approximation), and the emitted electrons propagate along the z-axis and are far from each other in time, then the probability of having two or more electrons emitted within the coherence time is very small [29]. In this situation it is possible to approximate the wave function using the quasi-monochromatic wave model, where with respect to Eq.26 we recover an expression of the wave function similar to Eq.14:

$$\psi(z,t) = \psi_0\, exp(i2\pi kz) exp\,(-i\omega t) exp\,(-i\delta(t)) \tag{29}$$

here, the electrons initial phase term $\delta(t)$ is time dependent and is taken constant within a coherence time τ_0, for example $\delta(t) = \Delta_n$ for the time interval $n\tau_0 \leq t < (n+1)\tau_0$. However, from one coherent period to the other the value of Δ_n varies randomly. This is the so-called random phase approximation.

Now, assuming the amplitudes of the two waves are the same, using the wave function as defined in Eq. 29 to calculate $\gamma_{12}(\tau)$ we obtain

$$\begin{aligned}\gamma_{12}(\tau) &= \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} = \langle exp(-i\omega t)\, exp(i[\delta(t) - \delta(t+\tau)])\rangle \\ &= exp(-i\omega t) \lim_{T\to\infty} \frac{1}{T}\int_0^T exp\,(i[\delta(t) - \delta(t+\tau)])dt\end{aligned} \tag{30}$$

To evaluate the integral we need to distinguish between two case: the first being $\tau > \tau_0$, i.e., the time delay exceeds the coherence time, and the second $0 < \tau < \tau_0$. In the first case, the phase difference $\delta(t) - \delta(t+\tau)$ is given by the difference between two random terms, which results in another random number. The integral over random phases averages to zero, so that $\gamma_{12}(\tau) = 0$ and the interference in totally incoherent.

In the second case, for $0 < \tau < \tau_0$, we have that

$$\delta(t) - \delta(t+\tau) = \begin{cases} 0 \ for\ 0 < t < \tau_0 - \tau \\ \Delta_n - \Delta_{n+1}\ for\ \tau_0 - \tau < t < \tau_0 \end{cases} \tag{31}$$

This applies to every coherent period in the quasi-monochromatic electron wave. If we now sum up all the coherent periods, we can rewrite the last expression of Eq.31 as

$$\gamma_{12}(\tau) = \exp(-i\omega t) \lim_{N\to\infty} \frac{1}{N\tau_0} \left\{ \sum_{n=0}^{N} \int_0^{\tau_0-\tau} dt + \int_{\tau_0-\tau}^{\tau_0} \exp(i[\Delta_n - \Delta_{n+1}]) \right\} \tag{32}$$

here, the sum over the random phases integrals averages to zero so that

$$\gamma_{12}(\tau) = \exp(-i\omega t) \frac{1}{\tau_0} \int_0^{\tau_0-\tau} dt =$$

$$= \left(1 - \frac{\tau}{\tau_0}\right) exp(-i\omega t) \tag{33}$$

This result shows that the visibility of the interference fringes, i.e., the contrast, $|\gamma_{12}(\tau)|$ decreases linearly with the delay time and disappears beyond the coherence time. This means that in order to observe the interference between the two wave the path difference between them must be smaller than $L = v\tau_0$, v being the electrons' velocity.

In the case of two waves with slightly different wavelengths ($\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \Delta\lambda$) the longitudinal coherence is defined as the length over which the two waves become completely out of phase with each other. We define L_C as the coherence length and we derive the relation

$$2L_C = N\lambda = (N-1)(\lambda + \Delta\lambda) \tag{34}$$

where N is the number of period where the two waves become in-phase again. If N is sufficiently large we obtain

$$N \approx N - 1 = \frac{\lambda}{\Delta\lambda} \tag{35}$$

and

$$L_C = \frac{\lambda^2}{2\Delta\lambda} \tag{36}$$

For the electron wave packet it can be shown that this expression can be rewritten as

$$L_C = \frac{\lambda^2}{2\Delta\lambda} = \frac{1}{2} v\Delta t = \frac{\lambda E}{2\Delta E} \tag{37}$$

where Δt is the uncertainty time obtained from the uncertainty principle. Therefore, if we now consider the coherence time to be equal to the uncertainty time ($\tau_0 = \Delta t$), at L_C we expect 50% contrast in the interference. This last definition was first investigated Mollenstedt and Ducker [17].

1.2.3. Spatial Coherence

To define the spatial coherence length (or transverse coherence length) we consider two waves with the same wavelength originating from two source points. A schematic view of this is sketched in Figure 1, where the solid red line and the dotted green lines are the wave fronts originating from the first and second source, respectively. The spatial coherence length, L_T, is defined as the lateral distance along a wavefront over which there is a complete dephasing between the two waves. Being θ the angle sustained by the two sources with respect to the point where the intensity is measured, if it is small by looking at Figure 1 we can geometrically estimate that

$$2L_T \approx \frac{\lambda}{\theta} \tag{38}$$

and it possible to notice how the transverse coherence length is inversely proportional to θ.

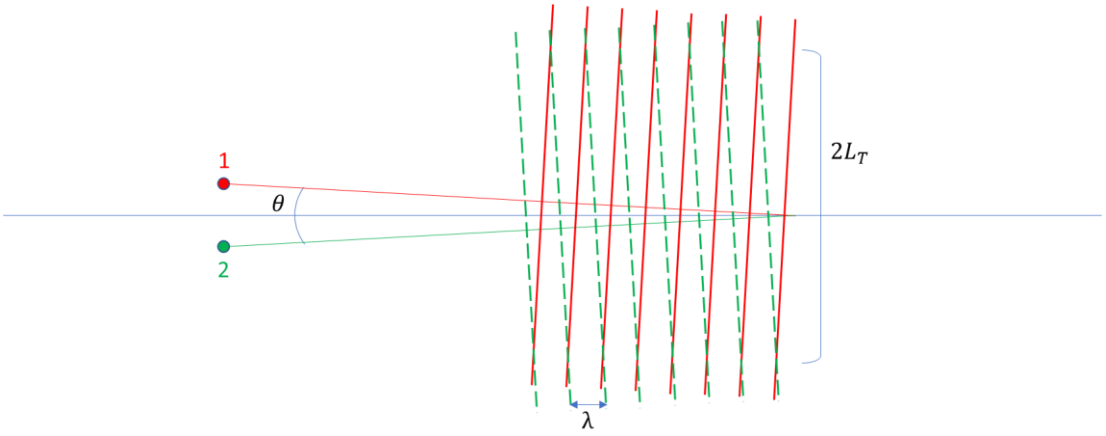

Figure 1: Spatial or transverse coherence. This illustration has been readapted from [29]

We can further generalize the above discussion by using the Young's two slit interference experiment as mentioned before. For sake of simplicity we use the two dimensional case, where the two narrow slits have the same width, are separated by a distance Δ and are placed symmetrically relatively to the source. The slits are illuminated by a one-dimensional source of finite width with intensity distribution $I(x)$. In every electron microscope the electrons are typically emitted from areas about the size of an atom, while the source dimensions are much bigger than an atom, so most electron sources can be seen as ideally incoherent, i.e., they consist of a statistically independent close-packed array of emitters. For such an incoherent extended source, where each atomic point-source generates an independent interference pattern at the detector and what is recorder is actually the sum of the intensities of these interference patterns. If we consider a point-source distant x from the optical axis, the two waves that originate at the slits can be written as

$$\psi_1(x,x',z,0) = A(x)\ exp(i2\pi kz)exp(-i2\pi k_x x')exp\ (-i\delta_1) \tag{39}$$

$$\psi_2(x,x',z,0) = A(x)\ exp(i2\pi kz)exp(i2\pi k_x x')exp\ (-i\delta_2) \tag{40}$$

where x is the transverse coordinate at the source, x' the transverse coordinate on the detector and δ_1 and δ_2 are the phases of the two waves at the slits, and their difference is determined by the path difference from the point-source to the slits, so that $\delta_1 - \delta_2 = \frac{2\pi}{\lambda} x \frac{\Delta}{Z_S}$ where Z_S is the distance along the optical axis of the source from the slits plane. If we only consider the interference intensity recorded at the center of the detector (so $x' = 0$), then the arrival time is the same, thus $\tau = 0$. According to Eq.23 the intensity contribution from the point-source at x is given by

$$\langle I(x)\rangle dx = 2I(x)[1 + \mathrm{Re}\{\gamma'_{12}(x)\}]dx \tag{41}$$

where in this case

$$\gamma'_{12}(x) = exp\left(i2\pi \frac{x\Delta}{\lambda Z_S}\right) \tag{42}$$

The overall intensity can be obtained by integrating over the width of the source

$$\gamma_{12}(\Delta, x' = 0) = \int_{-\infty}^{\infty} I(x) exp\left(i2\pi \frac{x\Delta}{\lambda Z_S}\right) dx \tag{43}$$

And extending this to the three dimensional case, with two dimensional source and slits we have

$$\gamma_{12}(\Delta_x, \Delta_y, x' = 0) = \frac{1}{I_0} \int_{-\infty}^{\infty} I(x,y) exp\left(i2\pi \frac{(x\Delta_x + y\Delta_y)}{\lambda Z_S}\right) dxdy \tag{44}$$

where $I_0 = \int_{-\infty}^{\infty} I(x,y)dxdy$ is the integrated source intensity. The above result is known as the Van Cittert-Zernike theorem in optics [30] and according to this theory the wavefront from a small incoherent source will appear mostly coherent at a large distance from the source ($Z_S \gg 0$). A possible explanation of this can be that since the electrons are confined within the source before their emission, once emitted their directions are determined by the source angle, sustained over the detector point. The range of electron momenta is proportional to the source angle and it is large when the detector is close to source. The recorded intensity by the detector is dominated by the signals coming from the closest source points and according to the uncertainty principle the source point can be determined with a high degree of accuracy. However, in the case the detector is placed far from the source, the momentum distribution is small, so it will be very difficult to pinpoint the contributions coming from specific source points. Thus, for a large source distance all source points appear to be the same and each contributes almost equally to the intensity.

If we now resume the one-dimensional source, and in this case it has a finite size D, the intensity at the source $I(x)$ is 1 for $|x| \leq D/2$ and 0 elsewhere, then

$$\gamma_{12}(\Delta, x' = 0) = \frac{1}{D}\int_{-D/2}^{D/2} exp\left(i2\pi\frac{x\Delta}{\lambda Z_S}\right)dx = \frac{sin\left(\frac{\pi D\Delta}{\lambda Z_S}\right)}{\left(\frac{\pi D\Delta}{\lambda Z_S}\right)} = sinc\left(\frac{\pi D\Delta}{\lambda Z_S}\right) \tag{45}$$

and the first zero of γ_{12}occurs when $\frac{D\Delta}{\lambda Z_S} = 1$ from which we can derive the relation

$$\Delta = \frac{\lambda Z_S}{D} \approx \frac{\lambda}{\theta} \tag{46}$$

given that $\frac{D}{Z_S} \approx \theta$. In the case $\Delta = L_T = \frac{\lambda}{2\theta}$ then

$$\gamma_{12}(\Delta = L_T, 0) = \frac{sin\left(\frac{\pi}{2}\right)}{\left(\frac{\pi}{2}\right)} = 0.64 \tag{47}$$

In summary, the transverse coherence length or coherence width is about $\frac{\lambda}{2\theta}$. The value of L_T can then be increased using a well-collimated beam, i.e., the source (ideally incoherent) subtends a small angle. This can be done by reducing the source size or by increasing the distance (here represented by Z_S). Naturally, by doing this the detected intensity will reduce, so a trade-off between spatial coherence and intensity is done when designing the source.

Chapter 2 - Instrumentation

In this chapter I will describe the most important technical details of the instrumentation that I used throughout my PhD. For ease of read and since this is not a dedicated manual, I will summarize only the most important technical aspects. I will first describe the TEM and then the dual beam machine.

2.1 The Transmission Electron Microscope (TEM)

The structure of a modern TEM is far more complex with respect to the one proposed and built by Ruska in the early 1930s. Figure 2 shows a schematic illustration of a modern S/TEM, where are sketched and pointed out the major components and attachments.

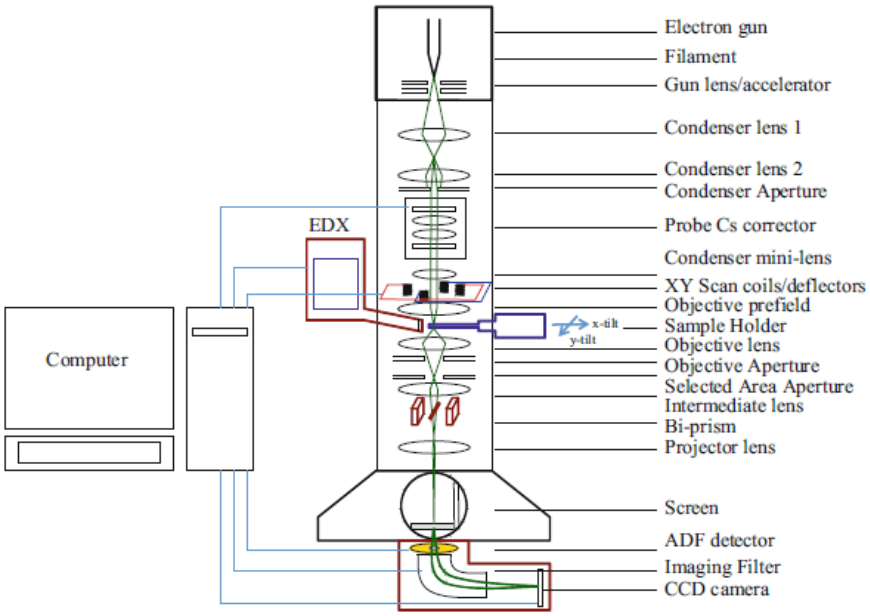

Figure 2: Schematic representation of a modern S/TEM where the main components such as lenses, apertures and attachments are shown. Except for the Probe Cs corrector, the S/TEM here represented has almost the same structure of the Talos X S/TEM present at the CIGS facility in Modena. This illustration has been taken from [29]

The microscope column can be divided in three optical systems: (i) the illumination system, (ii) the objective lens and (iii) the projection system. Each optical system and especially the function of the various lenses that comprise them will be explained in further details in the following subsections.

Apart from the lenses, there are some key components whose function is quite general and isn't specifically tuned and design for a particular section of the microscope. Electron beam deflectors, that are used throughout the TEM for optical alignment by steering the electron beam to the optical axis, are usually installed in strategic places. In most of the cases a beam deflector is a pair of saddle yoke magnetic coils or in their simplest form just a pair of electrostatic plates. These deflectors can both correct for or introduce a tilt or shift the beam. At the level of the electron gun this operation is referred as gun tilt/shift, in the illumination system right before the sample plane it called beam tilt/shift and in the projector system it either

image shift or diffraction shift (no tilt in this last case, however a tilt in the image plane is technically a shift in the diffraction, so diffraction shift is actually a tilt).

From Figure 2 it is also possible to see that below or above (only for the selected area plane) some of the most important lenses there the aperture planes. TEM aperture are metal diaphragms of different sizes that are mounted on linear translation stages (called aperture holders). It is possible to move the holder once inserted on both planar directions (y is the directed along the holder axis and x is perpendicular to it) with high precision even up to hundreds of micrometers without breaking the vacuum. The movement is nowadays controlled through the user interface, but it is not so difficult to find machines that are still being used where the movement control is manual by means of micrometric screws nearby the mounting point of the aperture holder. As one can easily imagine, apertures are used to limit the beam, i.e., select the electrons allowed to enter a lens by limiting their number as well as select their direction.

Add-ons, i.e., optical elements and detectors that can be added to the TEM column to expand the range of phenomena that can be studied in the TEM, such as aberration correctors (the aberrations in a TEM will be discussed later), monochromators, energy filters and spectrometers, detectors for energy-dispersive X-Ray (EDX) and the electron biprism have improved the capabilities of the instrument and in some case also allowed to study what was previously thought to be reserved for theoretical studies. The most prominent and important "add-on" is for sure the aberration corrector, which has allowed modern machines to reach a sub-Angstrom spatial resolution.

2.1.1 The Illumination System

The Illumination system can be called as the probe forming system and it is actually formed by the electron source and the condenser system. Since both parts are very important, I will separate their description in the two subsection that follow.

2.1.1.1 Electron Sources

Electron sources or emitters are in general metals or metallic solids, to avoid charging during the continuous emission. In fact, electrons are emitted by solids by overcoming the electron potential barrier by either photoemission, Thermionic emission, Schottky emission or Field emission. The potential barrier for electron emission is strictly related to the work function, i.e., the potential difference between the vacuum and the Fermi level (E_f) in absence of external fields. Conventional TEMs sources were first constructed using thermionic sources such as hairpin tungsten and lanthanum hexaboride (LaB_6) filaments, while nowadays they are Schottky or Field Effect emission sources. Photoemission is used only in TEMs used for time-resolved experiments, where the source is a photocathode illuminated by pulsed laser beams.

Regardless of the source used, the most important properties that need to be considered during the selection process of a new machine are the source brightness, size, energy spread and emission stability.

The brightness of a source is defined as the emission current density ($j = \frac{\Delta I}{\Delta S}$) per unit solid angle ($\Delta\Omega = \pi\alpha^2$):

$$\beta = \frac{\Delta I}{\Delta S \Delta\Omega} = \frac{j}{\pi\alpha^2} \tag{48}$$

and it is constant at all points along the optical axis, from the source to the detector, even when the lens aberrations are considered. The source brightness is a key factor that needs to be considered whenever a small probe is required. Since in most TEMs the beam convergence is fixed by an aperture during experiments, high brightness ensures a sufficient probe current density for electron imaging or diffraction pattern recording.

The source size is the effective object size seen by the first condenser lens. It is possible to improve the lateral coherence (spatial) of the illuminated condenser aperture by reducing the source size. This has also the effect of improving the quality of the illumination.

The source energy spread ΔE, as the name implies, defines the monochromaticity of the beam emitted by the electron source. It is determined by two factors: (i) the distribution of the emitted electrons' energy and (ii) the Coulomb interaction between the emitted electrons, that when measured along the beam is called Boersch effect [31]. The energy spread varies depending on the emission phenomenon used, for example for thermionic sources the measured energy spread ranges from 1.5 to 3 eV while for field effect sources it has been reduced 0.2-0.7 eV. As explained in section 1.2.2 the source energy spread affects the temporal coherence of the electron beam, moreover it also affects information transfer in high resolution S/TEM.

Lastly, the emission stability can be defined as the variance of intensity over a period of time. Any fluctuation in the electron source brightness is detrimental as it results in noise in the signals acquired by the detector, so for experiments that require extended acquisition times a very stable source is necessary.

Tables that compare the typical values of Brightness, source size, energy spread and stability between the various types of sources can be found in almost every manual on TEMs, such as [28,29]. For greater details on both properties and function it is worth to check the manuals just listed, however here I summarize the working principles and main advantages of each source:

- **Thermionic emission source**: This kind of source emits electron by heating the cathode/source. The most common cathode material are tungsten or LaB_6. The tungsten cathode is a wire 0.1-0.2 mm in diameter bent like an hairpin and the ends are soldered on electrical contacts. The wire is directly heated by passing a current through it. On the other hand, LaB_6 sources consist of small, pointed crystals. In this case the heating is indirect since their resistance is too high. With respect to

tungsten these cathodes require a lower vacuum to reduce the damage induced by the positive ions bombardment coming from the heat source. Furthermore, they also provide a higher gun brightness and a smaller energy spread. Regardless of the source material, the thermionic gun is actually made of three electrodes: the cathode, that we just discussed; a Wehnelt electrode that acts as a suppressor and allows to keep the emitted current constant and to limit the angles at which the electrons are emitted; the third element is the grounded anode. The shape of the Wehnelt electrode can greatly influence the brightness and other gun parameters.

- **Schottky emission source**: The emission source is a tip of tungsten covered by a layer of Zirconium Oxide (ZrO). The tip radius is $\approx 0.1 - 1\mu\mathrm{m}$ and it is obtained by first etching a tungsten rod, which is then coated with a ZrH_2 powder and heated in a Oxygen rich vacuum chamber for few hours. The oxide layer lowers the working function of the tip, making easier the emission of electrons. Even in this case there is a Wehnelt suppressor electron, but the tip apex protrudes of few tenths of a millimeter. In this case the electrons are emitted by heating the tip and by placing an extractor electrode which is at 4-8kV. The potential of the extractor is not strong enough for field emission, so Schottky sources can be seen as hybrid sources where the most dominant effect is still the thermionic one. However, the Schottky emission gun has a smaller energy spread with respect to thermionic ones because it is not widened by the Boersch effect, and the current density is almost two orders of magnitude higher. In essence, a Schottky source has a higher brightness, a smaller energy spread and a smaller source size with respect to a thermionic source, which explains why they subsided the thermionic ones.
- **Field Effect emission source**: Also known as field emission gun (FEG), it consists of a pointed cathode tip and two or more anodes. Even in this case the material of choice is tungsten since it is easily shaped as needed, but the emission is quite sensitive to the state of surface layers. In fact, during operation the work function changes due to diffusion of impurities, to surface reactions or to the adsorption of gasses. Furthermore, even in this case the ion bombardement damages the tip, and to reduce the damage as much as possible the source should be maintained in ultrahigh vacuum. The source is made starting from wires of 0.1mm in diameter that are spot-welded on a tungsten hairpin which is then electrolytically etched to form a tip of radius 0.1 μm. It is possible to anneal the tip to eliminate adsorbed gas atoms. The simplest form of a FEG features a three electrodes system: one cathode and two anodes. The cathode is the tip, while the first anode acts as the extractor and the second one further accelerates the electrons to the required energy. In this case the emission is not achieved by heating the source to high temperatures (>1800K), but owing to the field strength of the extractor the electrons are emitted by quantum-mechanical tunneling effect. Nowadays standard FEG sources work at temperatures in the range of 1500K, but in some cases with particular source configuration the cathode emits at room temperature and

such sources are known as cold-FEG. Field emission sources have the highest brightness and the lowest energy spread with respect to the other types of sources. They have a high coherence and are source of choice of modern S/TEM, particularly when the machines are built for high resolution or techniques that require a very coherent beam (for example Electron Holography).

Regardless of the source type the emitted electrons paths will form a crossover at a certain distance from the source, and the size of the crossover represents the (virtual) source size seen by the condenser system. For thermionic sources the crossover is after the Wehnelt anode, while for field effect ones it is after the second anode.

2.1.1.2 The Condenser System

The condenser system is equipped with at least two condenser lenses (three in the case of 300keV TEMs) and it is designed to be able to

- Focus the electron beam on the specimen to have sufficient image intensity even at high magnification.
- Keep the illumination uniform in every condition (so that the current density is also uniform). This allows to reduce specimen drift due to heating and to limit radiation damage and contamination in non-irradiated areas. For TEM techniques the electron beam should be kept as parallel as possible to maintain the required coherence.
- Vary the illumination aperture or convergence angle to allow for both convergent beam-based techniques and parallel beam-based techniques.
- Produce a small electron probe (0.2-100nm in diameter) for highly local spectroscopy techniques (EELS and EDX) but also to be able to function as a Scanning Transmission Electron Microscope (STEM- but this usually also require an $x - y$ scan coil/deflectors to be mounted above the objective lens. Modern TEM can function also as a STEM).

The first condenser lens, or C1, images the source crossover onto the object plane of the second condenser lens, or C2. It acts as a strongly demagnifying lens (meaning that is reduces the beam size seen by the other lenses). In a two condenser lenses illumination system the electron beam is controlled by varying the C2 focal distance together with the optimal condenser aperture. For TEM techniques illumination, the C2 lens transfers the image of the virtual source to the front focal plane of the objective pre-field lens which is above the sample and whose main job is to either make the illumination parallel for TEM or further reduce the probe size for STEM. In fact, for STEM the C2 lens and the objective pre-field lens work together with the C1 lens to provide both a very demagnified probe and a large convergence angle at the sample plane.

However, the two condenser setup has a number of drawbacks: first it has little flexibility in controlling the beam for TEM and second it is impossible to have a fixed object position after the C2 (this is desirable to improve the instrument usability and stability).

The solution for the first issue is the use of an auxiliary lens, known as condenser mini-lens (CM lens), which is not as versatile as the other two. Its main jobs are the following: for TEM illumination to image the crossover formed by the last condenser lens and to image it onto the front focal plane of the objective pre-field lens which then forms a parallel beam on the specimen; on the other hand, for STEM where a focused probe is necessary it works with the objective pre-field lens so that the crossover is imaged onto the specimen. Furthermore, the CM lens allows to change the size of the beam for TEM illumination or the convergence angle in STEM illumination with higher flexibility, since when it is not present the only way to change either of the two parameters is by changing the aperture size of the condenser aperture, and the number of available apertures is usually limited to a maximum of 4.

The solution to the second problem, which is having a fixed object position after the last condenser lens, and partially also to the problem of beam control flexibility is to add a third condenser lens. In fact, when the C1 lens is changed to change the spot size (smaller spot size means higher coherence, but less intensity) the crossover formed by C1 in front of C2 changes position, so C2 has to be readapted and the image position after C2 can change or the convergence changes. By introducing a third condenser lens it is possible to add further degrees of freedom in the beam control (high control in beam size and convergence angle), moreover it allows to fix the object position.

2.1.2 The Objective Lens

The Objective lens can be regarded as the most important lens of the microscope. Along the optical path it is positioned right after the specimen and it is designed with a short focal distance and a large magnification. This is actually the very first lens that contributes to the image formation and it is from this lens that the highest performance will be demanded. The natural astigmatism of the objective lens has to be so small that the objective stigmator (every important lens of the microscope has a stigmator paired to it to immediately correct for the astigmatism introduced by the lens itself) should only compensate for the astigmatism caused by contamination of the objective aperture and other perturbing effects.

Nowadays, in the most commonly used objective lens configuration the sample is inserted in the middle of the gap of a pair of symmetrical pole pieces. This configuration of sample and lens is commonly known as immersion lens, moreover instead of being called objective lens it is called condenser-objective lens. Here the field on the illuminating side of the specimen acts as the prefield objective lens (which will not be needed) and as already explained earlier it works with the condenser system to adjust the illumination on the sample (thus making the condenser-objective lens both a probe forming lens for the sample and the first image

forming lens). The remaining field is used for image formation, and it designed so that once the specimen is at the exact focus (in object position) then the image is produced at infinity (so we have infinite magnification). The sample can be brought into focus by either moving it along the z-axis or by changing the lens strength. The simplest and most effective model that is used to describe the condenser-objective lens is a particular case of the widely known in the EM community Glaser's bell-shaped magnetic field model. Glaser's bell-shaped magnetic field model is used to describe thick lenses such as all the real electromagnetic lenses found in the TEM, and can straightforwardly calculated the position of cardinal points and planes of thick lenses. Furthermore, it describes the process of image formation for a lens in the ideal thick lens case, but fails to describe real lenses for which numerical calculations are more appropriates. The model is well described in almost all manuals and handbooks on TEM, so I will not dwell too far in its description and only show the most important parts to get a basic understanding of it. In the bell-shaped model the magnetic field inside the lens along the z-axis (the propagation direction) has the shape:

$$B_z = \frac{B_{max}}{1 + (z/a)^2} \tag{49}$$

where B_{max} is the maximum field in the magnetic lens center and a is the half width at half maximum of the magnetic field z-axis distribution. Since all magnetic lenses are almost cylindrically symmetric the most appropriate system to use in the next few calculations is the cylindrical one (ρ, θ, z). It can be demonstrated that the radial component of the force acting on the electrons that pass through the field reported in Eq.49 is [28]:

$$m_{rel}\frac{\mathrm{d}^2\rho}{\mathrm{d}t^2} = -\frac{e^2}{4m_{rel}}\rho B_z^2 \tag{50}$$

where $m_{rel} = m_0(1 + E/E_0)$ is the relativistically corrected electron mass. For paraxial rays (i.e., rays that pass at small values of ρ from the optical axis of the lens) it is possible to assume that the electrons won't accelerate along the z direction so that $v_z = const = v = \frac{\mathrm{d}z}{\mathrm{d}t} = c\sqrt{1 - \frac{1}{(1+E/E_0)^2}}$, so that Eq.50 can be rewritten as

$$m_{rel}\frac{\mathrm{d}^2\rho}{\mathrm{d}t^2} = m_{rel}\frac{\mathrm{d}^2\rho}{\mathrm{d}z^2}\frac{\mathrm{d}^2z}{\mathrm{d}t^2} = m_{rel}v^2\frac{\mathrm{d}^2\rho}{\mathrm{d}z^2} = -\frac{e^2}{4m_{rel}}\rho B_z^2 \tag{51}$$

which can be rewritten as

$$\frac{\mathrm{d}^2\rho}{\mathrm{d}z^2} = -\frac{e}{8m_0V_{a-r}}\rho B_z^2 \tag{52}$$

where, as introduced at the start of Chapter 1 V_{a-r} is the relativistically corrected accelerating potential and $V_{a-r} = V_a\left(1 + \frac{E}{2E_0}\right)$. We can find the trajectory $\rho(z)$ in the meridional plane, i.e., the solution to Eq. 52, by

first introducing the reduced coordinates $b = \rho/a$ and $s = z/a$ and a dimensionless parameter k^2 defined as

$$k^2 = \frac{eB_{max}^2 a^2}{8m_0 V_{a-r}} \tag{53}$$

So that is it possible to rewrite Eq. 52 as

$$\frac{\mathrm{d}^2 b}{\mathrm{d}s^2} = -\frac{k^2}{(1+s^2)^2} b \tag{54}$$

which can be further simplified by considering the relations

$$s = cot\phi \;\; ; \;\; \mathrm{d}s = -\frac{\mathrm{d}\phi}{sin^2\phi} \;\; ; \;\; 1 + s^2 = cosec^2\phi \tag{55}$$

where ϕ varies from π at $z = -\infty$ to $\frac{\pi}{2}$ at $z = 0$ and then $\phi = 0$ for $z = +\infty$. So that Eq.54 can be rewritten as

$$\frac{\mathrm{d}^2 b(\phi)}{\mathrm{d}s^2} + 2cot\phi \frac{\mathrm{d}b(\phi)}{\mathrm{d}s} + k^2 b(\phi) = 0 \tag{56}$$

which in the case of a parallel incident ray at the initial condition $\rho = \rho_0$ for $z = -\infty$ (or $b(\phi = \pi) = \rho_0/a$ and $\frac{\mathrm{d}b(\phi)}{\mathrm{d}s} = 0$) we obtain that:

$$b = \frac{\rho}{a} = -\frac{\rho_0}{a\omega} \frac{\sin(\omega\phi)}{\sin(\phi)} \quad \text{with} \quad \omega = \sqrt{1 + k^2} \tag{57}$$

where ω is known the lens strength parameter. Figure 3a shows how the electrons trajectories vary as a function of ω. Glaser himself that noticed that a lens with a bell-shaped magnetic field with $k^2 \geq 3$ could work well as a condenser-objective lens. Figure 3b shows the condenser-objective lens in cross-section and the electron beams path through it, for lens excitation $k^2 = 3$. Here the front focal plane (FFP) and back focal plane (BFP) are conjugate, for example if a beam arrives parallel in the FFP then it is focused on the sample and again parallel in the BFP. The lens is thus operating in the “telefocal condition”.

For $k^2 \geq 3$ the lens is highly excited and in this condition it is possible to achieve a very short focal length and high magnification. This is important because since angular magnification is inversely proportional to the lateral magnification [29], i.e., the highest angle that the rays form with respect to the optical axis after passing through the objective aperture is $\alpha = \alpha_0/M$, where α_0 is the objective aperture (previously referred as angular magnification) and M is the magnification (or lateral magnification). Here, α_0 is given in first approximation by r/f, where r is the radius of the aperture and f is lens focus length, while M is usually between 20 and 50 times [28]. This ensures that outgoing rays travel at very small angles α to the optical axis in all the lenses that are after it. Moreover, since most of the principal aberrations (which will be described in paragraph 2.1.4) are proportional to α^n, where n varies on the kind of aberration, then they can be neglected for the subsequent lenses, even though they normally have higher values. The condenser-objective

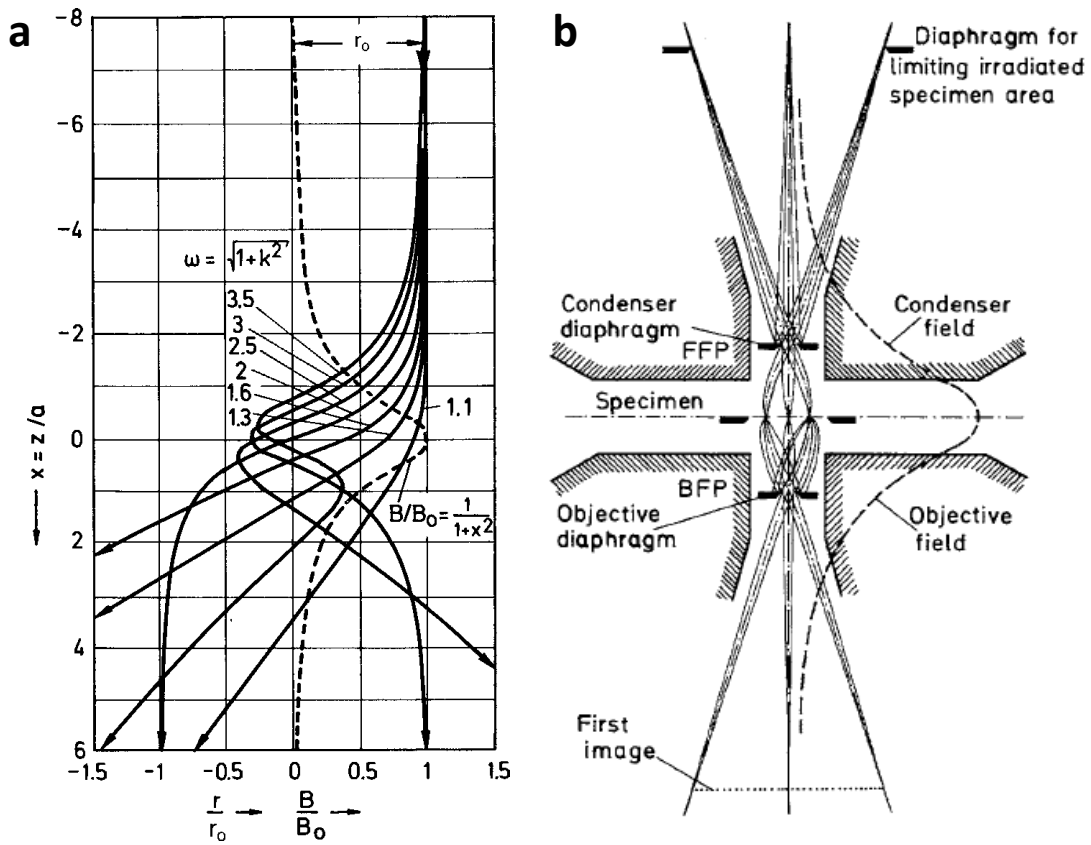

Figure 3: a) Electron trajectories for rays incident parallel to the optical axis at different values of lens strength ω. b) Cross-section of the condenser-objective lens, here are also shown the electron trajectories and field shape. Figures taken from [28].

lens configuration has been widely used, and the lenses used for it have become increasingly symmetric since this allows for convenient switching between TEM and STEM on the same specimen region. In this case, aberrations such as the spherical and chromatic, which will be described later, remain relatively constant when $a \approx f$, f being the focal length.

2.1.3 The projection system

In modern TEMs, the projection system has at least three intermediate lenses and at least one projector lens placed above the electron detector. The first intermediate lens is also called diffraction lens and it enables to switch from imaging mode and diffraction mode. The final image recorded on the electron detector is highly magnified from passing through all the lens and the total magnification is given by the product of the magnifications of the individual lenses:

$$M_{tot} = M_{obj} \times \prod_i M_{int\_i} \times \prod_l M_{proj\_l} \tag{58}$$

When working at medium or high magnification, the objective lens is operated at nearly constant strength and both its image position and object position are fixed. The last projector lens has also his image position fixed at the detector's position. What all this means is that the change in the total magnification (that needs to be able to vary from few thousand times to several millions during operation) is provided mainly by the intermediate lenses and the first projectors. On the other hand, for magnifications in the order of few

hundreds to few thousands the objective lens is switched off and the lens called “objective minilens”, placed after the objective lens and that is usually switched off, is used for the first stage magnification. Another possibility, again in the low and low-medium magnification range, it to use the special lens called Lorentz lens to provide the initial magnification when the objective lens is switched off. The name of this lens arises from the fact that it was added to the microscope to study magnetic samples and other materials that are sensitive to the strong magnetic field of the objective lens.

For diffraction an important parameter is the camera length, i.e., the effective distance between sample and detector, that allows to quantify the information contained in the diffraction pattern. The camera length can be calculated as

$$CL = f_{obj} \times \prod_i M_{int\_i} \times \prod_l M_{proj\_l} \tag{59}$$

2.1.4 Lens Aberrations

Almost every lens, both in light optics and electron optics, is aberrated. We define as lens aberration any effect introduced by the lens that causes the beam path to differ from the point-to-point imaging described by the paraxial equation. Aberrations are commonly categorized into two types: chromatic and geometric aberrations. Chromatic aberrations arise from the wavelength dispersion, electrons with different velocity (energy) will have different paths inside the magnetic lens, even when they arrive with the same initial trajectory. Naturally, this kind of aberration doesn’t affect the beam when the electrons are monochromatic. On the other hand, geometric aberrations are always present and arise from the defects in the lens, from the use of magnetic fields with low rotational symmetry and even in an ideal magnetic lens they are caused from the higher order terms beyond the first-order included in the paraxial equation.

For rotational symmetric lenses up to the third order, the prominent geometric aberration can be grouped into five isotropic and three anisotropic aberrations. The order of the aberration is calculated with respect to the product of α (the angle subtended by the rays with the optical axis) and r (the distance in the object plane of the point-source from the optical axis). The isotropic aberrations are:

- Spherical aberration
- Astigmatism
- Field curvature
- Distortion
- Coma

while the anisotropic ones are:

- Anisotropic coma

- Anisotropic astigmatism
- Anisotropic distortion

Of these, the most relevant ones are the spherical aberration, coma, astigmatism, distortion and curvature of image field. These are known as the Seidel aberrations in optics [28,29,32]. It is useful to point out that for high resolution experiments distortion and curvature of image field can normally be neglected. Furthermore, it must be noted that in some case the blurring introduced by the so-called diffraction effect can be erroneously mistaken for an aberration, however it arises from the presence of an aperture nearby the lens.

The spherical aberration has the effect of reducing the focal length for electron rays passing through the outer zones of the lens. This is because the magnetic field of the lens is not completely constant on the plane transverse to the optical axis, but its strength increase going away from the axis. This means that the rays that aren't along the optical axis will feel a stronger field and so a stronger focusing effect that brings them to a premature focus beyond which then broaden and form a disc at the image plane, instead of a spot. It has been demonstrated that the radius of the disc formed in the image plane by electrons that cross at the object plane the optic axis at angle α is

$$r' = C_S \alpha^3 M \tag{60}$$

where C_S is the spherical aberration coefficient (in mm) and M the lens magnification. It is clear that the radius rapidly increases as a function of α, so a small aperture is positioned prior to the lens to limit it, however due to diffraction effects the aperture can't be too small. Since C_S is always positive for a magnetic lens and it is large even at the condition for optimal lens design and excitation, the spherical aberration has been the most important factor that limited improvements in the S/TEMs spatial resolution. In the 90s spherical aberration correctors where successfully developed [33–35], bringing a renewed interest in transmission electron microscopy and the 80pm resolution that modern microscopes are able to achieve would without them.

For the objective lens the typical values of C_S are in the range of 0.5-2 mm and they depend on the lens excitation k^2. C_S decreases with increasing lens strength and the minimum focal length is found for $k^2 = 3$.

Coma is actually divided in two kind, off-axis and in-axis. Off-axis coma makes the rays emanating from an off-axis point in the object plane to show as a circle and at a greater distance from the axis in the image plane. The radius of one of such circles is proportional to $r\alpha^2$. Axial coma on the other hand is actually not a Seidel aberration and it arises from the incomplete axial symmetry of the lens, so due to imperfections in the lens. It is proportional to α^2, antisymmetric and introduces a complex wave-front distortion. In both cases of coma, the image of the beam becomes cone-shaped or comet-shaped, from which the name originates. The vertex angle of the cone for off-axis coma is 60°.

For astigmatism there is not just one kind, but every perturbation order has one apart for the first order. In general astigmatism gives a rotation-angle-dependent change in focus with n-fold symmetry, where n is the order. For the simplest case, twofold astigmatism, this results in having the object seemingly in focus along one direction and out-of-focus along the orthogonal one. This means that a circle in the object plane will become an ellipse in the image plane. The effect of two-fold astigmatism is proportional to αr^2 Two-fold astigmatism is routinely corrected using stigmators, which produce weak quadrupolar magnetic fields. However, these fields while they correct for two-fold astigmatism they strengthen four-fold astigmatism. Even in this case there is axial astigmatism which is due to imperfections in the lens field and construction.

The fourth aberration is distortion, which cause a displacement of off-axis points. In this case it is proportional to r^3 so that a square in the object plane will appear as square pizza dough whose vertices were pinched (pin cushion distortion) or buckled in (barrel distortion). It is also possible that the points don't just translate along the radial direction, but they can also rotate and this is known as spiral distortion. Naturally, on a circle distortion has the effect of either enlarging or reducing the circle.

The last of the Seidel aberrations is curvature of field, it is minor with respect to the others, proportional to αr^2 and it has the effect of blurring the edges of the field of view.

2.1.5 Relevant approximations and techniques

In this section I will show the theory required to understand and interpret the experimental data of the techniques used to characterize the electron phase plates that we made and used for our research project. We will first see how to interpret and simulate diffraction patterns and then we will see most commonly used technique to measure the phase of the electron beam after it has interacted with the sample.

2.1.5.1 *Fresnel – Fraunhofer diffraction*

In optics, both light and electron ones, the diffraction of an object can be divided between two regimes, the Fresnel-diffraction regime and the Fraunhofer-diffraction regime. The first one is also known as near-field diffraction, while the second one is known as far field diffraction.

In both cases it is possible to approximately calculate the wave function at a certain distance z from a source point (X, Y). So that from the wavefunction it is also possible to then evaluate the intensity distribution of the diffraction.

For the Fresnel case, we first need to start by considering that if our source point is on the plane A and the detection point (x, y) on plane A', then the distance between the two point can be approximated as

$$r = \sqrt{z^2 + (X - x)^2 + (Y - y)^2} \approx z + \frac{(X - x)^2 + (Y - y)^2}{2z} \tag{61}$$

where z has been assumed to be greater than all other in-plane distances. If we now assume that the wave function in plane A is $\psi_0(X,Y)$, then the wave function at a distance z is approximately given by:

$$\psi(x,y) = -i\frac{e^{i2\pi kz}}{z\lambda}\int\int \psi_0(X,Y)e^{i\frac{\pi}{\lambda z}[(x-X)^2+(y-Y)^2]}dXdY \tag{62}$$

which is known as the Fresnel propagation equation. This equation can be used to explain a series of electron diffraction effects, e.g., fringes around the edge of a sample or aperture in imaging mode. These fringes are often used to correct for astigmatism in the electron microscope. It is then worth to examine the Fresnel diffraction from the straight edge of an opaque aperture. The aperture covers half of the space in the X direction and the incident wave is a plane that propagates along the z direction with $\psi_0(X,Y) = 1$ for $X > 0$. A schematic view of the situation is represented in Figure 4. Considering all this we can rewrite Eq. 62 as

$$\psi(x,y) = -i\frac{e^{i2\pi kz}}{z\lambda}\int_0^\infty\int_{-\infty}^\infty e^{i\frac{\pi}{\lambda z}[(x-X)^2+(y-Y)^2]}dXdY \tag{63}$$

To evaluate this integral we need to introduce the Fresnel integral in its general form defined as

$$\int_0^S e^{\frac{i\pi}{2}s^2}ds = \int_0^S \cos\left(\frac{\pi}{2}s^2\right)ds + i\int_0^S \sin\left(\frac{\pi}{2}s^2\right)ds = C(s) + iS(s) \tag{64}$$

In the Amplitude-Phase Diagram $C(s) + iS(s)$ produces the so called "Cornu spiral"(Figure 5a). In the limit for s to positive or negative infinite, $C(s) + iS(s)$ approaches the values $\left(\frac{1}{2} + i\frac{1}{2}\right)$ for positive infinite and $-\left(\frac{1}{2} + i\frac{1}{2}\right)$ for negative. Owing to this, it is possible to rewrite Eq. 63

$$\psi(x,y) = (1-i)\frac{e^{i2\pi kz}}{2}\int_{s_0}^\infty e^{i\frac{\pi}{2}[(s)^2]}ds = (1-i)\frac{e^{i2\pi kz}}{2}\left[\int_0^\infty e^{i\frac{\pi}{2}[(s)^2]}ds - \int_0^{s_0} e^{i\frac{\pi}{2}[(s)^2]}ds\right] =$$

$$= (1-i)\frac{e^{i2\pi kz}}{2}\left[\left(\frac{1}{2} + i\frac{1}{2}\right) - C(s_0) + iS(s_0)\right] \tag{65}$$

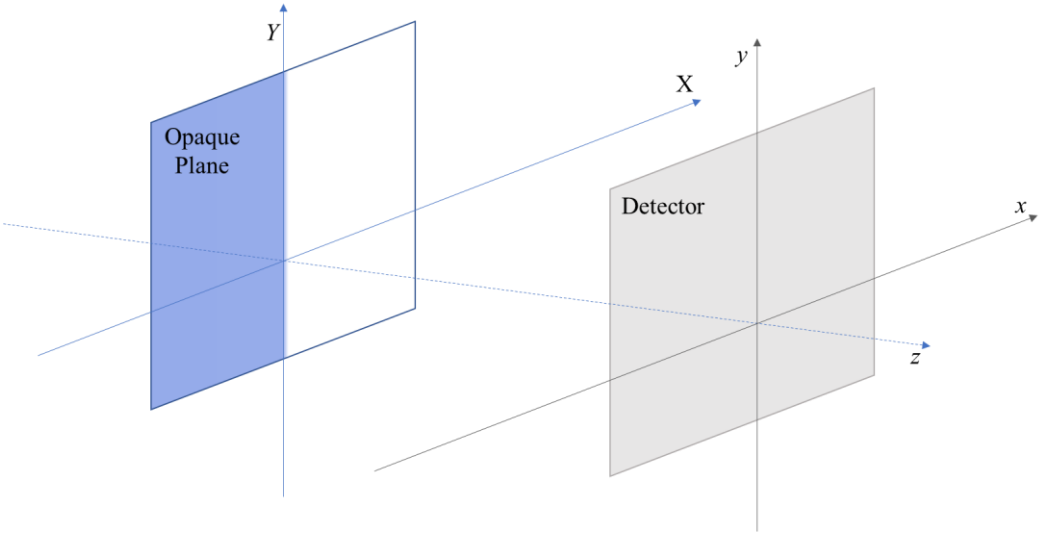

Figure 4: Schematic representation of the half-opaque aperture and detector for Fresnel diffraction

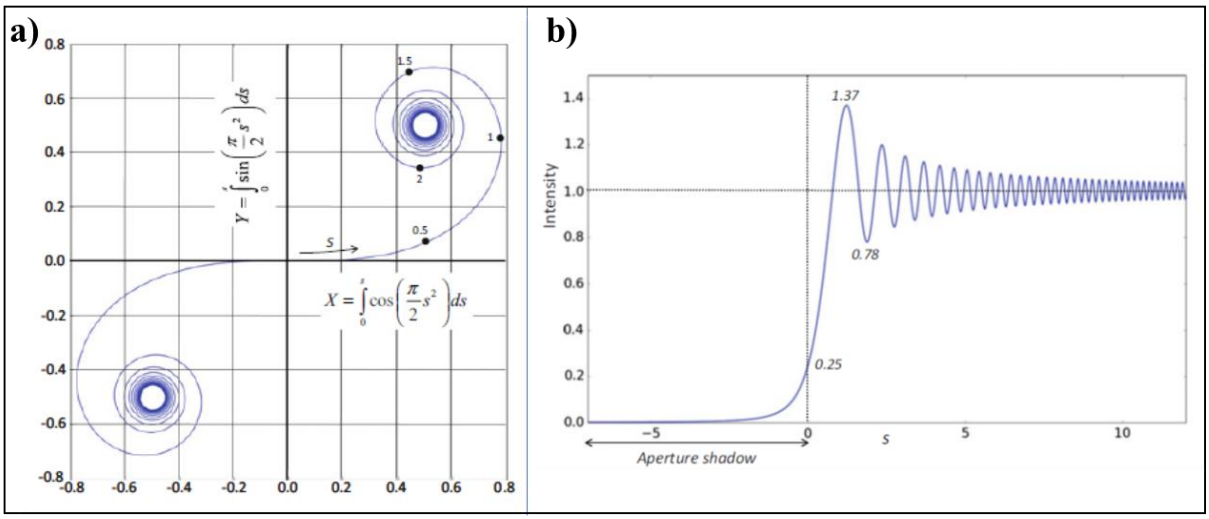

Figure 5: a) Amplitude Phase Diagram of $(C(s) + iS(s))$ with varying s, it is possible to observe the so called Cornu spiral. b) shows how the intensity at the detector point (x,y) calculated from Eq.65 varies with varying s_1. Both graphs are taken from [29]

where $s = \sqrt{2(x-X)^2/\lambda z}$, $s_0 = \sqrt{2x^2/\lambda z}$ and $\int_{-\infty}^{\infty} e^{i\frac{\pi}{\lambda z}[(y-Y)^2]} dY = \sqrt{\frac{\lambda z}{2}}(1+i)$. Figure 5b Shows how the intensity, calculated as $I(x,y) = |\psi(x,y)|^2$ varies in function of the parameter s_0. It is possible to see how, in the case of an opaque aperture with a straight edge, the intensity along the x direction on the detector oscillates (it is the value of x that varies s_0) forming he known fringes. As said earlier, fringes can be observed in out of focus electron images, and the number of fringes is determined by the coherence of the source, the more coherent it is the more fringes there will be.

In the far-field approximation (Fraunhofer), the electron detector is placed at effective long distance from the object, so that in this case the distance between the source point (X, Y) on the plane A and the detection point (x, y) on plane A', can be approximated as

$$r = \sqrt{z^2 + (X-x)^2 + (Y-y)^2} \approx \sqrt{z^2 + x^2 + y^2} = R \tag{66}$$

and the product between the distance and the wave vector as

$$kr = k\sqrt{z^2 + (X-x)^2 + (Y-y)^2} \approx kr - \frac{x}{\lambda R}X - \frac{y}{\lambda R}Y \tag{67}$$

In this case to estimate the wavefunction we start from the general formulation of the Huygens-Fresnel integral [29,30]:

$$\psi(x,y) = -i\int\int \frac{\psi_0(X,Y)}{r\lambda}\left(\frac{1+cos\theta}{2}\right)e^{i2\pi kr}dS \tag{68}$$

where S is the surface area of the wavefront, θ is related to the diffraction direction. For far-field electron diffraction most of the diffraction occurs in the forward direction, so that $cos\theta \approx 1$. With this and the two approximations introduced in Eq.66 and Eq.67 we can rewrite Eq.68 as

$$\psi(x,y) = -i\frac{e^{i2\pi kr}}{r\lambda}\int\int \psi_0(X,Y)e^{-i2\pi(k_xX+k_yY)}dXdY \tag{69}$$

where $k_x = \frac{x}{\lambda R}$ and $k_y = \frac{y}{\lambda R}$. This equation is known as the Fraunhofer diffraction equation and it relates the wave function at the far field to the exit-face wave function at the downstream face of the sample. It can also be noticed that the integral is actually a Fourier transform integral, meaning that while carrying out the calculations we can apply the properties of Fourier Transforms. For example, in the Fourier representation two propagation steps can be easily concatenated, since they correspond to multiplicative phase factors [36].

2.1.5.2 *Phase Object approximation*

One useful approximation that is extensively used in electron microscopy and that helps to understand how the wave function of the electrons changes after it passes through a thin sample is the phase-object approximation in the high-energy regime.

For a purely electrostatic sample whose internal electrostatic potential is $V_i(x,y,z)$ with $V_i \ll V_a$, so that it can be considered as a small perturbation for the electrons. Then, if ψ_0 is the wave function of the electrons before entering the sample, then the perturbed wave function after the interaction can be written as [36]

$$\psi = \psi_0\chi \tag{70}$$

where χ, the perturbation related transmission function, in the case ψ_0 is a plane wave can be calculated from

$$\nabla^2\chi + \frac{4\pi i}{\lambda}\frac{\partial\chi}{\partial z} + \frac{4\pi^2}{V_a\lambda^2}V_i\chi = 0 \tag{71}$$

The weak phase object approximation consists in neglecting the term $\nabla^2\chi$ so that we can rewrite the previous equation as

$$\frac{\partial\chi}{\partial z} = \frac{i\pi}{\lambda V_a}V(x,y,z)\chi \tag{72}$$

We can calculate the expression of $\chi(x,y,z)$ by integration between two generic planes z_i and z_f:

$$\chi(x,y,z_i) = exp\left[\frac{i\pi}{\lambda V_a}\int_{z_i}^{z_f}V(x,y,z)dz\right] \tag{73}$$

In the case we also consider the vector potential, so the also the magnetic field, of the sample, then we can rewrite Eq.72 as

$$\frac{\partial\chi}{\partial z} = \frac{i\pi}{\lambda V_a}V(x,y,z)\chi - \frac{ie}{\hbar}A_z\chi - \frac{ie^2\lambda}{4\pi\hbar^2}A_z^2\chi - \frac{\lambda e}{2\pi\hbar}\vec{A}\cdot\vec{\nabla}\chi \tag{74}$$

where A_z is the *z*-component of the vector potential and we added three further terms. In the case the first term, $-\frac{ie}{\hbar}A_z\chi$, is different from zero then the other two can be neglected. In this way we recover the non-relativistic corrected standard formula of χ:

$$\chi(x,y,z_i) = exp\left[\frac{i\pi}{\lambda V_a}\int_{z_i}^{z_f} V(x,y,z)dz - \frac{ie}{\hbar}\int_{z_i}^{z_f} A_z(x,y,z)dz\right] \tag{75}$$

It must be noted that it is possible to recover the same result with other important approximations as suggested in reference [36].

What Eq.75 tells, in summary, is that when the electron wave passes through a very thin sample according to the phase object approximation the electron wave function will have an addition phase term given by

$$\Delta\varphi = \frac{\pi}{\lambda V_a}\int_{z_i}^{z_f} V(x,y,z)dz - \frac{e}{\hbar}\int_{z_i}^{z_f} A_z(x,y,z)dz \tag{76}$$

however nowadays the relativistically corrected form is more accepted so that

$$\Delta\varphi = C_E\int_{z_i}^{z_f} V(x,y,z)dz - \frac{e}{\hbar}\int_{z_i}^{z_f} A_z(x,y,z)dz \tag{77}$$

where

$$C_E = \frac{2\pi}{\lambda}\frac{e}{E}\frac{E_0 + E}{2E_0 + E} \tag{78}$$

and E_0 is the electron energy at rest and E is the energy of the moving electrons. C_E is known as the electron interaction constant and as it is possible to notice it depends on the value of the electron accelerating potential, it is smaller the higher the accelerating potential is.

The phase object approximation is widely used in electron microscopy, mostly to analyse the results from Off-axis electron holography, which is presented in the next sections.

2.1.5.3 *Further notes on Wave Interference*

Before I can introduce Off-Axis electron holography, it is worth to reintroduce the concept of wave interference, which I already discussed in the first Chapter of this thesis. In fact, thanks to the well-known properties of wave interference it is possible to obtain information regarding both the amplitude and phase of a generic wave (or wave of interest). This is done by analysing the interference pattern of the wave of interest with a "known" reference wave. From a physical point of view a hologram is generated by the interference between the wave function of interest $\Psi_I(\vec{r})$, i.e.:

$$\Psi_I(\vec{r}) = A_I(\vec{r})e^{i\varphi_I(\vec{r})} \tag{79}$$

where $A_I(\vec{r})$ and $\varphi_I(\vec{r})$ are, respectively, the interest beam's phase and amplitude, and a reference wave, $\Psi_{ref}(\vec{r})$. The idea of holography is to write in 2D the interference between two waves propagating in the 3D space. This is possible due to the fact that in vacuum the wave equation, for example, written in the form of Helmholtz equation, constrains the wave behavior outside a specific 2D plane. The 2D plane, therefore, contains all the degrees of freedom of the wave and a generic description in any other point of space. We consider a specific plane with coordinates $\vec{\rho} = (x, y)$and an out of plane direction z. The total wave function in 3D is then:

$$\Psi_{holo}(\vec{r}) = \left|\Psi_I(\vec{r}) + \Psi_{ref}(\vec{r})\right| \tag{80}$$

but in a specific plane:

$$\Psi_{holo}(\vec{\rho}) = \left|\Psi_I(\vec{\rho}) + \Psi_{ref}(\vec{\rho})\right| \tag{81}$$

The intensity is therefore:

$$I_{holo}(\vec{\rho}) = |\Psi_I(\vec{\rho})|^2 + \left|\Psi_{ref}(\vec{\rho})\right|^2 + 2\,Re\left[\Psi_I(\vec{\rho})\Psi^*{}_{ref}(\vec{\rho})\right] \tag{82}$$

Or alternatively:

$$I_{holo}(\vec{\rho}) = |\Psi_I(\vec{\rho})|^2 + \left|\Psi_{ref}(\vec{\rho})\right|^2 + 2|\Psi_I(\vec{\rho})|\left|\Psi_{ref}(\vec{\rho})\right|\cos\left(\varphi_I(\vec{r}) - \varphi_{ref}(\vec{r})\right) \tag{83}$$

where $\varphi_{ref}(\vec{r})$ is the phase of the reference beam. The role of the reference wave is to make the phase $\varphi_I(\vec{r})$ visible as an intensity modulation. The reference wave must have a very simple form, such as a plane wave or a spherical wave (sometimes substituted by a parabolic approximation). It is customary to refer to the process as "inline" or "in-axis" holography if the two waves propagate in the same direction. If the two propagation directions are different, then, we refer to it as "off-axis" holography.

2.1.5.4 *"image" holography for object phase reconstruction*

If one considers $\Psi_I(\vec{r})$ as the wavefunction obtained after the electron beam interacts with a partially electron transparent sample with an unknown phase, holography can be used to extract such phase information.

This kind of holography, so-called off-axis holography, is performed by splitting the beam wavefront into two parts by means of an electrostatic biprism. In electron microscopy, an electrostatic biprism is a charged wire. In most cases, one of the beams (after the electrostatic biprism) traverses through the specimen. The biprism causes tilts to the beams, and thus they interfere with each other. One of the two waves, the object wave, interacts with the sample and gains a phase that is proportional to the sample physical features. Meanwhile the other part of the beam, the reference wave, apart from the tilting introduced by the biprism, is not affected. The intensity in a specific plane of the resulting hologram between the generic beam of interest in Eq. 79 and a tilted plane wave is described by the following expression:

$$I_{holo}(\vec{r}) = \Psi_{holo}^2(\vec{\rho}) = \left|\Psi_I(\vec{\rho}) + \Psi_{ref}(\vec{\rho})\right|^2$$

$$= 1 + A_I^2(\vec{\rho}) + 2A_I(\vec{\rho})\cos(\varphi_I(\vec{\rho}) + \vec{g}\cdot\vec{\rho}) \quad (84)$$

where $\vec{g}$ is the in-plane component of the wave vector of the plane wave and is determined by the tilt angle introduced by the biprism. In this expression, three different contributions to the total intensity can be distinguished: the reference image intensity, the specimen image intensity, and a set of cosinusoidal fringes. The latter's local phase shifts and amplitudes are respectively given by the phase and amplitude of the electron wavefunction in the image plane. Information about the phase and the amplitude of the wave function of interest can be extracted from the electron hologram image. This information can be obtained following this procedure: applying a Fourier Transform (FT), removing the artifacts, reconstructing the hologram by means of Inverse Fourier Transform (IFT), and finally removing the 2π discontinuities in the phase (i.e., unwrapping the phase), such that eventually a complex image of the hologram can be obtained. The FT of the expression in Eq. 81 can be written in the following form:

$$FT[I_{holo}(\vec{r})] = \delta(\vec{k}) + FT[A_I^2(\vec{r})] +$$
$$+\delta(\vec{k}+\vec{g}) \otimes FT\left[A_I(\vec{r})e^{i\varphi_I(\vec{r})}\right] + \delta(\vec{k}-\vec{g}) \otimes FT\left[A_I(\vec{r})e^{-i\varphi_I(\vec{r})}\right] \quad (85)$$

where $\delta(.)$ is Dirac's delta and $p \otimes g$ stands for the convolution of p and g. In this expression, the first two terms are the FT of the reference and the sample wavefunctions, respectively, which create a peak at the origin $\vec{k} = 0$. The last two terms, however, are peaked at $\vec{k} = \pm\vec{g}$, and they correspond to the FT of the desired image wavefunction for the minus sign and its complex conjugate for the plus sign. The images peaked at $\vec{k} = \pm\vec{g}$ are known as the *sidebands,* while the image in the origin can be referred to as the *centre-band*. In addition, the larger the value of $\vec{k}$, i.e., the higher the tilt of the reference wave is, the further away from the origin these sidebands will be. By observing the expression of the sidebands in Eq. 85, one can notice that they include the amplitude and phase information of the wavefunction of interest. To recover the complex wavefunction, one of the two side peaked patterns is selected and isolated by applying a circular mask, mathematically shifted to the origin of the reciprocal space and then inverse Fourier transformed. The applied mask must have soft edges, along with a radius which is approximately one-third of the distance between the sideband and the origin, given that the radius of the center-band is twice that of the sideband [37].

The amplitude and phase of the complex wave function of interest in the image plane are defined as:

$$A_I(\vec{\rho}) = \sqrt{Re^2 + Im^2} \quad (86)$$

$$\varphi_I(\vec{\rho}) = \tan^{-1}\left(\frac{Im}{Re}\right) \quad (87)$$

Where Re and Im are the real and imaginary apart of the complex wave-function, respectively.

The reconstructed amplitude image is very similar to an image obtained by energy-filtered bright-field TEM since in the holographic interference formation the effect of inelastic scattering is negligible [38]. On the other hand, the phase image may need to be further processed to remove the 2π phase discontinuities which appear at the positions where the phase shift exceeds 2π. These are because of the FT and IFT operations which generate a phase image that is initially calculated as the modulo of 2π. To correct the discontinuities, the phase image can be either "unwrapped" with proper algorithms [39] or it might be corrected to remove the "false" constant phase gradient that is introduced by different positioning of the center of the sideband.

Long-range phase modulations arising from inhomogeneities in the charge, thickness of the biprism wire, lens distortions and charging effects (e.g., at the apertures) can all introduce artifacts in the reconstructed wavefunction given that the phase information is stored in the lateral displacements of holographic interference fringes.

To account for these effects, a reference hologram is usually taken from a vacuum region without changing the electron-optical parameters of the microscope. The corrections to obtain the distortion-free phase of the wave of interest are then carried out by performing a complex division of the specimen wavefunction by the vacuum wavefunction in real space and calculating the phase of the resulting complex wavefunction.

For thin and weakly diffracting specimens it is possible to neglect dynamical diffraction. Eq. 77 and Eq. 87 allow us to retrieve information on the electrostatic potential and magnetic vector field of the specimen. In non-magnetic material the phase retrieved via off-axis holography is then given by

$$\varphi_I(x,y) - \varphi_0(x,y) = C_E \int_{z_i}^{z_f} V(x,y,z)dz \tag{88}$$

where φ_0 is the phase of the complex wave prior its interaction with the specimen. In the particular case in which no external charge distributions or applied electric fields are present within or around the specimen, then only the mean inner potential V_{mip} of the material coupled with the specimen thickness variations contribute to the electrostatic phase shift

$$\varphi_I(x,y) - \varphi_0(x,y) = C_E V_{mip} t(\vec{\rho}) \tag{89}$$

where $t(x,y)$ is the specimen thickness at position (x,y). Eq. 89 enables us to determine the mean inner potential if the specimen thickness is known (usually through complementary techniques) or alternatively the specimen thickness if the specimen has a uniform composition.

In other cases, the most common way is to first construct a model of the situation in study to write an analytical expression of the potential distribution $V(x,y,z)$, and from that calculate the phase shift using Eq. 77. By comparing the simulation and the experimental results it is then possible to fine tune the model to properly reproduce the realistic case.

2.1.5.4.1 *Case study of a Charged Tip*

One key example that is relevant for this thesis is the case study of a charged tip, first studied by Matteucci *et al.* [40]. The model starts from considering two oppositely charged linear segments, placed along the y-axis in a symmetric position with respect the xz-plane of an (x, y, z) coordinate system where z is oriented along the optical axis. Each segment has a length of $2c$ and their centers are at a distance $2h$ from each other as reported in Figure 6.

The analytical expression of the potential distribution can be obtained from

$$V(x,y,z) = \frac{1}{4\pi\varepsilon_0}\int_{-c}^{+c}\frac{\sigma}{\sqrt{x^2+(y-h-t)^2+z^2}}\mathrm{d}t + \frac{1}{4\pi\varepsilon_0}\int_{-c}^{+c}\frac{-\sigma}{\sqrt{x^2+(y+h-t)^2+z^2}}\mathrm{d}t \tag{90}$$

which gives

$$V(x,y,z) = \frac{\sigma}{4\pi\varepsilon_0}\left[sinh^{-1}\left(\frac{c-(y-h)}{\sqrt{x^2+y^2}}\right) - sinh^{-1}\left(\frac{-c-(y-h)}{\sqrt{x^2+y^2}}\right) + \right.$$
$$\left. + sinh^{-1}\left(\frac{-c-(y+h)}{\sqrt{x^2+y^2}}\right) - sinh^{-1}\left(\frac{c-(y+h)}{\sqrt{x^2+y^2}}\right)\right] \tag{91}$$

from which it is possible to notice that the potential distribution is rotational symmetric around the y-axis and it vanishes on the $y = 0$ plane. The equipotential surfaces behave approximately as a family of hyperboloids near and around the extremities of the two charged lines. This model can be reasonably used, at least in first approximation, to model and study the situation of a charged tip in front of a conductive plane. This is valid because one can retrieve the oppositely charged tip through the image charge method.

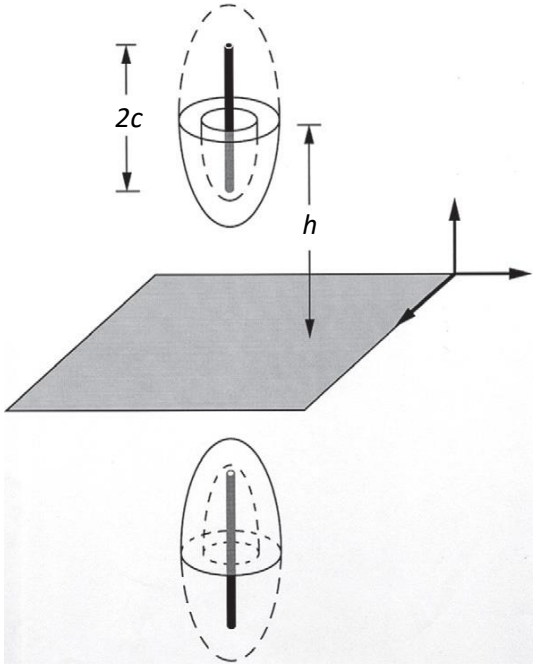

Figure 6: Geometry of the theoretical model to calculate the field near a charged microtip. The image is taken from [40].

Using Eq.77 with $A_z = 0$, $z_f = +\infty$ and $z_i = -\infty$ it is possible to calculate the analytical expression of the phase shift induce to the electron wavefunction [40]:

$$\Delta\varphi(x,y) = 2C_E \frac{\sigma}{4\pi\varepsilon_0}\left\{[-c+(y-h)]ln\left(\sqrt{x^2+[c-(y-h)]^2}\right) + |x|sin^{-1}\left(\frac{-c+(y-h)}{\sqrt{x^2+[c-(y-h)]^2}}\right) + \right.$$

$$+[-c-(y-h)]ln\left(\sqrt{x^2+[c+(y-h)]^2}\right) + |x|sin^{-1}\left(\frac{c+(y-h)}{\sqrt{x^2+[c+(y-h)]^2}}\right) +$$

$$+[c+(y+h)]ln\left(\sqrt{x^2+[c+(y+h)]^2}\right) + |x|sin^{-1}\left(\frac{c+(y+h)}{\sqrt{x^2+[c+(y+h)]^2}}\right) +$$

$$\left.+[c-(y+h)]ln\left(\sqrt{x^2+[c-(y+h)]^2}\right) + |x|sin^{-1}\left(\frac{-c+(y+h)}{\sqrt{x^2+[c-(y+h)]^2}}\right)\right\} \quad (92)$$

The computer simulation of the equiphase lines obtained by the coherent superposition of the object wave and a plane reference wave is reported in Figure 7. Here, between two successive dark and white curves is present a phase difference of 2π. It is also important to notice how the equiphase lines seem to enter the tip shadow, a striking different behaviour with respect to the equipotential lines reported in Figure 2 of [40]. The authors recognized that at the time it was the first example where such a difference between equiphase and equipotential lines occurred.

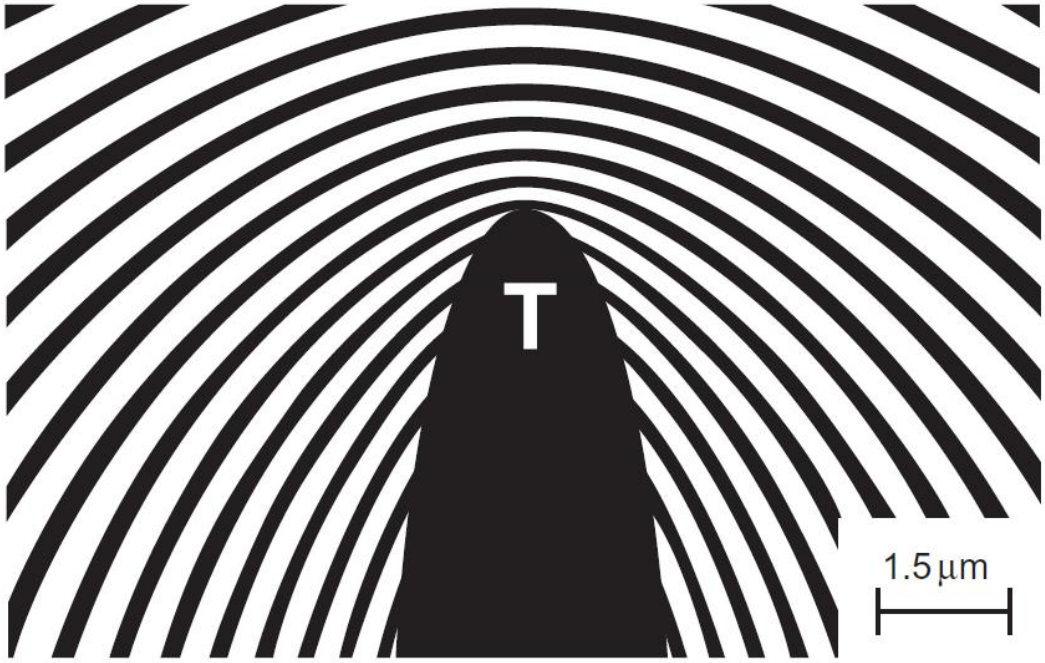

Figure 7: Computer simulation of the equiphase lines around the tip T. Image taken from [40].

An experimental validation of their model is also present in the paper, Figure 5 of [40], thanks to which they refined their model to include the fact that reference beam is also modulated by the field of the tip, which extends microns away from the tip itself, thus making the reference wave a perturbed reference wave (which

is always considered nowadays in studies that involve off-axis holography). This allowed them to reach a satisfactory agreement between experimental and theoretical results.

This example is paramount for this thesis as it is the starting point for the realization of the "unwrapper" element for an electrostatic OAM sorter. The theory of the Ideal OAM sorter is reported in Chapter 3, while the theory of the electrostatic one and its realization is reported in Chapter 5.

Greater details on Off-Axis Holography and examples of its possible applications can be found in book chapters such as [41–43]. As it is shown in Chapter 4, another possible used of the weak phase approximation and off-axis holography is to create holographic phase plates to create structured beam, a practice also known as beam shaping.

2.2 The Dual-Beam (SEM-FIB)

The dual-beam machine, as previously introduced, is in most of the cases made with an SEM and a FIB. The first is used mainly for imaging purposes, while the second is used to for both imaging and sample analysis and modification. In Figure 8 we can see how the two machines are installed in the dual beam: the SEM column is mounted vertically, while the FIB column is commonly inserted so that is forms a 52° angle with respect to the vertical direction. The sample stage has the ability to translate along three direction (x, y, z with z being the vertical axis), to rotate around the z axis and to tilt to bring the sample perpendicular to the FIB column.

In the next few paragraphs for each machine, I will first show the main components and how the beam interacts with the sample, then we will see most of the prominent applications and uses of both machines.

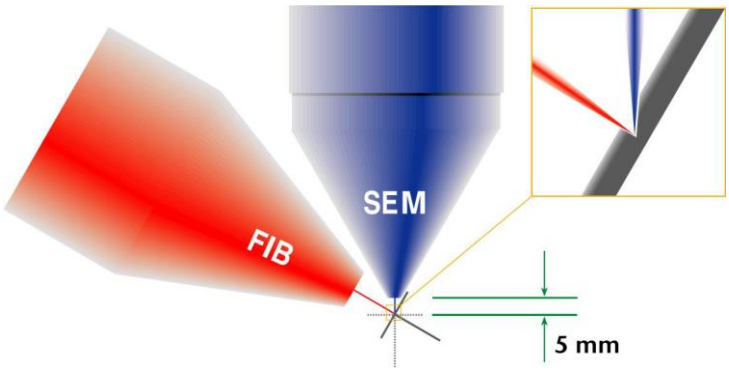

Figure 8: SEM and FIB nozzle configuration inside the dual-beam main chamber. As it is possible to see the two are at an angle with eachother, that in most of the cases is approximately 52°.

2.2.1 The Scanning Electron Microscope (SEM)

The SEM, like the TEM, allows to examine and analyse the morphology and chemical composition of a sample. Naturally, there are certain features that are similar if not identical between the two machines. The electron

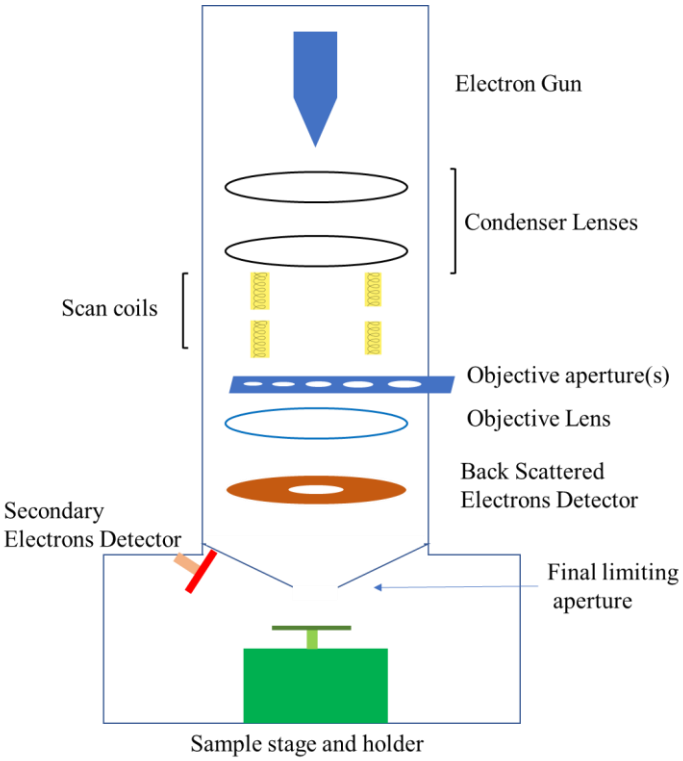

Figure 9: Schematic representation of the main components inside an SEM

source used in a SEM is practically identical to a TEM one (described in section 2.1.1.1) and nowadays almost all SEMs have either a Schottky or F-E source. The illumination system is also very similar, it commonly features two condenser lenses and an objective lens. In this case, the objective lens is not an immersion one and here is used to focus the beam on the sample. The apertures in the illumination system are very important in the SEM since they are used to control the spherical aberration of the beam and to exclude scattered electron that may go back into the illumination system. Moreover, the apertures are used to modify the beam current and the beam angle (subtended with the optical axis), resulting in an enhancement of the depth of field (which will be explained later). In the SEM, as the name might suggest, the electron beam is controlled so that it can scan a certain area of the sample. The deflection is accomplished by using paired coils, however this allows to scan only along one direction, for x-y scanning a set of two orthogonal pairs of are employed. In the column there are normally two sets of these so that the beam always passes through the optic axis of the objective lens. Most of the components just described and their approximate location inside the SEM column are reported in Figure 9. What is missing from the sketch in Figure 9 are the stigmators that are fundamental to correct the elliptical cross section of the beam into a circular one, like for the TEM. The elliptical shape of the beam again arises from imperfections and defects in the lenses and aperture contaminations. [44,45]

Unlike the TEM, thanks to the typical lower energies of the electron beam used in the SEM (electrons are accelerated at energies between 1 to 30 keV) the electron-sample interaction is mainly confined to sample's surface or near surface. Furthermore, the samples can even be a few mm thick, while in the TEM they are usually required to be no more than few hundreds of nm.

In the SEM when the incident beam arrives at the specimen surface penetrates into the sample for some distance before the electrons collide with an atom or atom's electron or get scattered. In doing so, the electron beam excites a certain region of the sample, known as primary excitation region, from which a variety of signals are generated. Most of the signals generated were already described in the early paragraphs of section 2.1, so here we will briefly see the main ones that are used to study the sample. Before doing so, it must be noted that the size and shape of the interaction region largely depend on the beam energy and the sample's atomic number and density. For example, at the same accelerating voltage, the shape of the interaction volume is tear-drop like for lighter elements (sketched in Figure 10) and hemispheric for high Z elements. Moreover, with increasing electron energy the volume and depth of penetration increase. In Figure 10 we can also find the approximate region of origin inside of the interaction volume of the signals used to image and analyse the sample in the SEM. From the first few nanometers of the interaction volume Auger electrons are emitted, which thanks to their characteristic energy can provide useful chemical information on the surface layers of the sample. Usually, Auger electrons are measured in dedicated SEM machines and not in dual beam ones.

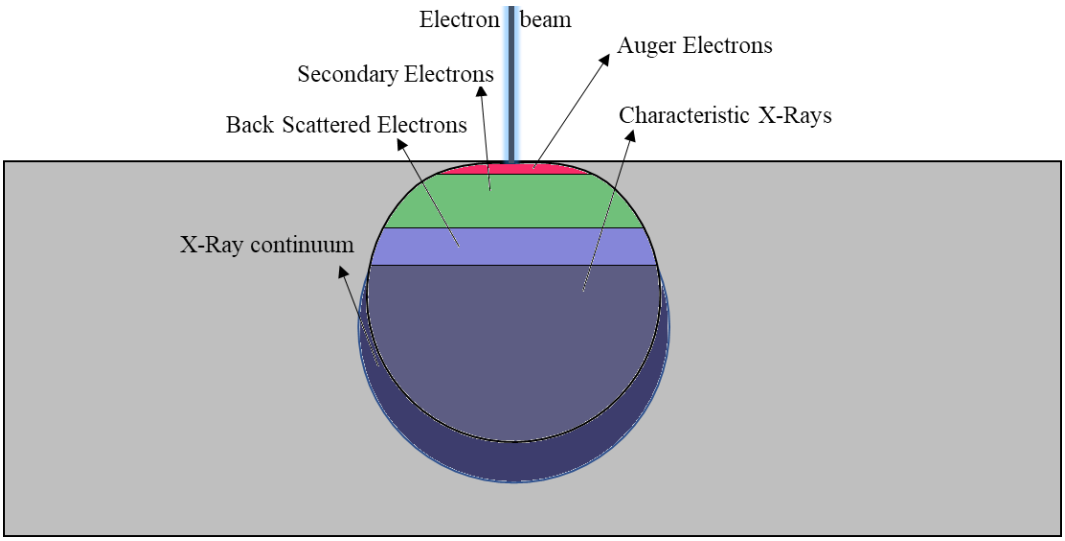

Figure 10: A sketch showing the tear drop-shaped volume of interaction of the beam with the sample and the approximate zone of origin of the signals commonly measured in the SEM.

Next, we have the region from which are emitted the Secondary Electrons that can escape the sample and reach the detector. This region is slightly bigger than the one from which Auger electrons are emitted from. Secondary Electrons (SEs) are loosely bound electrons that are emitted either when the primary beam reaches the sample and ionizes some of the atoms in the interaction volume (SE1) or during other inelastic scattering processes that happen in the interaction volume (SE2). SEs have a low energy, typically in the range of 3-5 eV and never greater than 50eV, so even though they are emitted from the ionized atoms in the whole excitation region, only the ones emitted within few nanometers of the material surface will reach the detector, while the others are reabsorbed in the material. For this reason, SEs accurately give topographic information regarding surface texture and roughness, all with a resolution of 10nm or better. Topographic

contrast is also dependent on the incidence angle between the beam and the surface, when the surface is perpendicular to the beam the zone from which SEs are emitted so the detected intensity will be lower with respect to one that is inclined.

Lastly, is the region from which Backscattered Electrons (BSEs) are scattered from. This region is considerably larger than it is for SEs, and because of this the lateral resolution of a BSEs image is considerably worse than one with SEs (1μm against 10nm). A BSE is an electron which has undergone a single or multiple scattering events and that escapes from the surface with an energy greater than 50 eV. BSE are one of the two main signals used to image the sample, alongside secondary electrons. Roughly 10-50% of BSEs are back scattered toward the electron column final aperture (the percentage depends on the incident angle of the primary beam) and most of these retain around 60-80% of the primary beam energy. In this case, elements with a higher atomic number yield to a higher BSEs signal, thus providing an atomic number dependent contrast in the SEM image. For example, the percentage of BSEs emitted with respect to the number of incident electrons (known as backscattered electrons yield) is around 6% for Carbon, while it is approximately 50% for heavier elements such as tungsten or gold. In examining a sample, the BSEs carry information about features deep underneath the surface and can produce topographical images that differ from SEs ones, because BSEs can be blocked by some regions of the specimen that SEs can be drawn around.

Like in the TEM, the electrons' energy is enough for characteristic X-Ray to be produced. This means that even in the SEM it is possible to carry out EDX measurements to acquire information on the composition and chemical structure of the sample. X-Rays are commonly used to study big regions of the sample since the lateral resolution is slightly better than the one of BSEs, but it is still in the order of few hundreds of nanometers.

SEMs where first built to collect sample's images with secondary electrons, so it comes natural that the first detectors that where implemented and that we will discuss about are SEs detectors. The first detector for SEs was developed by Everhart and Thornley [46]. This device consists of three components: a scintillator that converts the electron signal in light, a light pipe (or nowadays a fiber-optic cable) that brings the light outside of the sample chamber and feeds it to a photomultiplier tube (PMT) that converts the light in electrons again and amplifies the electron's signal so that it can be visualized on a screen. The number of photons generated by the scintillator depend on the scintillator material and on the energy if the impinging electrons, so a bias of 10kV is applied to the scintillator (or in some case an electrode in front of it) so that the electrons that strike it have sufficient energy. Since such high voltage is able to affect also the primary electron beam, usually a faraday cage biased at $\approx 300V$ is placed over the scintillator so that the secondary electrons emitted from the sample are attracted, but the primary beam's electrons are not. The PMT, on the other hand, has a high amplification factor, a logarithmic response so that it can process signals covering a large intensity range, adds low now to the original signal, responds rapidly to signal level changes, and most

importantly nowadays is relatively low in price. As a whole, the Everhart-Thornley (ET) detector has high efficiency and speed, flexible, cheap to construct and require relatively low maintenance, and for this is has been present in almost every SEM built to date.

The most common placement of the SE detector is inside of the main chamber at an angle with respect to the sample. An asymmetric detector geometry has the advantage that in the acquired image topographic elements such as edges, corner, steps or generic surface roughness elements are shadowed or highlighted depending on their relative position with respect to the detector. The naturally present contrast in the image is intuitive, reliably interpretable and produces micrographs that are aesthetically pleasing. However, most of the SEs emitted from the sample are emitted normal to the surface, so that for a typical sample that is almost flat the Everhart-Thornley (ET) detector will be able to collect about 15-30% of the available SE1 and SE2 signal [44]. If we now associate such poor performance with the fact that the detector also receives BSEs and SEs created from the impact of BSEs with the chamber (defined as SE3), then in the final signal only a small percentage of will come from the sample, so that the signal-to-noise ratio is degraded and the contrast reduced.

Modern SEMs are design to improve image resolution, one of the solutions found by engineers is to use a “snorkel” objective lens, that also defines the final aperture and whose magnetic field extends to the sample. In this case, the distance between the sample and the final aperture, known as working distance (WD, a parameter that is relevant for both SEMs and FIB and that will be discussed later), is quite short and the main ET detector is actually placed in SEM column few mm off the optical axis, before the objective lens (a secondary ET detector can be placed in the chamber to image the sample at long WD). This configuration has many advantages over the previously explained one and is known as “through the lens” (TTL) detector. As previously mentioned, secondary electrons are mostly emitted normal to the surface and travel back into the column, moreover, the magnetic field of the objective lens captures the SEs and further channels them back in the column. In this configuration, in some cases the ET detector instead of having the Faraday cage that protects the detector from the primary beam electrons feature a Wien filter. The Wien filter consist of a magnetic field normal to the electric field of the scintillator, whose strength is tuned so that the primary beam remains on-axis, while SEs are directed at the detector. This system can be fine-tuned to be able to select the electrons that arrive to the detector by their energy. A more complex version of the Wien filter equipped ET detector is able to selectively collect also the signal from either SEs or BSEs. In the TTL configuration the ET detector is able to collect 70 to 80% of the SE signal from the specimen [44], meaning that the resulting image has higher contrast and sample information with respect to the “in-chamber” detector. This means that the detector in this configuration is more sensitive to sample properties such as elemental composition, electronic properties or sample charge, while it is less sensitive to topography.

BSEs specific detector can also be either in the sample chamber or TTL, in this case, the faraday cage (when present) is negatively biased to repel any secondary electrons from reaching the detector. Only those electrons that travel in a straight path from the specimen to the detector form the backscattered image. Apart for the commonly employed ET detector, in some case for BSEs solid state detectors can be used and here the signal is directly converted into a current. Furthermore, as previously mentioned the contrast that is observed in the backscattered image is Z dependent.

In the paragraph dedicate to SEs detector the working distance was mentioned as a key parameter for both SEMs and FIB microscopes. It is worth remembering that the WD can be defined as the distance between the final aperture and it is fundamental since it defines the depth of field. The large depth of field of SEMs in many applications put them above optical microscopes. The depth of field in simple terms is the portion of the image that appears acceptably in focus [44], it is linearly proportional to the working distance and inversely proportional to the final aperture diameter (which in some SEMs can be changed between fixed values mounted on an aperture strip) and magnification. Depth of field in the range of 20 to 40μm can be easily achieved at low magnification, while in optical microscopes with the same magnification it is usually one order of magnitude less [45].

Since during my thesis I mainly used the SEM of the dual machine to acquire simple topographic images of the TEM samples prepared with the FIB, for thorough explanations of analytical techniques in SEMs I refer to [44,45]

2.2.2 The Focused Ion Beam (FIB) microscope

The FIB, from a structural point of view, is very similar to the SEM. The main difference is naturally in the source, and the most common FIB source is a Liquid Metal Ion Source (LMIS) [44,47]. This type of source can produce an ion beam of approximately 5 nm in diameter and it comprises a Tungsten needle attached to a reservoir that contains the metal source material. Several metallic elements can be used, however the most common one is Gallium (Ga), this because there are a series of advantages that made Gallium the metal of choice:

- Low melting point: easy to liquify and minimizes any reaction with the tungsten needle;
- Low volatility and low vapor pressure: this ensures that when in use the supply of metal is constant and it can be used in its pure form instead of an alloy (like what happens for Gold sources), moreover both yield to a long source lifetime;
- Low surface free energy: this promotes a viscous behaviour on the tungsten substrate;
- Excellent mechanical, electrical and vacuum properties;
- Can be emitted with a small energy spread and high angular intensity;

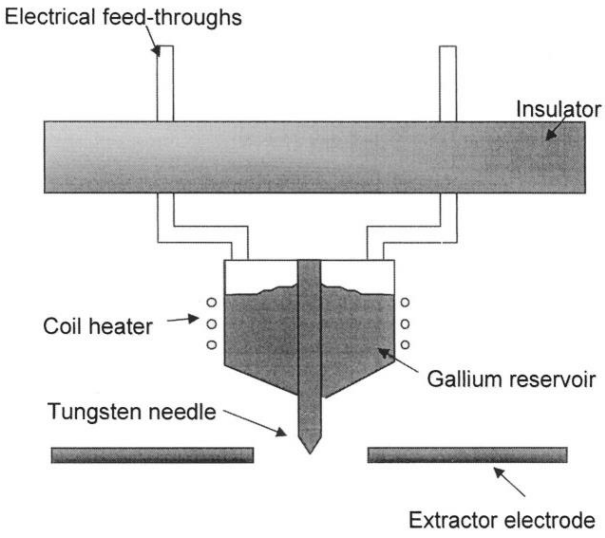

Figure 11: Schematic represantation of a simple Ga-LMIS. Taken from [47].

The ion emission happens in two steps: first the reservoir is heated so that *Ga* wets the tungsten needle that has a tip radius of 2 to 5 μm, a strong electric field ($\approx 10^8\ V/cm$) is applied to the needle and it causes the liquid Gallium to form a cone (known as "Taylor cone") with tip radius of about 2 to 5nm around the tungsten tip, once the electrostatic and surface tension force are in balance. Second, an electrode known as extractor ionizes and pulls the Ga^+ ions, achieving field evaporation of the metal with current density in the range of $10^8\ A/cm^2$. It is worth mentioning that once heated, Gallium remains molten at ambient conditions for weeks when the machine is used regularly or for at least a week or two when used sporadically. In FEI (now Thermofisher Scientific) machines another electrode is inserted between the source and extractor. This detector is known as suppressor, it is normally kept at few hundred Volts and it is used to keep the emission current constant, since it varies as the molten Gallium depletes with use. The typical value of the emission current is between 1 and 3 μA to reduce the energy spread and yield to a stable beam.

As previously mentioned, the rest of the ion column is quite similar to the SEM column as it also features a condenser lens, an objective lens, the $x - y$ deflectors and a strip of apertures between the lenses. However, here the apertures are the main tool used to define the final beam current that hits the sample. Final beam currents ranging from few pA to 20 or even 30 nA can be achieved. The energy spread of the focused ion beam that reaches the sample is generally greater than the energy spread of a focused electron beam accelerated at the sample potential, and it is approximately 5 eV. This is mainly attributed to the ions dimensions, in fact, since ions are much more massive than electrons, space charge effect limit the apparent source size and increase the width of the energy distribution of the emitting ions. It is then possible to evince that chromatic aberration is the limiting factor in resolution and probe size of FIB systems.

Nowadays, machine manufactures, i.e., the companies that produce FIBs and dual beam, have been experimenting with news FIB sources and materials. From the point of new material choices a key example is manufacturer Raith that apart for standard Ga-FIB/SEM machines in its portfolio has LMIS sources with Gold, Silicon or Silver and many more. Other companies, such as Zeiss, build dual beam machines where

instead of an SEM there is a second FIB column where in this case the ions used are Helium ions, and in this case He is used for both imaging and sample modification. However, the most popular new breed of FIBs are Plasma FIBs(P-FIB), where here the accelerated ions are noble gasses ions such as Ar^+, Xe^+, Ne^+. The main advantage of P-FIBs over standard Ga-FIBs is the faster milling speed, which in some case is almost double of Ga-FIBs. Tescan and Thermofisher have the biggest shares in market of P-FIBs.

For an FIB the working distance is typically larger than in SEMs. This allows for a long depth of focus (as previously mentioned for the SEM) and it permits the introduction of samples with varied topography.

The ions interact with the target material (sample's material) differently with respect to electrons. The ion-solid interaction produces secondary ions (SIs), x-rays, secondary electrons, backsputtered ions, neutral ions and target material's clusters. The penetration depth is definitely smaller with respect to electrons with comparable energy, in fact the penetration depth is almost 2 orders of magnitude smaller than the electrons' one. Moreover, a common event in the ion-sample interaction that doesn't produce a signal is ion implantation in the target material, this results in a change in the material's properties and topography and will be discussed in the next paragraph. The signals that are generated make the FIB the ideal complementary machine for an SEM to study a sample surface. Furthermore, since the ion-sample interaction is always destructive, the ions can be used to mill the sample, so that, for example, one can mill the first few hundreds nanometers of a multilayer sample to obtain a cross-section image of the sample and study if there are defects in the sample structure. This example is just one of the possible dual-beam applications and it is one of the main reasons why such machines are widely used both for academic and industrial research. Nowadays, dual beams are even employed in the production line of IT manufactures for the production of CPUs, GPUs, and memory modules, just to name a few. Apart from imaging and milling, the FIB machine can also be used to deposit materials through a process almost identical to Chemical Vapour Deposition (CVD). In the next few paragraphs, a more in-depth overview of the three main applications will be provided starting from imaging.

As previously mentioned, the ion-sample is surface sensitive and the standard mode for imaging is through the detection of the emitted SEs, but SIs can also be used. On average per incident ion the number of emitted SEs is one to three orders of magnitude greater than the number of SIs [47,48]. However, even though the signal intensity of SIs is lower, it provides vital complementary information particularly for non-conducting regions of the sample. In fact, the non-conducting regions of a sample accumulate a net positive charge under the exposure of the imping Ga^+ ions. The net positive charge inhibits the escape of SEs inducing a dark contrast in the image, while SIs are not influenced by the charge and the resulting image will clearly show the details of the feature. This shows clearly that by acquiring complementary SEs and SIs images it is possible to quickly find non-conducting features and at the same time still image the topography of said features. Another interesting characteristic that can be imaged only when the incoming probe is an ion beam is the

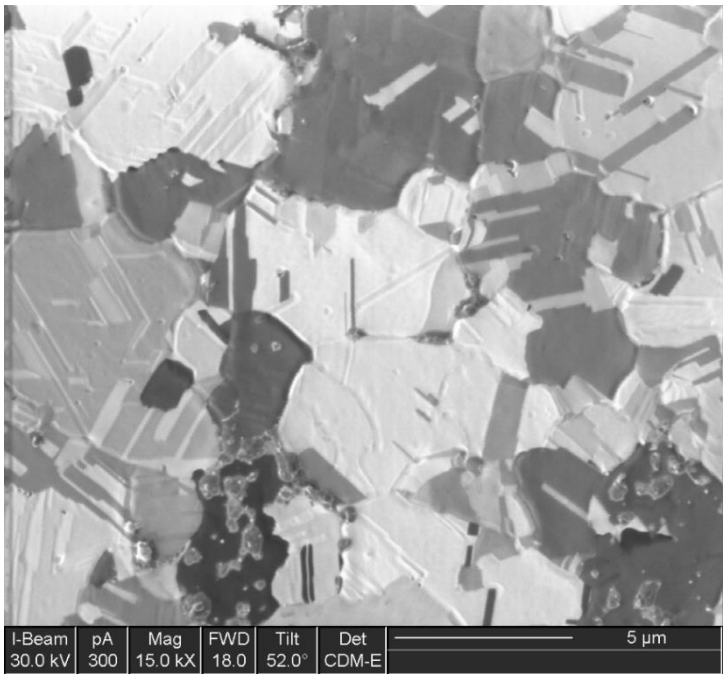

Figure 12: SE image of a polycristalline sample probed with ions. The contrast is mainly due to ion channeling where darker regions correspond to deeper interaction regions so stronger channeling.

varying contrast that can be observed in SE images of polycrystalline samples. The varying contrast present in the SE images of polycrystalline samples is a direct consequence of the phenomenon known as ion channeling [47]. Through channeling the ions may penetrate greater distances along low index crystalline directions compared to non channeling directions or amorphous materials. Thus, the ion range inside the sample is increased so that the interaction volume is enlarged and moved further from the surface. An example of an SE image where the ion channeling contrast is clearly visible is shown in Figure 10, here the darker regions correspond to single crystal regions that is aligned to a low index direction. The darker a region appears the deeper the ions have penetrated, this because the deeper the SEs are generated the lower is the number of SEs that escape from the surface and reach the detector.

Apart from introducing contrast during imaging, Ion channeling also affects the milling speed. However, it is necessary to first understand how the FIB mills the sample's surface. Milling takes place as a result of physical sputtering of the sample, i.e., a series of elastic collisions where momentum is transferred from the incident ion to the atoms within a collision cascade region of the sample. Surface and near-surface atoms, or cluster of them, may be ejected/detach from the sample as a sputtered particle if the transferred kinetic energy is sufficient to overcome the surface binding energy (SBE) of the target material. Some of the sputtered particles are ionized and these are what we previously referred as SIs. An important parameter is the sputtering yield, *Y*, defined as the number of ejected particles per incident ion. For typical FIB energies *Y* is in the range of $10^{-1} - 10^{2}$, depending on target and incident angle [49]. Y can be calculated using the formula [50]:

$$Y(E_0) \cong \frac{1}{4} \Gamma_m \frac{\alpha N S_n(E) \Delta x_0}{E_0} \tag{93}$$

where α is a dimensionless function incorporating the incident angle, the mass ratio M_2/M_1 between the target atom mass and the incoming ion mass, and the incoming ion energy E. Moreover, E_0 is an arbitrary minimum energy of the sputtered particles, Δx_0 is the depth interval for which the atoms are set in motion with energy greater than E_0, $S_n(E)$ is the nuclear stopping cross-section, N is the atomic density and Γ_m is defines as

$$\Gamma_m = \frac{m}{\phi(1) - \phi(1-m)} \tag{94}$$

where $\phi(x) = \frac{\mathrm{d}[\log(\Gamma(x))]}{\mathrm{d}x}$, $\Gamma(x)$ is the Gamma Function and $m = m(E)$ is an energy dependent coefficient that is near zero for low energies, 0.5 at medium energies and asymptotically nears 1 at high ion beam energies. In Eq.93 is present only the nuclear contribution to the stopping cross-section since at typical operating FIB energies the nuclear contribution to the stopping power dominates over the electronic one, which can be omitted. Once the sputtered particles leave the surface they possess an average energy between 2 and 5 eV.

A term that was previously mention, but not well explained is collision cascade, this should not be confused with the interaction region since the latter is a volume, while the collision cascade is the ensemble of all the processes that happen after the incoming ions collide with the target atoms and generates a series of secondary energetic collisions. The nature of the cascade is strongly related to the mass ration M_2/M_1, where, as said before, M_2 is the sample's atomic mass and M_1 the ions atomic mass, and it can be divided in three regimes. The first one, or Regime 1, is called single knock-on regime and it occurs when either $M_2 \ll M_1$ or E is low. As the name suggests, the recoiled atoms do not receive enough energy to start the cascade and here the sputtering is minimal and mainly confined to the first few surface layers. In Regime 2, either E is moderate or $M_2 \approx M_1$ (or both) and it is known as linear cascade regime. It is referred as linear because the recoiled atoms receive enough energy to start a cascade, but the density of moving atoms is low enough so that both multiple collisions and collisions between atoms are seldom. In this regime is possible to apply the binary collision approximation to the incoming ions, so that each collision can be considered between the ion and a single target atom and the path between collisions for the moving ion is straight. This regime is where the FIB generally operates. Lastly, the third regime, also known as spike regime, and here $M_2 \gg M_1$ and/or E is large. Here the majority of the atoms within a spike-shaped volume move during the collision cascade.

In the linear regime the dominant mechanism of energy loss involves elastic interactions between ion and a screened nucleus. This can be modelled using a two body billiard ball collision model and here the "collision" is the distance of closest approach governed by interatomic potentials between the incident ion and the target atom. Here, physical phenomena tend to exhibit periodic fluctuations based on electronic structure and atomic radii of the atoms under consideration. Moreover, since it is possible to consider the binary

collision approximation in the linear regime, then the elastic energy losses can be assumed to mainly be due to a series of uncorrelated binary collisions and it is then possible to model the resulting phenomena with Monte Carlo computer simulations. The main software package that is used to model the ions behaviour in the sample is “Transport of Ions in Matter” (TRIM- that is part of the SRIM software package [51]), which can be efficiently used to physically model and generate data regarding the final 3D spatial distribution of ions in the target material, or to generate all the kinetic data in relation to the energy loss of the ion to the solid. For example, TRIM can quantitatively model phenomena such as target damage, average sputtering yield per incident ion, ionization and phonon production. However, TRIM treats all targets as amorphous materials, so it fails to take into account the possible contribution of channeling or other orientation dependent phenomena.

As introduced in the discussion of the different sputtering regimes, the incoming ion beam energy is an important factor that affects the sputtering yield. In fact, from a statistical point of view more energic particles (in this case ions) will have more collisions, thus increasing the sputtering yield. Furthermore, material properties of the target can influence sputtering, for example the melting temperature is a direct indicator of the bond strength and materials with high melting temperature have a low sputtering yield with respect to low melting temperature ones. Thus, it is possible to use the well-known parameters in the periodic table to predict milling behaviour. Lastly, probably the most important parameter that affects the sputtering yield is the ion beam’s incidence angle. For normal incidence the emission of sputtered particles follows a cosine distribution, however as the angle of incidence is increased the direction of maximum emission shifts away from the incoming beam direction. This is schematically represented in Figure 13. A simple explanation for this is that although the dominant mechanism controlling the sputtering yield is the surface bonding energy, the position of the collision cascade is also relevant, and it can be influenced by the incident angle. In fact, at higher incident angle the collision cascade is confined in a region close to the surface

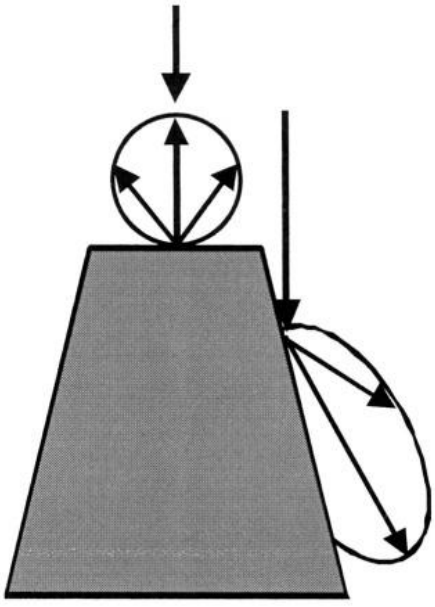

Figure 13: Schematic representation of the emission direction angular distribution at two incidence angles. Taken from [18].

which facilitates sputtering. Simulations and empirical studies [47] have shown that the sputtering yield increases up to a certain angle of incidence and then rapidly diminishes approaching almost zero. The maximum in the sputtering yield occurs between 75-85°, depending on the material. The incidence angle is definitely more important than the ions' energy since, for example, at normal incidence an increase in energy increases both the sputtering yield and the collision cascade depth (it will be located deeper from the surface), however after a certain energy the sputtering yield flattens and stays constant at increasing energy. Nevertheless, the flattening of the yield's curve can be delayed by increasing the incidence angle.

Another process that is strongly related to the incident direction of the ion beam is backsputtering. Similarly for backscattered electrons in an SEM, the incoming ions can be either scattered directly or after a certain number of collisions and exit the sample. The backsputtering yield increases with increasing angle of incidence and increasing mass ratio M_2/M_1. Furthermore, if we now take into account the previous discussion on the sputtering yield and its formula it is possible to notice how, in general, materials with a higher sputtering yield have also a correspondingly high backsputtering yield. At last, the previously introduced ion channeling also affects the sputtering yield, usually it has the effect of reducing it since the collision cascade region moves further from the surface.

Apart from sputtering, there are other relevant processes the first of which is ion implantation. Ion implantation usually is performed in the energy regime where electronic stopping tends to dominate, i.e, high ion beam energies, which are far greater than the typical energies at which the FIB operates. In FIBs ion implantation is mainly affected by the ion beam dose, i.e., the number of ions that impact and are absorbed into the target per unit area. As the target material is being bombarded with moderate energy ions, some of the incident beam ions are implanted and withheld by the target, while due to sputtering the surface is constantly receding. In fact, sputtering affects the implant profile as well as the limiting the concentration of impurity atoms, thus creating a steady-state condition between impurity implantations and removal. Once this steady-state is reached the concentration of impurity atoms (or implanted ions) reaches a maximum state. Furthermore, the concentration limit is inversely proportional to the sputtering yield, thus materials with high sputtering yield will have lower impurity concentration limits. [47,52] It must be noted that implantation often leads to changes in the material properties, for example the SBE may change. The preferential sputtering parameter r, the ratio of the sputtering rate of the target to the sputtering rate of the impurities, can also change. In the absence of crystalline orientation effects and low ion doses the impurity concentration distribution can be approximated to a Gaussian curve. However, for a given beam energy as the dose increases, sputtering also increases and the impurity concentration distribution has a maximum near the surface and falls off over a distance analogous to the ion range. Lastly, the steady-state surface composition for a single element target may be described by the following equation [53]:

$$\frac{N_A}{N_b} = \frac{r}{(Y-1)} \tag{95}$$

where N_A and N_B are the concentration per unit volume of the implanted atom and the target atom respectively, as introduced before r is the preferential sputtering factor and Y is the sputtering yield. Thus, phenomena such as high doses, preferential sputtering, segregation, atomic mixing and chemical effects are all relevant for the steady-state composition.

The second detrimental phenomenon is the amorphization of the target sample. Amorphization is caused by ion implantation and is a metastable phase induced in crystalline materials by ion bombardment and depends on unit cell size, complexity of chemical ordering and the width of an intermetallic phase field [53]. The more complex and bigger the material unit cell is, the larger the amorphous layer will be, while smaller ones are difficult to amorphized. Alloys or materials where the atomic packing arrangement is less stringent than in materials such as stochiometric intermetallics will remain crystalline. For example, Silicon amorphizes when FIB milled, while Copper doesn't (except for specific crystallographic directions). Heating due to FIB milling may also be a factor contributing to the Amorphization, however heating is for most materials deemed to be negligible and confined to the collision cascade region.

Lastly, a further relevant process that influences the final result of FIB milling is redeposition. This topic will be described also in Chapter 3, where we will discuss how synthetic holographic amplitude phase plates can be produced to shape electron beam, however it is important to first introduce it here. Redeposition happens in close proximity to the active milling site where sputtered material and backsputtered ions may deposit on the nearby surfaces. This kind of surface degradation must be kept into account during FIB milling, particularly in the creation of high aspect ratio structures. Thus, controlling or at least knowing how and where redeposition will occur is mandatory for successful and rapid production of high quality specimen for TEM. Redeposition is a function of a multitude of physically and chemically variables, the most relevant of which are: kinetic energy of the atoms leaving the surface, sticking coefficient of the target material, geometry of the feature milled and the sputtering yield of the target material. In fact, when an atom leaves the target material as a sputtered particle it has a finite energy and it can produce secondary interactions with targets that lie in their trajectory. Depending on the energy of impact and the sticking coefficient of the material, the sputtered atom may redeposit or not. The sticking coefficient is a statistical measure of a material's affinity to adhere to a surface and it has been observed that for the low energy range at which FIB sputtered particles operate at there is little difference between different materials. [47] In fact, in a FIB parameters such as geometry of the feature to be milled and sputtering yield are more critical in redeposition effects. As the hole is made deeper, the effect of redeposition becomes stronger and after a certain point the redeposition rate equals the sputtering rate (number of atoms being sputtered per unit time), limiting the aspect ratio. In essence, the rate of redeposition is ultimately a function of the dynamic equilibrium

between the sputtering rate and the capacity of the vacuum system to remove the sputtered atoms from the chamber. So, if the vacuum system is able to evacuate the sputtered material at a rate which is greater or equal to the sputtering rate, then theoretically there can be almost no redeposition. One other possibility to counteract redeposition is by introducing in the milling area a reactive gas species, a process known as reactive gas enhanced racing, where the gas allows to be volatilized and more easily removed by the vacuum system.

Gas delivery systems, however, are mainly used in conjunction with the ion beam (but also with the electron beam in dual-beam machines) to produce site specific deposition of metals or insulators. The typical metals that can be deposited via FIB are Tungsten, Platinum and Cobalt. The deposition is an ion beam assisted Chemical Vapor Deposition (CVD) where the precursor organometallic gas is introduced in the chamber through an inlet capillary that is usually positioned $\approx 100\mu\mathrm{m}$ above the sample surface. In general, the inlet should be positioned as close to the are as possible and must be oriented such that the precursor's quantity impinging on the area is maximized. The gas molecules are adsorbed on the surface in proximity of the inlet and get decomposed only when the ion beam (or the electron beam) strikes. Material build-up results from the continuous adsorption and decomposition in the scanned region. Thus, the precursor must have a sufficient probability to stick to a surface of interest in sufficient quantity and it has to decompose by an energic ion beam. It is necessary to point out that the deposition process consist of a fine balance between sputtering and deposition. In fact, if the ion beam current is too high that the sputtering process will be dominant and milling will occur. In any condition, for deposition to happen the precursor needs to decompose more rapidly than it is sputtered. For optimal FIB deposition rate and material composition a combination of FIB current, pixel dwell time (time spent by the ion beam on a single pixel), raster refresh time (the time which passes before the beam returns to a specific) and precursor flux must be chosen. In fact, for a given precursor flux and focused ion beam current, a pixel dwell time and refresh time must be defined to completely decompose as much of the available precursor as possible (as well as to sputter the minimum amount of material as possible) and to allow sufficient build up of the precursor at each pixel before the ion beam returns to the pixel, respectively. Nowadays the manufacturers ship machines with preinstalled recipes for either material deposition or etching that user can then modify or directly use. All FIB deposited films contain not only the desired metals but also incorporate impurities from the incompletely decomposed precursor (such as Carbon and Oxygen), but also Ga from the focused ion beam (if Ga is the LMIS material). The percentage of contamination can vary significantly depending on the deposition conditions, however both Ga and C always constitute a significant amount of the total material composition. The deposition of insulating material is also of great technological importance as FIB induced deposition of Silicon Dioxide can be used to edit Integrated Circuit devices where space is limited. Other possible applications of FIB induced deposition include surface protection for SEM or S/TEM samples, lithography mask repair and fabrication of 3-D structure with nanoscale resolution (and microscale dimensions). As already mentioned, in dual beam

machines also the electron beam can be used for material deposition since the electron beam energy is sufficient to decompose the precursor. Electron beam deposition is able to small high fidelity and precise structure, however the deposition rate is usually a fraction of the one with the ion beam and for metal films deposition the deposited material is almost insulating (impurities content is far higher with respect to the ion beam deposited counterpart which still behaves as a metal).

Chapter 3 – Vortices and Singularities, The Orbital Angular Momentum (OAM) and the OAM Sorter

Vortex waves naturally occur in both classical mechanics and quantum mechanics and the common feature that ties them all is the fact they are waves with helical phase fronts and circulating current density. Vortex waves are one of the most important type of structured waves, i.e., wave fields which differ from the common plane waves or spherical waves or simple gaussian waves. As we saw in section 1.2 of Chapter 1, plane waves and superpositions of them are possible solutions of the Schrödinger equation. However, the wave after the beam has interacted with the sample typically cannot be approximated to a simple plane wave, but it can be described as a superpositions of plane waves. It is also possible to use diffractive elements to change the functional shape of the wave to a generic coherent superposition of plane waves. In both cases we speak about "structured waves", the most interesting and generic forms of structured waves arise from the interference of three or more plane waves. Wave fields consisting of three or more interfering plane waves may contain phase singularities [54], namely dislocations of the phase fronts . Phase singularities indicate the presence of vortices [55–57] as the phase of the wavefunction ($\varphi = Arg[\psi(\vec{r})]$) is well defined around them, and generically it has a nonzero increment for a contour enclosing the singularity:

$$\oint \nabla Arg[\psi(\vec{r})] \cdot d\vec{r} = 2\pi\ell \tag{96}$$

where ℓ in an integer winding number that provides continuity to the phase modulo 2π, it also called topological charge of the vortex and $\ell = 0, \pm 1, \pm 2, \ldots$ Phase singularities appears at points of destructive interference ($\vec{r} = \vec{r_s}$) where the amplitude of the wavefunction vanishes ($|\psi(\vec{r})| = 0$), while the phase is indeterminate. Phase singularities appear as points in 2D and lines in 3D. The typical behaviour of the wave function near the phase singularity is

$$\psi(\vec{r}) \propto |\vec{r} - \vec{r_s}|^{|\ell|} \exp(i\ell\theta) \tag{97}$$

where θ is the azimuthal angle around the $\vec{r} = \vec{r_s}$ point. Waves with such properties are known as vortices because the probability current density $\vec{j} \propto Im(\psi^* \nabla\psi) = |\psi|^2 \nabla Arg[\psi]$ swirls around the phase singularities. The swirling current suggests that vortices should possess an Orbital Angular Momentum (OAM). Vortices are one of the most generic forms of wave fields and have appeared in studies on optics in the 50s [58–60], and even earlier to describe atomic orbitals with angular momentum [61,62] or eigenmodes of the Schroedinger equation in a magnetic field [63]. However, it was only in 1974 when Nye and Berry [55] performed a systematic study on vortices while studying ultrasonic wave pulses. In 1992 a study by Allen and colleagues opened up the possibility to generate optical vortices in free space [64], in while it was only in 2006 that Bliokh and colleagues proposed theoretically vortices in electron-optics [65,66].

Before we further delve in the properties of vortices it is necessary to remind what Angular Momentum is in both classical and quantum mechanics.

3.1 The Angular Momentum

The angular momentum in classical mechanics is defined as

$$\vec{L} = \vec{r} \times \vec{P} \tag{98}$$

And as we can see it orthogonal to the plane formed by the position vector (or operator) and the linear momentum vector (or operator).

In quantum mechanics further properties can be observed. In cartesian coordinate the angular momentum operator commutes with the free space Hamiltonian operator, thus:

$$\begin{cases} [H, L_x] = 0 \\ [H, L_y] = 0 \\ [H, L_z] = 0 \end{cases} \tag{99}$$

On the other hand the three components of the angular momentum do no commute with each other:

$$\begin{cases} [L_x, L_y] = i\hbar L_z \\ [L_y, L_z] = i\hbar L_x \\ [L_z, L_x] = i\hbar L_y \end{cases} \tag{100}$$

We can also introduce the L^2 operator, with

$$L^2 = L_x^2 + L_y^2 + L_z^2 \tag{101}$$

and it also commutes with the three components of $\vec{L}$, so that

$$\begin{cases} [L^2, L_x] = 0 \\ [L^2, L_y] = 0 \\ [L^2, L_z] = 0 \end{cases} \tag{102}$$

thus, it is possible to notice how H, L_i and L^2 are a set of commuting operators.

In absence of spin, we define the angular momentum in the position representation as:

$$\vec{L} = \frac{\hbar}{i}(\vec{r} \times \vec{\nabla}) \tag{103}$$

where $\vec{\nabla}$ is the vector differential operator.

Considering that the main propagation direction is the z direction and that distances along x and y are much smaller with respect to those along z (i.e., we are in the paraxial approximation), we can focus on the z-component of the angular moment and it can be expressed as:

$$\hat{\mathrm{L}}_{\mathrm{z}} = \frac{\hbar}{\mathrm{i}}\left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}} - \mathrm{y}\frac{\partial}{\partial \mathrm{x}}\right) \tag{104}$$

while in cylindrical coordinates (ρ, θ, z) it can be expressed as

$$\hat{L}_z = \frac{\hbar}{i}\frac{\partial}{\partial \theta} \tag{105}$$

The Laplacian operator in cylindrical coordinates can be written as:

$$\nabla^2 = \left[\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}\right] \tag{106}$$

So that the Hamiltonian operator in cylindrical coordinates can be written as:

$$\mathrm{H} = -\frac{\hbar^2}{2m}\nabla^2 = -\frac{\hbar^2}{2m}\left[\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2}\right] = -\frac{1}{2m}\left[\frac{\hbar^2}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) - \frac{{L_z}^2}{\rho^2} + \hbar^2\frac{\partial^2}{\partial z^2}\right] \tag{107}$$

and finally we can write the time independent Schrödinger equation in cylindrical coordinates as:

$$-\frac{1}{2m}\left[\frac{\hbar^2}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) - \frac{{L_z}^2}{\rho^2} + \hbar^2\frac{\partial^2}{\partial z^2}\right]\psi = E\psi \tag{108}$$

The vortex wavefunction is a solution of Eq. 108 for a free particle (such as an electron moving in free space) and it takes the form:

$$\psi_{\mathrm{V}}(\rho, \theta, \mathrm{z}) \propto \mathrm{A_V(r)exp[i(}\ell\theta + k_z z)] \tag{109}$$

where $\mathrm{A_V}(\rho)$ also contains the radial distribution of the wave and may very slowly change with z for diffracting beams. The angle-dependent phase term, $\ell\theta$, suggests the presence on an helicoidal wavefront. Figure 14 portrays in a) the three-dimensional representation of a wavefront associated with an Electron Vortex Beam (EVB) where the hue represents the probability distribution ($|\psi(r)|^2$), with yellow denoting the regions of highest transverse probability. In Figure 14b) it is possible to appreciate the semi-classical (Brohmian) electron trajectories of electrons propagating in free-space in a Laguerre-Gaussian mode (one the two most common solutions of the time independent Schrödinger equation, that will be discussed in section 3.2) [67].

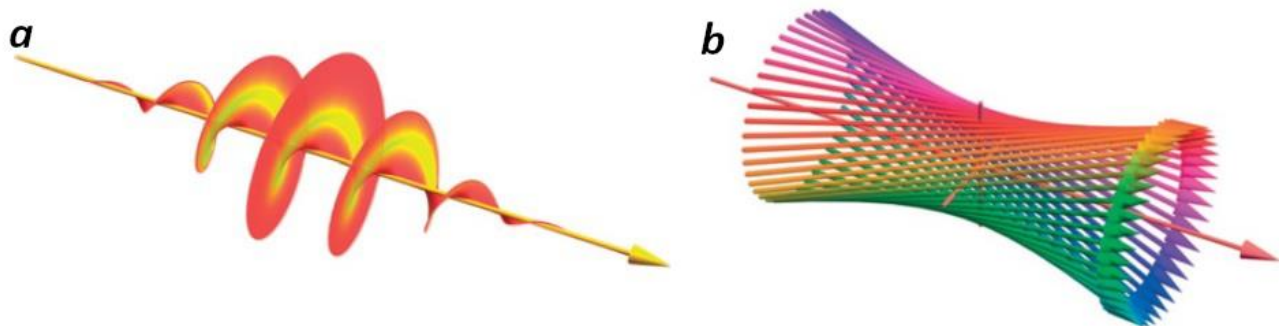

Figure 14: Three-dimensional representation of a) an equiphase surface of an EVB carrying OAM $\ell = 1$ where the hue from red to yellow indicates an increasing amplitude of the wavefunction, while b) shows the semi-classical trajectories of the electrons forming the EVB where the hue in this case is associated to the phase of each electron. Taken from [67].

Vortex beams, Eq. 109, are eigenmodes of the z-component of the quantum-mechanical Orbital Angular Momentum (OAM) operator, Eq.105, with eigenvalues $\hbar\ell$. This means that vortex beams carry a well-defined OAM $\hbar\ell$ per particle along their main propagation axes. This aspect can be proved by calculating the expectation value (Eq.11) of the OAM for this kind of wavefunction:

$$\langle \hat{L}_z \rangle = \frac{\langle \psi_V | \hat{L}_z | \psi_V \rangle}{\langle \psi_V | \psi_V \rangle} = \hbar\ell \tag{110}$$

Unlike photon, one further striking feature of EVBs that arises from the fact that electrons are charged particles is that they also feature a vortex induced magnetic moment $M = \ell\mu_B$, where μ_B is the Bohr magneton. Thus, they can be used for magnetic field sensing [68,69], which has prompted a greater interest the nowadays flourishing field of EMCD [70–75](the theory of this technique will be introduced in section 3.4). Furthermore, electrons can interact with other electromagnetic waves (photons), as well as radiate photons via the Vavilov-Cherenkov or other effects.

Furthermore, as was said below Eq.97 the probability current density around phase singularities swirls around. However, another property of vortex beams can be found by looking at the probability current density defined as

$$\vec{j} = \frac{\hbar}{m} Im(\psi^* \nabla \psi) \tag{111}$$

If we substitute to the generic wavefunction the vortex beam wavefunction reported in Eq.109, a non-zero azimuthal component of the current results and it is expressed as:

$$j_\ell(r) = \frac{\hbar}{m}\left(\frac{\ell}{r}\hat{\theta} - k_z\hat{z}\right)\gamma_\ell(r) \tag{112}$$

where $\hat{\theta}$ is the unit vector of the azimuthal coordinate, $\hat{z}$ is the unit vector along the z direction and $\gamma_\ell(r)$ is a factor proportion to the square of the radial distribution of the wave function, $\mathrm{A_V(r)}$. Equation 112 further implies that the vortex beam has a magnetic dipole associated to it.

To summarize, multiple important physical points have to be made about the OAM of vortex wave states:

I. In vortex beams the z-directed OAM carried is intrinsic (or more precisely it has an intrinsic component), as demonstrated in [76,77], which means that it is independent of the choice of coordinates. This is contrast to the angular momentum of classical point particle which is only extrinsic, since it depends on the choice of the coordinate origin. The only classical analogy is the angular momentum of an extended object.

II. The expectation value of the OAM in vortex beams is aligned with the mean momentum: $\langle \boldsymbol{L} \rangle = \ell \frac{\langle \boldsymbol{P} \rangle}{\langle P \rangle}$. Again, this is in contrast to what happens in classical mechanics where the angular momentum is orthogonal to the linear momentum.

III. Remarkably, focusing on electron vortex beams one may notice a resemblance between non-relativistic scalar electrons in a vortex beams and massless particles with spin ℓ. It is well known that the spin angular momentum (SAM) of massless relativistic particles is aligned with their momentum, so that helicity is a well-defined quantum number. As said in the previous bullet point vortex beams carry similar OAM with well-defined longitudinal component, which we may denote as orbital helicity. The SAM of electrons, however, is limited to $|\boldsymbol{S}| = \frac{\hbar}{2}$, while as demonstrated in Eq.110 the intrinsic OAM can take arbitrarily large values. Thus, the effect of spin is limited (moreover spinpolarized sources have only been proposed and never realized [78]) while the OAM of the electrons in the vortex beam affects their dynamics and their interactions with external fields, atoms and other particles.

IV. Both the intrinsic OAM and the swirling probability current density do not contradict the rectilinear propagation of either plane waves or of classical particles in free space. Also, while the centroid of a vortex state follows a rectilinear trajectory in free space, the trajectories of the particles are Bohmian (semiclassical) as shown in Figure 14 and they can be curvilinear.

V. The intrinsic OAM of a vortex state is not a collective effect, but it persists on the single-particle level [79].

3.2 Common vortex solutions of the Schrödinger equation in cylindrical coordinates

The cylindrical coordinates solution reported in Eq.109 is relatively generic, here I will describe two particular solutions that allow for a convenient analytical description and offer a good approximation to the electron states produced in electron microscopes.

3.2.1 Bessel Beams

The first relevant solution of the Schrödinger equation in cylindrical coordinates, Eq. 108, is known as Bessel beam and here the eigenmodes are monoenergetic beams which propagate along the z axis and are axially-symmetric. The generic Bessel beam wave function is given by:

$$\psi_\ell^B(\rho, \theta, z) = J_\ell(k_\rho \rho) e^{i\ell\theta}\, e^{i(k_z z)} \tag{113}$$

where ρ, θ, z are the cylindrical coordinates, J_ℓ is the ℓ-th order Bessel function of the first kind, ℓ is an integer number and the OAM topological charge. k_ρ and k_z are the transverse and the longitudinal components of the wave vector, respectively, and $k^2 = k_\rho^2 + k_z^2 = \frac{2m\omega}{\hbar} = (\frac{2\pi}{\lambda_{dB}})^2$, where m is the electron

mass, $\hbar$ is the reduced Planck constant and λ_{dB} is the electron's de Broglie wavelength (Eq.3). Eq.113 satisfies Eq. 108 if $E = \frac{\hbar^2}{2m}k^2$. It is necessary to point out that up to now all the calculations which led to Eq.113 were done in the non-relativistic limit (since the Schrödinger equation, Eq.108 , is non-relativistic), however, as stated from the beginning, the electrons inside a TEM are relativistic. A complete discussion regarding Bessel beams with a relativistic approach that starts from Dirac equation was first reported by Bliokh *et al.* [80] and later also explained in Chapter 4 of [81], with new developments on relativistic electron vortices in general in [82–84]. In the relativistic approach one finds that Bessel beams are vector beams (and not scalar as in the non-relativistic approach), they are eigenmodes of the total angular momentum operator and one can observe that the intrinsic spin-orbit interaction is significant. Nonetheless, it is possible to describe the behaviour of electron Bessel beam with a non-relativistic approach in the paraxial approximation, where the spin-orbit interaction becomes weak.

The Bessel wave function is a well-known non-normalizable solution of the Schrödinger equation of a free electron in cylindrical coordinates. Bessel beams were first mathematically modelled in optics by Durnin [85]. Experimentally, they were realized as photon quasi-Bessel beams, an approximation of the Bessel beams, which retained the same properties of the theoretical and complete ones over a finite distance [86]. Durnin and colleagues defined Bessel beams as beams "whose central maxima are remarkably resistant to the diffractive spreading commonly associated with all wave propagation" [87,88]. The zeroth order Bessel beam possesses the smallest central spot diameter and the longest depth of field compared to other ordinary types of beams [89]. In fact, the zero-order beam has no vortex and maximal probability density on the axis.

Bessel beams represent the simplest theoretical example of vortex beams. However, while the generic wavefunction decays as $\sim\frac{1}{\rho}$ for $\rho \to \infty$, it is not localized in the transverse dimension. In fact, the integral $\int_0^\infty \left|\psi_\ell^B\right|^2 \rho d\rho$ diverges and as previously mentioned the function cannot be normalized with respect to the transverse dimensions. This implies that theoretically an infinite number of particles or energy per unit z-length in the Bessel beam, which explains why only a good approximation of a Bessel beam, i.e., a quasi-Bessel beam, can be experimentally produced [86] [90]. Bessel beams can be considered as a coherent superposition of conical plane waves, or a set of plane waves propagating on a cone. Both the delocalization of Bessel modes and the absence of diffraction are a direct consequence of the fact that the wave vectors are distributed only azimuthally, while the radial component (k_ρ) and therefore k_z are fixed. Apart from being non-diffractive, Bessel beams are also "self-healing" so that (apart from an overall decrease in the intensity) they can recover their intensity profile. Lastly, from Eq.113 it is possible to notice that the probability density is both time and z independent and proportional to $J_\ell^2(k_\rho\rho)$.

3.2.2 Laguerre-Gaussian Beams

A clear disadvantage of Bessel beams is the fact they are delocalized. A possible way to overcame this and produce beams that are localized also along the transverse direction is consider a beam made as a superposition of multiple Bessel beams all with the same energy, but different wave numbers. This can be done by introducing a small uncertainty in the radial momentum component [91]. Another, and much simpler, way to generate transversely localized vortex beams is use Laguerre-Gaussian beams. Laguerre-Gaussian beams are also solutions of the paraxial Helmholtz equation in cylindrical coordinates. These kind of beams are derived by first applying the paraxial approximation to the Schrödinger equation in cylindrical coordinates (Eq.108) and then solving it. Owing to the paraxial approximation we can say that the electrons' momenta are mostly directed along the z-axis. Thus, in the first order approximation we have that $k_\rho/k \ll 1$, so that the Schrödinger equation can be simplified using the substitution $\frac{\partial^2}{\partial z^2} \approx -k^2 + 2ik\frac{\partial}{\partial z}$, which gives

$$
\begin{aligned}
&-\frac{1}{2m}\left[\frac{\hbar^2}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) - \frac{{L_z}^2}{r^2} + \hbar^2(-k^2 + 2ik\frac{\partial}{\partial z})\right]\psi = E\psi \\
&E\psi - \frac{1}{2m}\left[\frac{\hbar^2}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) - \frac{{L_z}^2}{r^2} + \hbar^2 2ik\frac{\partial}{\partial z})\right]\psi = E\psi \\
&-\frac{1}{2m}\left[\frac{\hbar^2}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) - \frac{{L_z}^2}{r^2} + \hbar^2 2ik\frac{\partial}{\partial z})\right]\psi = 0
\end{aligned}
\tag{114}
$$

The solutions to this are the Laguerre-Gaussian (LG) beams, whose wavefunction is defined as [92]:

$$
\begin{aligned}
\psi^p_{LG\,\ell}(\rho,\theta,z,t) = &\frac{C_{\ell p} z_R}{\sqrt{z_R^2 + z^2}}\left(\frac{\sqrt{2}\rho}{w(z)}\right)^{|\ell|} L_p^{|\ell|}\left(\frac{2\rho^2}{w^2(z)}\right) exp(i(k_z z + \ell\theta - \omega t)) \times \\
&\times exp\left(-\frac{\rho^2}{w^2(z)} + ik_z\frac{\rho^2}{2R(z)}\right) exp\big(-i(2p + |\ell| + 1)\xi(z)\big)
\end{aligned}
\tag{115}
$$

where p and ℓ are two discrete quantum numbers, and ℓ as before is the azimuthal index or topological charge for the OAM while p is the radial index which defines the $(p+1)$ radial nodes in the intensity distribution. $L_p^{|\ell|}$ is the generalized Laguerre polynomial [93], $C_{\ell p} = \sqrt{\frac{2^{|\ell|+1} p!}{(\pi(|\ell|+p)!)}}$ is a normalization factor, $w(z) = w_0\sqrt{1 + \left(\frac{z}{z_R}\right)^2}$ is the beam waist radius along the propagation axis z and in particular w_0 is the beam radius at focus, z_R is the Rayleigh range given by $z_R = \frac{k_z w_0^2}{2}$, $\xi(z) = ArcTan\left(\frac{z}{z_R}\right)$ and $R(z) = z\left[1 + \left(\frac{z_R}{z}\right)^2\right]$ is the complex wavefront's radius of curvature. It is worth noting that the characteristic transverse and longitudinal scales of the beam are w_0 and z_R, where

$$
z_R \gg w_0 \gg \frac{2\pi}{k} \tag{116}
$$

The last exponential term in Eq.115 is known as Gouy phase term [94–97]. The Gouy phase is closely related to the transverse confinement of the modes. Moreover, it is possible to demonstrate that the evolution along the optical axis is related only to the Gouy phase term and w_0, which makes this kind of Gaussian beams diffraction shape invariant: the beam evolves only by the scale factor $\sqrt{1+\left(\frac{z}{z_R}\right)^2}$. This is a weaker condition of the actual diffraction invariance of Bessel beams. On the other hand, LG beams are eigenstates of the Fourier transform operation and of the OAM, simultaneously, in this way they form a complete orthonormal basis. The Fourier spectrum of LG beams is smoothly distributed over different radial wave-vector components k_ρ, thus explaining why LG beams are properly localized and normalizable in the two transverse dimensions.

It is possible to notice that the phase shift induced by the Gouy phase term at any point along the z direction is given by

$$\Delta\varphi_G(z) = (2p + |\ell| + 1)ArcTan\left(\frac{z}{z_R}\right) \tag{117}$$

so that for a beam propagating through its focal point from $\frac{z}{z_R} \ll -1$ to $\frac{z}{z_R} \gg 1$ the Gouy phase term imparts to the beam a total phase shift equal to

$$\Delta\varphi_G = (2p + |\ell| + 1)\pi \tag{118}$$

A series of simulated LG beams with varying indexes are reported in Figure 15. The zeroth-order mode $\psi^0_{LG\,0}$ is the standard Gaussian beam, which can be regarded as an infinitely-long Gaussian wavepacket. Conversely, higher order modes exhibit a variety of structures related to the internal spatial degrees of freedom of localized electrons.

LG beams are of interest to scientists involved in both fields of magnetic materials and structured waves. For example, the LG wavefunction is functionally similar to the Landau states wavefunction, the eigenstates of

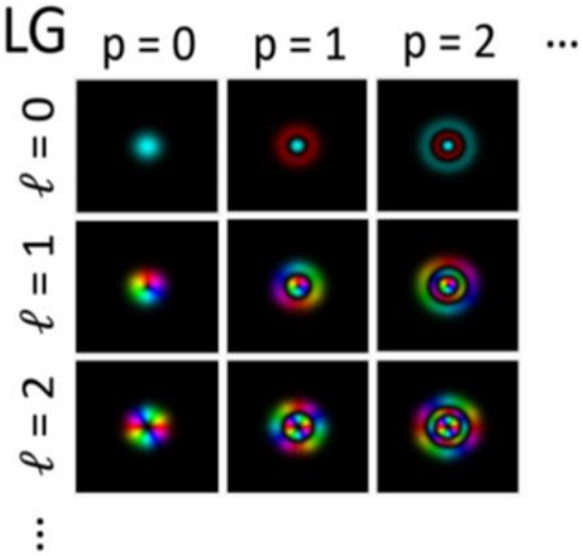

Figure 15: Examples of Laguerre-Gauss beams with varying indexes. The intensity and phase shift of the wavefunctions are represented by brightness and hue respectively.

the propagation of electron beams inside a constant magnetic field [98]. By properly tuning the LG beam waist it has been experimentally demonstrated that it is possible to couple them to Landau states [99]. Even though the generated LG beams were not the pure ones, this still was an experimental proof that opens up the possibility to observe a transition between the two states.

In conclusion, LG beams with different quantum numbers (p, ℓ) and Bessel beams with different ℓ are both complete set of orthogonal monoenergetic modes for the free-space Schrödinger equation in cylindrical coordinates (even though both are restricted to the paraxial approximation). Therefore, any free-electron state can be represented as a superposition of these modes. Vortex beams are then well suited when dealing with monoenergetic electrons with a well-defined propagation direction (like in the TEM) used to study problems with some sort of azimuthal symmetry.

3.3 Methods for Generating vortex beams and for measuring their OAM

Now that I have provided to the reader a basic understanding regarding vortex beams, it is important for me to report what are the main ways to both generate vortex beams and measure their OAM, focusing mainly on beams of accelerated electrons. Since Chapter 4 is meant to introduce the reader to beam shaping through nanopatterned apertures like amplitude plates or phase plates and in it I will provide a series of examples which are mainly focused on methods to generate vortex beams and measure them, for most of the methods that I will introduce in this section I will give a brief explanation and refer the reader to the corresponding sections in Chapter 4. It is important for me, however, to denote that a series of in-depth review on the topic of electron vortex beams have been written in recent years and those contain details that here may be omitted for sake of definitiveness [67,81,100–102].

3.3.1 Electron Vortex Beam generation

The first two methods that can be used to generate EVBs have been readapted from those used in photon optics. The first is the well-known Spiral Phase plate(SPP) [103–105], while the second is a diffraction grating with a fork dislocation [55–57]. The SPP idea was used by Uchida and Tonomura for the first experimental demonstration of the production of a free-electron vortex in a TEM [20]. The SPP works by taking advantage of Eq.89 ($\Delta\varphi = C_E V_{mip} t_{local}$) which implies that free-space electrons acquire and additional phase when they move through a material's region of thickness t_{local} and with mean inner potential V_{mip}. Thus, a region with spiralling varying thickness will create the corresponding spiral phase in the transmitted wave. Uchida and Tonomura employed a spiral-thickness-like region in a stack of graphite flakes to generate an EVB. However, in their experiment the phase change at the step was not an integer times 2π and as a result the output electron wave possessed a non-integer vortex [106,107], which can be seen as a superposition of several vortex states with different quantum numbers ℓ [108]. Few years later improved experimental

realizations of the spiral phase plate were reported in which an integer OAM was generated [109,110]. Greater details on the realization of the SPP for electrons can be found in section 4.3.1.1.

The diffraction grating with an edge dislocation (or pitchfork), on the other hand, owing to its edge dislocation is able to produce screw dislocations in diffracted beams when a basic Gaussian-like beam (a beam with a Gaussian-like intensity profile that can be easily obtained in a TEM by converging the beam with lenses or by circular diaphragms) impinges on it [111,112]. This works because the vortex structure represents a screw dislocation in the phase front of the wave field. Therefore, if the dislocation in the grating is of order ℓ_0, after the Gaussian-like beam is diffracted by it the n-th diffraction order will be a vortex beam (a quasi-LG one) with $\ell = n\ell_0$. This method was first used by Verbeeck and co-worker to generate high quality vortex beams with $\ell_0 = 1$ [21], and soon after by Mc Morran *et al.* for beams with $\ell_0 = 25$ [22]. The latter also showed that it was possible to generate beams with topological charges up to $\ell = 100$, that was naturally present in the $n = 4$ diffraction order. Few years later, it was thanks to this method that beams with intrinsic OAM of few hundreds and even thousands $\hbar$ were created [113]. Large OAM beams prove that spin angular momentum is not present in Vortex beams (as mention earlier) and are well suited for experiments on magnetic materials. Greater details on the realization of a diffraction grating with an edge dislocation are shown in section 4.3.1.2.

Figure 16 shows the schematic representation of how the SPP (a) and the fork-diffraction grating (b) are able to generate EVBs.

Another method that was first reported in photon optics for the generation of vortex beams is mode conversion [64]. This method allowed to generate for the first time free-space photon vortex beams. However, since (in my opinion) in it more suited to estimate the topological of a vortex beam, the description of this method is reported below in the section dedicated to techniques to measure the OAM of a beam.

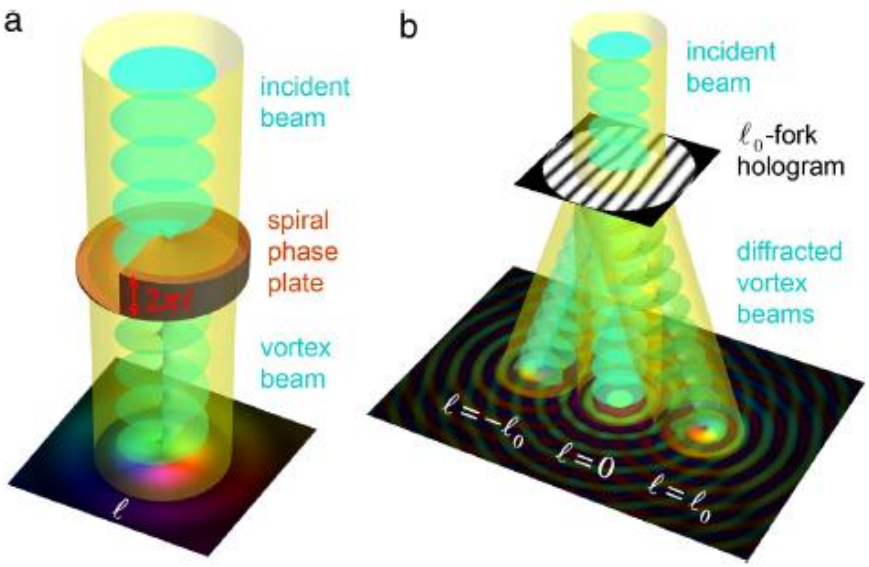

Figure 16: Schematics of basic methods for the generation of electron vortex beams analogous to those used in photon optic. a) shows a spiral phase plate, while b) a diffraction grating with a fork dislocation. Taken from [81]

There are other methods that can be used to generate EVBs inside a TEM, that however don't have a counterpart in optics. This is because these methods take advantage of the fact that electrons are charged particle and hence their phase can be influenced by either external electric field or external magnetic fields (as we already observed when we discussed the weak phase approximation and in particular Eq.77).

The first method uses a long and thin magnetic rod with a sharp end to generate an EVB [114–116]. In fact, this works because one can approximate one of end of the rod to a quasi-magnetic monopole [117,118], and during the scattering of an electron by a magnetic monopole a vortex beam is generated [66,119]. A magnetic monopole of strength α_m (where α_m is also the dimensionless magnetic-flux line strength of an equivalent magnetic flux line, and $\alpha_m = 1$ corresponds to two magnetic-flux quanta) can produce a vortex beam with $\ell = \alpha_m$ when an electron moves from a region without magnetic flux to a region with the flux. One can understand the working principle of this method by thinking in terms of angular-momentum conservation. From the classical point of view, the Lorentz force from the monopole bestows an additional OAM to the electron. The monopole magnetic field can be written as $\vec{B} = \left(\frac{\hbar c \alpha_m}{2e}\right)\frac{\vec{r}}{|\vec{r}|^3}$, and indeed the Lorentz force from the monopole is not central and it originates from the non-symmetric vector potential. However, it can be demonstrated [118] that when electron comes from $z \to -\infty$ and the scattered electron is observed at $z \to +\infty$, the electron OAM along the propagation direction must change from $L_z^{out-going} = L_z^{incoming} + \hbar\alpha_m$. From the point of view of quantum mechanics, using the eikonal approximation [36], a point-like electron approximately follows a straight line trajectory before passing at a radial distance $\vec{r}_0$ from the monopole. Then, the Lorentz force from the monopole deflects the electron, which gains a transverse (azimuthal) momentum component equal to $p_\theta = \frac{\hbar\alpha_m}{r_0}$. As a result of this, the electron also acquires an OAM equal to $L_z = p_\theta r_0 = \hbar\alpha_m$. This is valid as long as we consider the kinematic momentum operator (and OAM operator) instead of the canonical one, where $\hat{p}_{kin} = \hat{p} - \frac{e}{c}\boldsymbol{A}$ and $\hat{p} = -i\hbar\vec{\nabla}$ are the kinematical and canonical operators, respectively, and $\boldsymbol{A}$ is the vector potential. More in depth details on this method have also been reported in the review by Bliokh et al. [81]

The second method was recently developed and involves a pair of charged needles. This method was theoretically developed [120] and later realized [121] by Pozzi and Tavabi *et al.* The two charged needles are oppositely charged. The analytical expression of the electrostatic potential of a line of charges of constant charge density K that lies on the x, y-plane at $z = 0$ and is aligned to the y direction (has one end at point $(-a, 0)$ and the other at $(0,0)$, and is compensated by a neutralising charge in the same plane at (x_D, y_D) is given by [122]

$$V(x.y,0) = C_V\left[log\left(\frac{\sqrt{(a+y)^2+x^2}+a+y}{\sqrt{x^2+y^2}+y}\right) - \frac{a}{\sqrt{(x+x_D)^2+(y+y_D)^2}}\right] \quad (119)$$

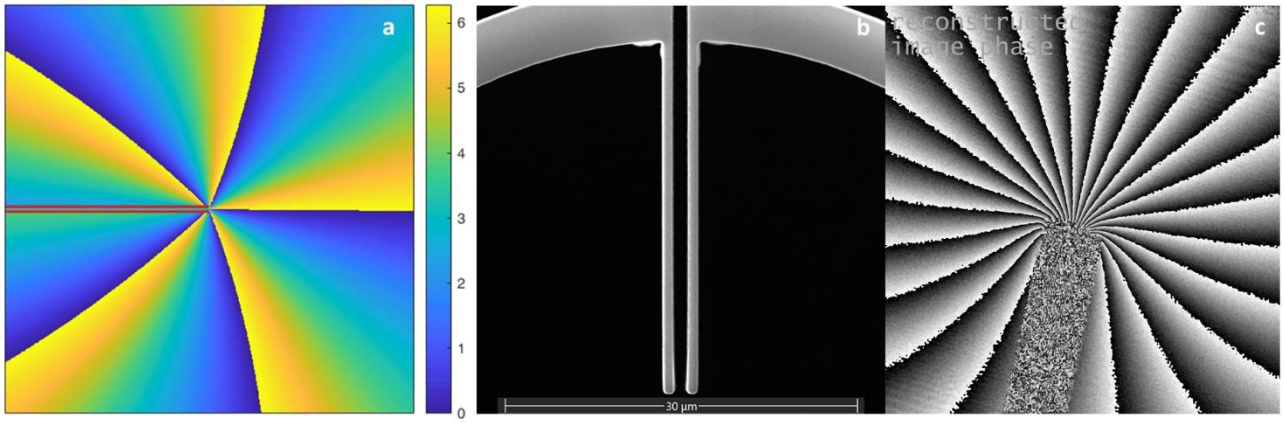

Figure 17: a) Bi-dimensional phase profile (Modulo 2π) obtained for two line charges of length $a = 200\mu m$ and $C_V = 1$ for 300kV electrons, calculated by starting from Eq.120 and Eq.121 with $\alpha = 0.2\mu m$. As one can desume from the colorbar, the darkest tone of blue corresponds to a phase equal to 0 while the brightest yellow to 2π. The two charged lines lie horizontally and have been enlarged and colored in red for improved visualization. In b) is shown a top view SEM image on device with improved design with respect to those used in Ref. [121]*, while c) reports the experimental phase profile obtained by applying $1V$ on one needle and $-1V$ on the other.*

where naturally the propagation direction is along the z-axis and $C_V = \frac{K}{4\pi\varepsilon_0}$. By inserting Eq.119 in Eq.77 we obtain that the associated electron-optical phase shift can be written as

$$\Delta\varphi(x,y) = C_E C_V[-(a+y)log((a+y)^2+x^2) + y\,log(x^2+y^2) + 2a + \\ +2x\,arctan\left(\frac{y}{x}\right) - 2x\,arctan\left(\frac{a+y}{x}\right) + alog((x+x_D)^2+(y+y_D)^2)\Big] \quad (120)$$

Because the electric potential is addictive, then one can obtain the phase attributed to two oppositely charged rods separated by a distance 2α:

$$\Delta\varphi_{TOT}(x,y) = \Delta\varphi(x-\alpha,y) - \Delta\varphi(x+\alpha,y) \quad (121)$$

In Figure 17a can be found the bidimensional phase profile, modulo 2π, generated by the two lines of charges with $C_V = 1$. The same results can be obtained by utilizing two inversely biased metallic thin needles. In Figure 17b is shown the top-view SEM image of a electrostatic phase plate that can be used to freely generate vortices (the design has been changed from that used in Ref. [121]), while in Figure 17c it is possible to observe the experimental phase profile obtained by polarizing the needles in Figure 17b with $\pm 1V$.

The key advantage of this approach over all the others is that it can be used to generate on-axis EVBs whose OAM can be continuously tuned by changing the charge density (in the case of charge lines) or the applied potential (in case of biased metallic needles). Furthermore, there is almost no loss of intensity, apart from the small fraction of the beam that is shielded by the charged wires (or needles) and their support. It is worth noting, however, that there are discrepancies from the ideal phase of a Vortex beam (as it can be noticed by Figure 17a and c where the lines at which the phase jumps from 2π to 0 aren't radially oriented, but have a slight curvature) which are caused by spurious charging of the needle/lines under the electron beam and the finite length of the wires. The first has the effect akin to that of a biprism, while the second affects the effective potential shape.

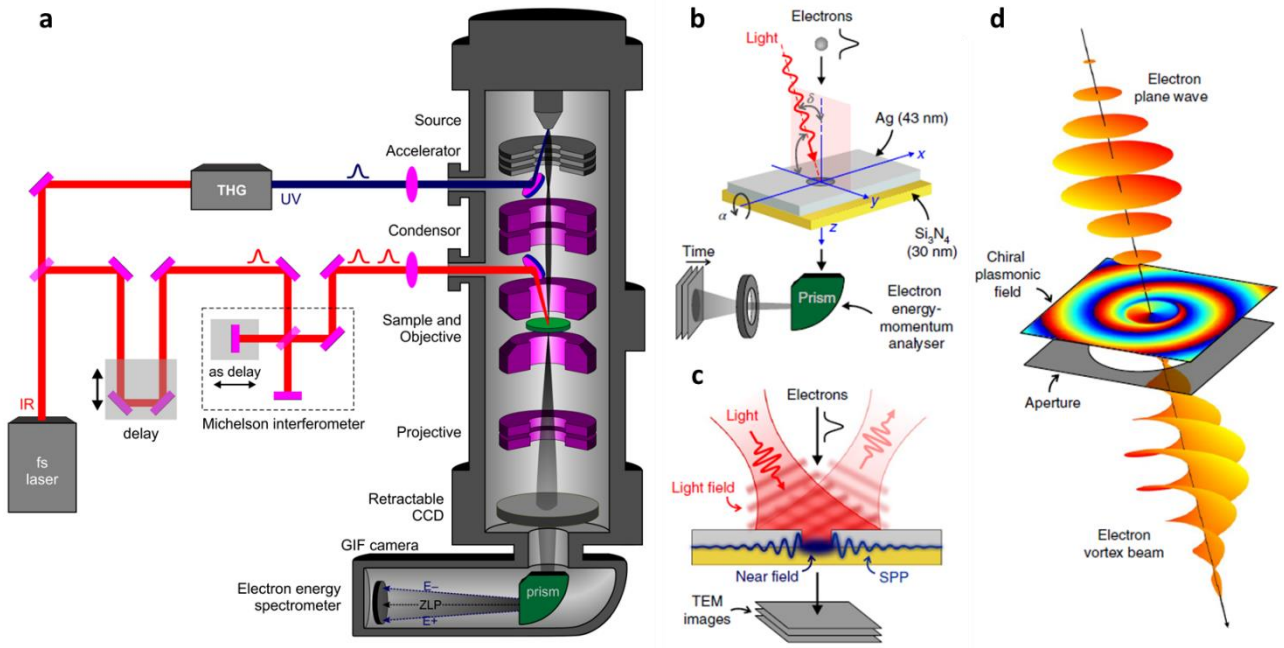

Figure 18: Generation of EVBs in an ultrafast TEM using a surface plasmon polariton near field. a) schematically reports the configuration of the ultrafast TEM used in [123] *to impart an OAM to an ultrashort single-electron wavepacket. b) shows the experimental geometry of the system used to demonstrate the validity of this approach. c) Schematic representation of the non-local holographic method used to image SPPs radially propagating at the Ag/Si_3N_4 interface away from the hole. d) schematic representation of the generation of an EVB after the ultrashort wavepacket interacted with the chiral plasmonic field. The images were all taken from* [123].

Lastly, another method that was recently developed uses optically excited femtosecond chiral plasmonic near fields to impart an azimuthally varying phase shift to ultrashort single-electron wavepackets [123]. Even though this method can be used only on ultrashort wavepackets, i.e., the wavepackets that are created in an ultrafast TEM [124–126], this approach is intrinsically scalable to smaller length scales (sub-wavelength regime) with respect to the other methods here reported. Furthermore, it still works in a situation where only a partial transverse coherence of the beam can be achieved. It is also possible to dynamically control the vortex beam properties thanks to the high degree of tunability of the optical field (and so the phase distribution of the plasmonic near field), which allows to coherently modulate both the longitudinal and transverse phase profile of the electron wavefunction. An interesting aspect of this method is that it is possible to effectively use it to impart an OAM on pulsed beams of heavier elementary of composite charged particles, where the transverse coherence is even stricter than in electrons.

Figure 18 can help to understand the working principles behind this method. In Figure 18a it possible to see how the ultrafast TEM was operated to demonstrate this method. A femtosecond laser emits optical pulses that thanks to a beam splitter are in part sent to the TEM source to generate the ultrashort electron wavepackets, while the rest is sent to the sample. Here the pulses that are sent to the sample are delayed via series of techniques, as shown, so that they are in phase with the pulse that reachs the electron-source. In Figure 18b and c it is possible to appreciate the geometry of the system used to impart an OAM and how the chiral plasmonic field was created. A 43nm thick layer of Ag was deposited on a thin 30nm membrane, $\mathrm{Si_3N_4}$, then a 0.8μm hole was bored, so that by shining the hole with either a circularly or partially circularly

polarized light (energy per photon $\hbar\omega = 1.57eV$) a surface plasmon polariton (SPP) with a chiral plasmonic field is generated at the hole edge and it radially propagates at the Ag/Si_3N_4 interface. Figure 18d intuitively depicts the generation of an EVB from the interaction of the ultrashort wavepacket with the chiral plasmonic field. A semi-analytical theory indicates that the inelastically scattered components of the wavefunction (in cylindrical coordinates) upon propagation through the plasmon-supporting structure can be written as [123]:

$$\psi_l(\rho,\theta,z;t) = \psi_{inc}(\rho,\theta,z;t)J_l(2|\beta|)\exp\left(il\arg(-\beta) + il\omega\left(z-\frac{v}{t}\right)\right) \quad (122)$$

where l corresponds to the number of plasmons absorbed/emitted by a single electron, ψ_{inc} is the incident electron wavefunction, v is the electron velocity, J_l is the l-th order Bessel function of the first kind and $\beta = \beta(\rho,\theta)$ is a complex function that describes the interaction of the electron with the light and SPP fields. Naturally, for greater details on the theory behind this method, the reader is referred to the article by Vanacore *et al.* [123] and its supplementary information.

3.3.2 OAM measurement

In addition to the ability to freely generate Vortex beams, it is necessary to also be able to measure the OAM of a generic beam. In fact, after interaction with a sample, especially magnetic or plasmonic ones, the electron beam, even when in a specific OAM state (plane wave or vortex), often ends up in a quantum state that is a superposition of many Vortex beams.

Like diffraction measures the momentum after a scattering event, we need a way to measure the OAM of a beam after it has interacted with a sample so that we can characterize the specimen.

It is worth mentioning, before I show some of the methods used to measure/estimate the OAM of a Vortex beam, that in real-life conditions due to aberrations or scattering by non-cylindrical-symmetric objects the pure cylindrical-symmetry of EVB might be lost. In this case, the direct proportionality between the topological charge ℓ at singularities and the OAM is lifted [127], thus the process of measuring the topological charge is not generically equivalent to measuring the OAM [128]. Ideally, one should be able to measure the azimuthal component of the current density j_θ(to which L_z is proportional to) independently from the radial component of the current density j_ρ and from the shape of the wave. However, all the methods developed so far fall short to satisfy these requirements, but they can still be used in most cases to estimate of the OAM.

The first method that was used to measure the topological charge employed projecting a vortex beam on a "fork" diffraction grating [129,130]. In fact, if the impinging beam is a vortex beam instead of a plane wave, the topological charge of the EVB that is generated on the n-th diffraction order is given by $\ell' = n\ell_0 + \ell$, where ℓ is the topological charge of the impinging beam and ℓ_0 is the dislocation order of the fork. It is easy to notice how in the case $n\ell_0 = -\ell$ the beam generated will be a non-vortex one ($\ell' = 0$) with maximum intensity on the center (instead of donut-like shaped intensity profile). This idea is called "phase flattening"

and is based on the fact that only a non-vortex state ($\ell' = 0$) can have intensity in the centre of the diffraction. Furthermore, when the beam is in a superposition of OAM states each component will contribute to some central intensity in the n-th order. This provides a way to attempt a direct decomposition of the beam. However, the result is strongly dependent on the radial extension of the beam which is not what a OAM measurement should be.

The topological charge of a vortex beam can also be determined by looking at the diffraction of a geometric aperture. In fact, the diffraction pattern of a generic geometrical aperture is altered by the unique phase profile of a EVB in such a way that the topological charge is revealed. As shown by Guzzinati and colleagues [130] a triangular aperture generates a triangular pattern of spots, where there are $|\ell| + 1$ spots per side and the orientation of the triangular patter depends on $Sign(\ell)$ (owing to the Gouy phase). Further details on how geometrical apertures can unveil further physical phenomena are reported in [131].

Another geometrical approach for the analysis of the OAM of a beam is offered by a multi-pinhole interferometer (MPI), as shown in [132]. The interferometer comprises of a set of small holes, and for OAM measurement the most efficient configuration is N pinholes evenly distributed along a circumference centered at the pole defining the OAM. Here, the diffraction pattern that is obtained when a EVB impinges on the MPI is determined by the relative phases between the pinholes, which allows to retrieve the value of ℓ. With this method it is possible to obtain a quantitative OAM decomposition of the beam, even in the case of EVBs with coherent or incoherent superposition of OAM states. In fact, by recording the diffraction pattern is possible to determine the autocorrelation function, $\mathcal{A}(\psi)$, of the interferometer. Since $\mathcal{A}(\psi)$ exhibits peaks at positions corresponding to the distances between similar objects, the peaks are strongly related to the azimuthal phase gradient sampled by the pinhole. Once the phase differences are measured, it is possible to obtain the OAM spectrum of the original beam by performing the OAM-harmonic decomposition of the diffraction [132]. There are, however, some drawbacks to this method. First, the pinholes need to be small enough to consider the phase inside each pinhole to be approximately constant, and they should be distant enough so that the peaks in $\mathcal{A}(\psi)$ do not overlap. This method is therefore highly inefficient in terms of intensity. Second, since the wave is "sampled" at just N positions, the MPI is not able to distinguish between vortex modes with topological charge ℓ and $\ell + N$, since they yield identical phase differences in the N pinhole positions.

The fourth method to measure the OAM of electron vortex beam is an indirect one and it exploits the beam's magnetic moment [133]. In particular, when an electron that carries a magnetic moment (like one in a vortex beam) passes through a hollow metallic cylinder, it induces eddy currents in the cylinder without altering the vortex state. From simulations it was estimated that the eddy currents generated are in the order of several picoamperes. This method has only been theoretically proposed, and never experimentally realized.

It is possible to measure the topological charge of the EVB or its OAM spectrum by directly manipulating the phase of the wave. The first example of this type of approach allows to measure the topological charge using a phase plate to introduce astigmatism in the beam (this can also effectively be done with the stigmators that are already present in the microscope). This method is based on the conversion between two modes of the electron beam in the microscope. The two modes are Laguerre-Gaussian, a vortex solution of the Schrödinger equation in cylindrical coordinates, and Hermite-Gaussian beam (another type of Gaussian beams) being a solution of the Schrödinger equation in cartesian coordinates with typical development along two orthogonal direction. It is possible to transform one mode into the other by using astigmatism. The general name for this approach is mode conversion, and it can be useful to characterize a LG vortex beams with trivial radial form. However, for more complex vortex states the method gives more ambiguous results. It was thanks to this approach that the first free-space photon Vortex beams were generated and measured by Allen *et al.* in 1992 [64]. These were obtained by using a cylindrical lens to transform an high-order LG mode into an high-order HG mode and backwards. In a TEM it is possible to obtain the same effect of a cylindrical lens in optics by increasing the astigmatism of a lens. This property has been exploited by Schattschneider *et al.* [134] as a mean to measure the OAM of an EVB. Mode conversion in TEMs can nowadays be performed in multiple ways and is generically referred as $^{\pi}/_{2}$ mode conversion. An in depth review on the topic that gives further details has been recently written [135]. In it is also shown a novel design, where by using some of the lenses present in the aberration corrector it is possible to obtain a $^{\pi}/_{2}$ mode converter. This approach, however, is limited in the event of a complex radial phase, such as in the case of a LG-beam with p different from zero.

Finally, the method of election of this work and probably the most quantitative si the OAM sorting. It involves directly manipulating the wavefunction by log-polar conformal-mapping. This method allows to directly measure the OAM spectrum of a EVB, but in general of any beam by employing a transverse-coordinate transformation that can convert a circular distribution into a straight line distribution. This means that, for example, an input image comprising concentric circles is then transformed into an output image of parallel lines. It also means that orbital angular momentum and its azimuthal gradient operator is transformed in a linear momentum whose spectrum can be easily measured by diffraction. This method was first demonstrated in photon optics by Berkhout *et al.* [136], with improvements by Mirhossein and co-workers [137], and readapted to study electron vortex beams by Grillo *et al.* [23] by means of a pair of holographic phase plates. The phase gradient introduced by these phase plate must be much larger than those of the beam itself, so that the evolution of the beam is completely controlled by these element to make the mapping. The results obtained in [23] and the theoretical calculation backing the possibility of the realization of an electrostatic OAM sorter reported by Mc Morran *et al*. [138] have prompted the results shown in this thesis. Key details on the holographic approach for an EVB OAM sorter are shown in Chapter 4

section 4.3.2, while the electrostatic approach is theoretically shown at the start of Chapter 5. We were recently able to experimentally realize it, as shown in the rest of Chapter 5 which references [139].

However, it is important to introduce the theory behind the ideal OAM sorter. As previously said this method maps each input circle onto an output line, and this gives the required deviation in ray direction and hence the phase profile of the transforming optical element. However, this phase distortion requires to be corrected by a second element. Therefore, the conformal-mapping system features two custom elements: one to transform the image (also known as "unwrapper") and one to correct the phase distortions (known as "corrector") that must be in the Fourier conjugate plane of the other. The first element performs a mapping from $(x, y) \rightarrow (u, v)$, being (x, y) and (u, v) the cartesian coordinates of the input plane and output plane, respectively. As before z is the propagation axis. It must be noted that this geometrical transformation (unwrapping) can be performed by a single element only if the mapping is conformal [140–142]. Here u and v are defined as [136]:

$$v = a\, arctan\left(\frac{y}{x}\right) \qquad (123)$$
$$u = -a\, ln\left(\frac{\sqrt{x^2 + y^2}}{b}\right)$$

The phase profile of the unwrapper element is given by

$$\varphi_U = \frac{2\pi a}{\lambda f}\left[y\, arctan\left(\frac{y}{x}\right) - x\, ln\left(\frac{\sqrt{x^2 + y^2}}{b}\right) + x\right] \qquad (124)$$

where λ is the wavelength of the incoming beam and f is the focal length of the Fourier-transforming lens between the two elements. The parameter a is a scaling terms with $a = \frac{d}{2\pi}$, where d is the length of the transformed beam. The parameter b translates the transformed beam in the u direction and can be chosen independently of a. The required phase correction can be calculated using the stationary phase approximation and is given by [136,142]:

$$\varphi_C = -\frac{2\pi ab}{\lambda f} exp\left(-\frac{u}{a}\right) cos\left(\frac{v}{a}\right) \qquad (125)$$

Figure 19 report the bidimensional phase profiles of the elements of the OAM Sorter.

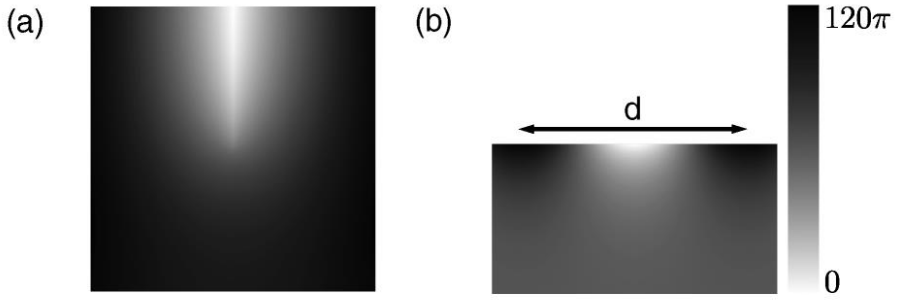

Figure 19: Phase profiles of (a) the transforming and (b) the phase-correcting optical element. d is the length of the transformed beam. In (b) only that part of the phase-correcting element is shown, that is illuminated by the transformed beam. Taken from [136].

This method for measuring the OAM of a beam (either of photons or of electrons) is very efficient since no part of the beam is discarded and is closest to an ideal OAM measurement since it decouples completely the radial and azimuthal degree of freedom. Furthermore, it is able to directly measure OAM spectrum of a generic beam with a near arbitrary superposition of OAM state (within the stationary phase approximation), i.e., a vortex beam with two or more topological charges. However, it suffers from some limits: first there is a significant modal cross-talk or OAM resolution that theoretically is already about 20% and in the experimental application is further enhanced both with photons [136,142] and electrons [23]; second, for both electrons and photons the use of holograms produces a loss of intensity, this being particularly relevant for the electron case where the transmitted intensity after the second element was less than 10% of the impinging intensity on the first. The first drawback refers to the sorting efficiency or sorting resolution. The sorting efficiency is expressed as a percentage and tells how much of the intensity that should be directed to the spot corresponding to a specific topological charge is measured in said spot. A 100% efficiency means no crosstalk so that the diffraction "spot" (I used the quotation mark in this case because the OAM spectrum is actually a series of elongated oval spots) is almost monodimensional along the OAM spectrum axis and doesn't influence the intensity of the nearby spots, while an efficiency of say 50% means that the "spot" is so broad that it becomes difficult to differentiate between two neighbouring spot. This can be compared to the concept of Rayleigh resolution (Eq.1) thus we can define the sorting resolution where $1\hbar$ in resolution means a sorting efficiency of approximately 77% (which coincides with the sorting efficiency of the device proposed by Berkhout). Three years after Berkhout's paper an improved version of the sorting device was proposed, which pushed the sorting efficiency to values greater than 92% [137]. We recently tried to create a similar system in electron-optics, the design and results that we obtained are reported in Section 4.3.3 of Chapter 4. Figure 20 shows how the Fan-Out approach is able to improve the sorting resolution. Figure 20a shows the intensity profile of the sorted beam at the focal plane, the beam is a superposition of two tilted plane waves obtained from the conformal transformation of two neighbouring OAM modes. This is what is typically obtained using Berkhout's OAM-sorter and it evident the overlapping of the intensity spots. Figure 20b shows that even if a magnifying lens is inserted after the corrector element to magnify the plane waves and narrow their correspond diffraction spots, however the spacing between the spots is also reduced accordingly so that the overlap (and sorting efficiency) remains unchanged. Indeed, what matters here, as in Rayleigh calculation, is that the phase ramps only once from 0 to 2π in the relevant intensity. Figure 20c shows that the tilting of multiple copies of the wavefront, is possible to increase the width of the transformed beam, while the angle of the tilt or the wavefronts across it remains constant. These converted modes can be then focused to a series of spots which have the same spacing as in Figure 20a, but narrower as in Figure 20b. This last point is clearly show in Figure 20d and e.

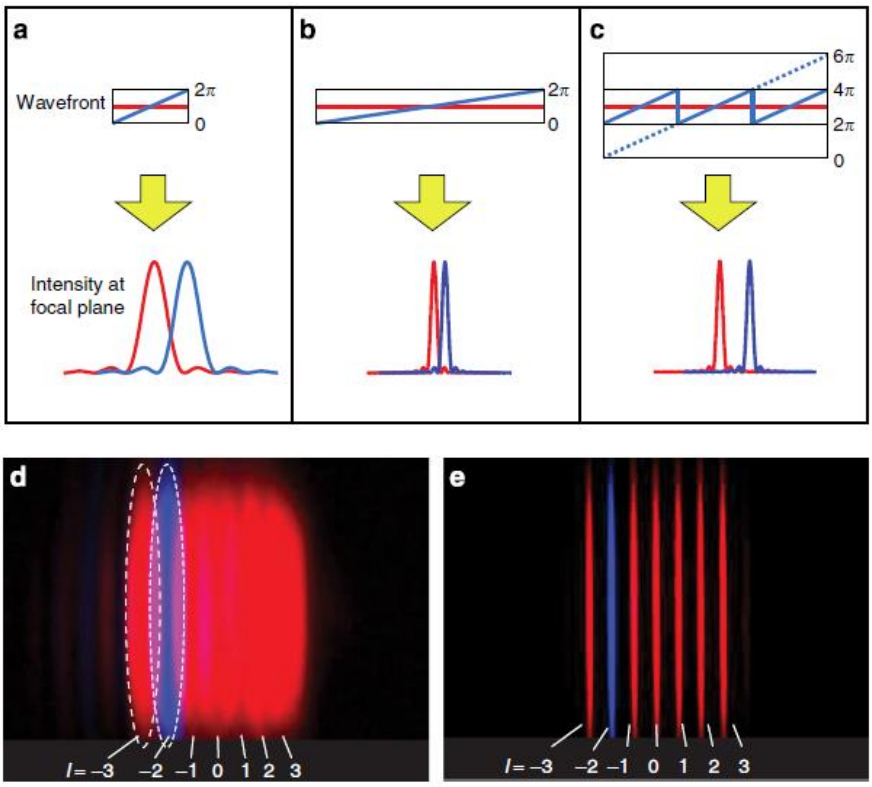

Figure 20: Fan-out approach introduced by Mirhosseini et al. [137] *to improve the sorting efficiency of the OAM sorter proposed by Berkhout et al.* [136]. Taken from [137].

The previously mentioned problem of loss of intensity, for electron beams at least, was due to absorption and other inelastic processes that happen when the electrons pass through the holograms. For details see Section 4.1.3.4 of Chapter 4. The loss of intensity can be almost completely avoided by using the electrostatic phase plate approach as proposed by Mc Morran and coworkers [138], which was later modified by Pozzi *et al.* [143], and as we recently demonstrated in [139] (chapter 5). We recently have also proposed theoretically a version based on magnetic elements [144](chapter 6).

3.4 Theory of Electron energy-loss Magnetic Circular Dichroism (EMCD)

As mentioned in section 3.1 Electron Vortex Beams (as well as other structured waves) have been used in Electron energy-loss Magnetic Circular Dichroism (EMCD) experiments since they allow to improve the spatial resolution, showing the promise for atomic-column resolution [145–148]. Recently, the research group that I'm part of proposed in series of scientific works methods on how an OAM sorter can be used to enhance EMCD measurements [72,75].

EMCD was first proposed and later realized by Schattschneider et al. [70,71]. The starting point of this technique are the similarities between X-Ray Absorption Near Edge Structures (XANES) and Energy Loss Near Edge Structures (ELNES) [149]. When one compares the double differential cross section for inelastic electron scattering to the cross section for the absorption of a photon of energy E one notices that the two are quite similar. In fact, in ELNES the expression of the double differential cross section for inelastic electron scattering

within first order time-independent perturbation theory and considering the dipole and single-particle approximation is given by the expression:

$$\frac{\partial^2\sigma}{\partial E\partial\Omega} = \sum_{i,f} \frac{4\gamma^2}{a_0^2 q^4}\frac{k_f}{k_i}\left|\langle f|\vec{q}\cdot\vec{R}|i\rangle\right|^2 \delta(E_i - E_f + E) \tag{126}$$

where E is the energy loss, Ω is the solid angle, $|i\rangle$ and $|f\rangle$ are the initial and final target states, $\vec{q} = \vec{k}_i - \vec{k}_f$ is the momentum transfer from the probe to the target in the interaction, a_0 is the Bohr radius, $\gamma = 1/\sqrt{1-\beta^2}$ is the relativistic factor, $\vec{R}$ is the position operator of the target electron. [it has to be pointed out that this is a semi-relativistic approximation valid for incident electron energies $\leq 200\ keV$ and for isotropic media].

On the other hand, in XANES in the dipole approximation the cross section for the absorption is written as [150]:

$$\sigma = \sum_{i,f} 4\pi^2\hbar\alpha\omega\left|\langle f|\vec{\varepsilon}\cdot\vec{R}|i\rangle\right|^2 \delta(E_i - E_f + E) \tag{127}$$

where $\alpha = e^2/(4\pi\epsilon_0\hbar c) = 1/137$ is the fine structure constant, E is the energy of the absorbed photon, $\vec{\varepsilon}$ is the polarisation vector of the absorbed photon. It is evident from the expression of Eq.126 and Eq.127 we see that within the dipole approximation the polarization vector $\vec{\varepsilon}$ in XANES is formally equivalent to $\frac{\vec{q}}{q}$. The correspondence between the two arises from the fact that the driving agent of the transition is an electric field. The oscillating field of the photon or of the closely passing electron acts directly onto the electrons of the absorbing atom and changes the charge distribution in direction of the field. For photon absorption the electric field is parallel to the polarisation vector, while for the inelastic electron scattering it is antiparallel to the momentum transfer. Furthermore, the selection rule states that the change in symmetry of the charge distribution forces a change of the angular momentum quantum number by ±1 since the dipole operator has odd parity.

In an absorption experiment with circular polarized light, such as in XMCD, the light polarisation vector is $\vec{\varepsilon}_{tot} = \vec{\varepsilon} + j\vec{\varepsilon}'$ (where j is the imaginary unit) with $\vec{\varepsilon} \perp \vec{\varepsilon}'$, which means that the two polarisation vectors are dephased of $\frac{\pi}{2}$. It is possible to recover a similar behaviour in electron inelastic scattering if the total momentum transfer is $\vec{q}_{tot} = \vec{q} + j\vec{q}'$ with $\vec{q} \perp \vec{q}'$ and $|\vec{q}| = |\vec{q}'|$. This condition on the momentum transfer can be commonly achieved in the two-beam diffraction geometry, i.e., when the sample is tilted by a few degrees from zone axis to excite specific reciprocal lattice vectors. By replacing in Eq.126 $\vec{q}$ with $\vec{q} + j\vec{q}'$, then it is possible to rewrite the double differential cross section for inelastic electron scattering as

$$\frac{\partial^2\sigma}{\partial E\partial\Omega} = \frac{4\gamma^2}{a_0^2 q^4}\sum_{i,f}\frac{k_f}{k_i}\left[\left|\langle f|\vec{q}\cdot\vec{R}|i\rangle\right|^2 + \left|\langle f|\vec{q}'\cdot\vec{R}|i\rangle\right|^2 - \langle f|\vec{q}\cdot\vec{R}|i\rangle\langle i|j\vec{q}'\cdot\vec{R}|f\rangle + \right.$$
$$\left. + \langle f|j\vec{q}'\cdot\vec{R}|i\rangle\langle i|\vec{q}\cdot\vec{R}|f\rangle\right]\delta(E_i - E_f + E) \quad (128)$$

Here, the first two terms in the square brackets are called Dynamic Form Factors (DFFs) and are commonly denoted as $S(\vec{q}, E)$. They describe inelastic scattering of an incident plane wave into an outgoing plane wave with wave vector transfer $\vec{q}$ and energy loss E. The DFFs appear in angle resolved EELS and in anisotropic materials are responsible for the direction dependence of the spectra. The last two terms in the square brackets are inelastic interference terms, known as Mixed Dynamic Form Factors(MDFFs) [151] and are denoted as $S(\vec{q}, \vec{q}', E)$. The MDFFs collapse into the DFFs when $\vec{q} = \vec{q}'$, due to symmetry relations it is always valid that $S(\vec{q}, \vec{q}', E) = S^*(\vec{q}', \vec{q}, E)$ and when both time inversion and spatial inversion symmetry are present it is valid that

$$S(\vec{q}, \vec{q}', E) = S(-\vec{q}, -\vec{q}', E) = S(\vec{q}', \vec{q}, E) = S(-\vec{q}', -\vec{q}, E) = S^*(\vec{q}', \vec{q}, E) = S^*(\vec{q}, \vec{q}', E) \quad (129)$$

which means that all the elements are real, thus no chiral signal is present. However, in presence of magnetic moments the time inversion symmetry is broken, which means that only spatial inversion symmetry is valid, so that

$$S(\vec{q}, \vec{q}', E) = S(-\vec{q}, -\vec{q}', E) = S^*(\vec{q}', \vec{q}, E) \quad (130)$$

and indeed the chiral signal is closely related to the imaginary part of the MDFF. It is then possible to write Eq.128 as

$$\frac{\partial^2\sigma}{\partial E\partial\Omega} = \frac{4\gamma^2}{a_0^2 q^4}\frac{k_f}{k_i}(S(\vec{q}, E) + S(\vec{q}', E) + 2Im[S(\vec{q}, \vec{q}', E)]) \quad (131)$$

For the first experimental realisation of this technique Schattschneider *et al.* [71] collected the EELS spectra at two different diffraction angles (corresponding to spot "+" and "-" in the diffraction plane which are shown in Figure 21) of an Iron (Fe) sample and compared the dichroism of the two spectra collected to the XMCD

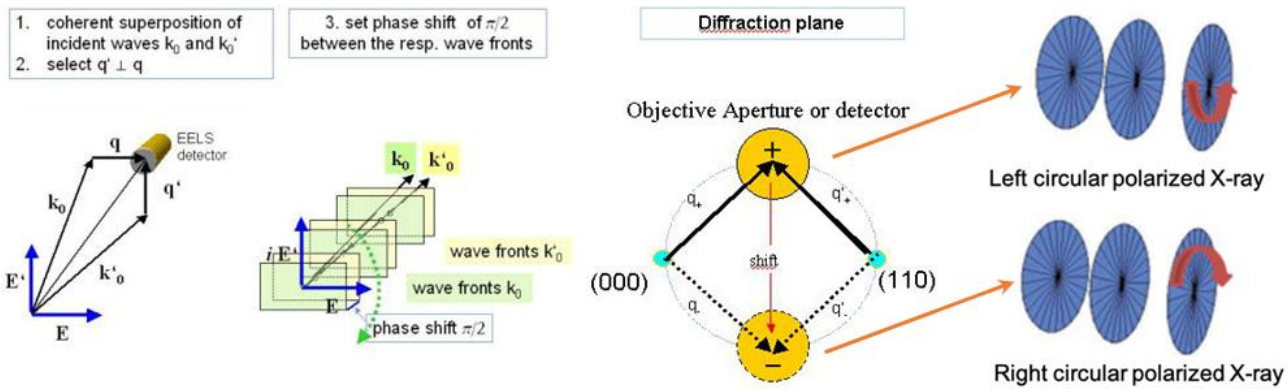

Figure 21: Schematic representation showing on the left the three steps necessary to reach the conditions to measure the chiral dichroism: 1. Reach the two beam condition; 2. Position the spectrometer entrance aperture such that in the reciprocal space is is at the same distance from the two main diffraction spots (such that $\vec{q} \perp \vec{q}'$ and $|\vec{q}| = |\vec{q}'|$). On the right are shown the two positions in the diffraction plane at which the detector aperture is positioned to collect the EELS signal, that can be compared to collecting the XANES spectra with two circularly polarized X-Rays. Figures taken from [71].

spectra and spectra dichroism collected on the same sample at a synchrotron facility. The results obtained with the two techniques were comparable. However, the dichroic signal of the TEM experiment was smaller than predicted and this was mainly due to the fact that the majority of the signal collected in the EELS spectra was non-dichroic Bragg scattering and thermal diffuse scattering. Furthermore, the integration of the signal over the finite spectrometer entrance aperture and the non-zero convergence angle reduces the dichroic effect. Another prominent effect is the sensitivity of the signal to the specimen thickness. A thickness variation of ± 1nm within the sampling area changes the dichroic signal by another 20%.

As mentioned at the beginning of this section, the use of EVBs in EMCD, mostly atomic-sized one [116,152], offer the possibility to reach atomic column size resolution. In principle, atomic-sized EVBs can be focus onto a single atomic column and induce atomic excitations, with different intensities for transitions in which the magnetic quantum number changes by $\Delta m = \pm 1$ [153]. For example, in magnetic transition metals differences in energy-resolved diffraction patterns are expected for $2p \rightarrow 3d$ transitions while using vortices with opposite OAM of $\pm\hbar$. This is because of the different populations of spin-up and spin-down $3d$ electronic states [68,154,155]. The use of atomic-sized EVBs can solve some of the issues of classical EMCD as they can be performed in zone axis conditions, which allows to increase the achievable spatial resolution [147,156,157]. However, this approach is difficult for three reasons:

I. It requires two measurements using EVBs with opposite OAM (usually $\ell = \pm 1$) in the same region, which cannot be easily performed due to sample and probe drift.
II. An EVB does not conserve its OAM while propagating in a crystal since the free-space cylindrical symmetry is broken by the crystal potential [158–160]. During propagation the electron probe loses its well-defined OAM acquiring different OAM components, thus reducing the intensity of the $\pm\hbar$ component and most importantly the dichroic signal [147].
III. It is experimentally challenging to prepare a high-quality coherent atomic sized EVB [152].

These limitations can be solved by using a standard electron probe and by performing a post-interaction analysis in terms of OAM on the inelastically scattered electrons [161,162]. Accordingly, an approach based on post-interaction analysis with an OAM sorter was theoretically proposed [72]. This approach has the advantage that it allows to record simultaneously the electron's OAM and energy loss spectra, where the two dispersion are orthogonal to each other. Moreover, the sample can be oriented along a high symmetry direction (zone-axis), allowing for atomic resolution. Lastly, the majority of the inelastic signal is recorded in this approach and the main limitation is the sorting efficiency, which should be near or even lower than the theoretical limit.

Original Research Works

Chapter 4

Forewords

The work here presented is meant to be a tutorial to introduce and guide research groups in learning what synthetic holography is, how to shape the beam (in this case the electron beam) with synthetic computer generated holograms and we also describe a series of examples mainly related to Vortex beams. For this work I wrote most of the manuscript, and checked and reported the theory behind synthetic holograms in such a way that it should be easily understood from even a freshly graduated student. Due to the broad spectrum of its contents and given that it still hasn't entered the publication phase, I choose to slightly modify its structure and move part of its text to other chapters to better adapt it to this thesis flow. The title of this manuscript will be:

"Theoretical and practical aspects in the generation of electron holograms"

Production of Holograms for Transmission Electron Microscopy: A tutorial on the design and preparation.

Introduction

The electron microscope has been primarily conceived to study matter at the highest lateral resolution but, with time, the quantum wave nature of the electrons has attracted the interest of scientists for both fundamental and applicative aspects. It is the wave nature of the electrons that allows us to draw an analogy with optics. In the limit of non-relativistic and monochromatic electrons, the Helmholtz equation equally well governs electrons and photons. Moreover, concepts like the refraction index and lenses can be transported from one context to the other and obtain similar results. In the case of light, refraction is an effect of the interference of the initial photon and those reemitted by atoms in the material. In contrast, for electrons the electrostatic and magnetostatic actions cause retardation (or anticipation) of the electron wave. Considering the similarities between the two research fields, the analogy is striking and has paved the road to the concept of electron microscopy. However, optics has been a flourishing field with a broad range of applications

beyond imaging in a microscope, and most of this progress (e.g., STED and superoscillation microscopy) has been triggered by the idea of structured light waves. This idea conceives the possibility for the wavefront and spatial intensity distribution of light to be controlled in an arbitrary way beyond what ordinarily offered by conventional optical elements. Recently, the concept of structured waves has been extended to the realm of matter waves and primarily to electrons. Structured electron waves include electron beams possessing helical wavefront, namely electron vortex beams, self-accelerating beams, non-diffracting beams, but also orbital angular momentum analyzers just to name a few. Moreover, one can fabricate conventional electron optics devices, i.e. lenses, diffractive elements and aberration correctors, by means of holographic approach. The key to gain so much freedom with electrons is the use of synthetic holograms as a tool to modulate the phase and amplitude of the electron wave. The term hologram, in general, comes from the Greek words for "whole writing", namely the ability to write both the intensity and phase of the wave. This is conceptually accomplished as a form of ideal interference of two waves since the interference pattern is related to the relative phase of the two waves.

Even if the original idea from Gabor was meant for electron optics [15], it is in light optics that holograms have seen wider applications, becoming a ubiquitous concept even for example in Canadian dollar notes. In electron microscopy, the nomenclature electron hologram is more often referred to the act of physically recording the interference of a wave perturbed by a semitransparent specimen and a plane wave. Conversely, it is of interest in this context the recreation of the perturbed wave starting from a calculated interference pattern and therefore, for the sake of clarity, this operation will be referred to as "*synthetic holography*", while the calculated patterns as "*computer-generated holograms*". The two operations are somehow inverse; when the fabricated interference pattern is illuminated, an electron beam that has the phase and amplitude of the semitransparent specimen is generated. This is the technique which was historically implemented when limited computing resources did not allow the researchers to mathematically apply numerical Fourier transforms. In this method, the recorded electron micrographs of the interference were illuminated with a laser to create the image of the object.

Furthermore, in order to physically construct the synthetic hologram, one needs to scale down the equivalent of the transparent electron micrograph at the electron's scale. Unfortunately, there is no electron optics analog of a completely transparent object. The best approximation is given by a thin layer of material with low electron absorption. Low Z materials like Carbon and Silicon Nitride (Si_3N_4) can be typically used to accomplish this step. Si_3N_4 can be produced routinely in very small membranes and inserted along the electron path, and standard techniques of nanofabrication allow to imprint thickness modulations on the lateral and depth scales of 50-100 nm. The positive inner potential of the substrate acts to accelerate the electrons producing an effective refraction index variation of $\Delta n/n = 10^{-5}$. For a somehow fortunate coincidence, this small refraction index variation over a typical modulation of 50 nm is sufficient to advance the electron wave of 300 keV by exactly one wavelength. Conversely, 100 nm of gold is enough to damp most

of the electron beam intensity. Therefore, all the needed tools to modulate both the phase and amplitude of an electron wave are nowadays available. One can notice the historic development from the early days of rough amplitude modulation of the wave to the fine and precise control over the amplitude and phase modulation and complex patterns that we can achieve today. This chapter provides a comprehensive overview on the theoretical and numerical calculations, fabrication, as well as analyzing the fabricated electron's holograms.

4.1 - Computer-Generated Holograms

4.1.1 Synthetic Holograms generation

Since the 60s, with the advancements in the computational power and image processing capabilities of computers, new types of holograms have emerged, the so-called *Computer-Generated Holograms* (CGHs). As *Lesem et al.* [163] stated back in 1968 when referring to holograms for 3-D display:

> "*A properly illuminated hologram forms for the viewer a picture which is identical with that which he would observe if he were looking at the scene himself. A computer generated hologram yields such a 3-D picture, without the original scene ever having to exist*".

From this simple explanation, it is possible to understand that CGHs allow to design and test the desired patterns with no need to create models for each re-iteration, and hence considerably reducing the total time required to make a fully working Synthetic CGH (S-CGH). The term synthetic has been used to underline that the last step involves producing a hologram which will be inserted in either a microscope or an optical bench. Figure 22 presents an intuitive representation of the two concepts (or kinds) of holography: the "conventional" image-holography (Figure 22a) and synthetic holography (Figure 22b), where, in this case, by sending a plane wave on the S-CGH the desired wave function is obtained.

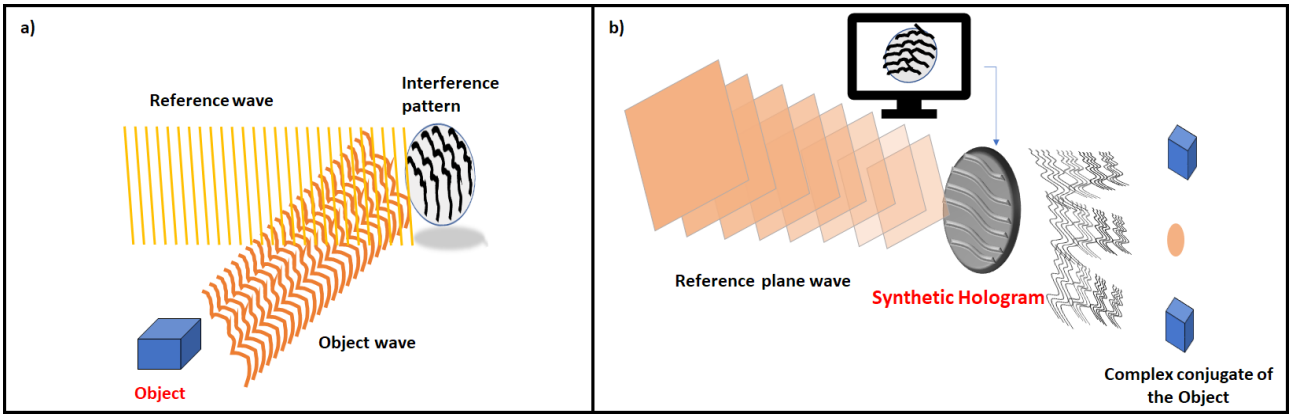

Figure 22: Infographic comparing traditional "image" holography, a), to synthetic holography,b).

The generation of the desired and reference wavefunctions, as well as the interferometric process, can be carried out using computer software properly written and developed for the task. In the research group that I'm part of for most of our works we rely on a modified version of *Stem_Cell* software [164], which allows us to construct the bi-dimensional phase and amplitude profiles of the desired wavefronts. In this software, the interferometric process, which in conventional holography consists of the overlapping of the two wavefunctions at a given plane, is carried out computationally. The CGH (i.e., the interference intensity pattern) generated by this set of operations is then exported to a file, which is later used to fabricate the S-CGH.

The typical longitudinal dimension of a S-CGH ranges from a few micrometers up to hundreds of micrometers. The smallest feature can go down to a few tens of nanometers. This means that the fabrication process

requires modern state-of-the-art machines and very well developed processes. These areas will be described in detail in the next sections. Moreover, a S-CGH works correctly and shows the desirable function when it is fully illuminated by the incident (reference) beam. The manufactured S-CGH, with a proper size, is mounted in one of the condenser apertures planes of an electron microscope. However, when the S-CGH is used, or during the testing, the imaging technique requires to switch-off the objective lens. This is necessary because the typical periodicity in the order of 100 nm implies a typical angle of 1 µrad. Hence, a very long focal length objective is required to image such small diffraction angles, which is usually not possible when the main imaging lens, the objective lens, of an electron microscope is operationally excited. Thus, the objective lens needs to be switched off. This is the case for the most commonly known Low Mag Mode of the TEM. However, this is not the only way to use the synthetic hologram. For example, when the synthetic hologram is positioned in the condenser system forming a focused probe, the diffracted beams are directly visible in the specimen plane. The practical application of holograms is a completely different topic that will not be investigated here in further details. When observed in Low Angle Diffraction (LAD) mode inside a TEM the S-CGH generates the wavefunction of interest. This is because S-CGHs act as *transmission diffraction gratings*: they split and diffract the beam into several other beams, each centered in a different position in the diffraction plane. A simple recipe for producing the hologram would be therefore to take the formula of the "image hologram" and invert it. By illuminating a piece of material that introduces the same amplitude (or intensity) modulation as in Eq.84 , it is possible to obtain an object where one of the diffractions would be the wave as it passed through the sample.

Synthetic holograms are more versatile as typically the interaction with the hologram can produce both phase and amplitude modulations. Depending on the type of interaction with the beam, holograms can generally be divided in three macro categories: (i) amplitude holograms; (ii) phase holograms; (iii) mixed (amplitude-phase) holograms. Moreover, we will see that depending on the hologram structure they can either encode (i) only the phase or (ii) both phase and amplitude of the interest wave. It must be stressed here that the character of the hologram (phase, amplitude or both) and the type of encoding are quite independent. A phase hologram, for example, can be made to encode both amplitude and phase of the interest wave, but with some restriction of the efficiency.

With respect to an incident plane wave, the transmittance function of a (amplitude-phase) S-CGH can be defined as:

$$T_H(\vec{\rho}) = A(\vec{\rho})e^{i\Delta\varphi(\vec{\rho})}, \tag{132}$$

where $\vec{\rho}$ is the transverse spatial coordinate with respect to the propagation direction of the beam, $A(\vec{\rho})$ and $\Delta\varphi(\vec{\rho})$ are the amplitude and phase modulations. The difference among the three types of holograms resides in the modulation that the hologram imposes on the incident wave. A S-CGH is defined as amplitude hologram when it modifies the amplitude of the incident wave by modulating $A(\vec{\rho})$ and the phase is kept unvaried, so $\Delta\varphi(\vec{\rho}) = constant$ [111]; on the other hand a hologram is defined as a phase hologram when

it modifies only the phase of the electron beam by modulating $\Delta\varphi(\vec{\rho})$, so $A(\vec{\rho}) = constant$ [112]. A mixed hologram, as the name might suggest, modifies both $A(\vec{\rho})$ and $\Delta\varphi(\vec{\rho})$ [165]. It is worth noting that a "phase" S-CGH will always have an absorption effect proportional to the material's thickness and its chemical composition. Even a pure-phase S-CGH will have an effect, however small, on the wavefunction's amplitude. An alternative way for pure-phase modification in the electron wave is to substitute material based holograms with structured electric or magnetic fields which can desirably modulate the phase. In this case, it is more challenging to design complex and arbitrary phase shifts, an example of this is reported in Ref. [139] (Chapter 5).

4.1.2 Different Types of Holograms

4.1.2.1 Amplitude holograms

In light optics, binary holograms (characterized by a local transmittance that is 0 or 1 only) are produced from partially transparent elements, e.g. gratings made of metals or any substrates that can block the light beam in some regions in the transverse plane. They are considered as the simplest type of S-CGH that can be fabricated. Amplitude modulation in the transmission is usually achieved by covering parts of the beam either with a material that can prevent the light from passing through (an opaque material that absorbs the beam), by deflecting the beam at a high-angle or reflecting a part of the incident beam.

In electron optics, the amplitude modulation is mainly due to the combined effect of inelastic scattering and high-angle scattering that happens when the electrons pass through the opaque hologram material. These scattering processes are usually stronger for heavy materials and thick substrates. The blocking of electrons is usually achieved using a thick sputtered layer of a high-Z element, such as gold or platinum. Doing so the wavefront amplitude is locally either fully preserved or completely blocked by the hologram. Since the hologram is absorbing/scattering the electrons, its action is non unitary, and thus the overall intensity is reduced by a factor that is proportional to the blocked area with respect to the incident beam cross-section. The amplitude holograms are, by definition, blocking part of the electron beam and therefore have limited efficiency. Since the absorption modulation is an amplitude-dominated effect, these kinds of holograms produce diffraction that is symmetric between positive and negative orders. Moreover, it is impossible to concentrate the intensity on a single diffraction spot and a large part of the intensity is directed to the 0^{th} order transmitted beam.

To gauge the absorption of a material, a useful whereas not exhaustive parameters, is the mean free path of plasmon excitation. The mean free path for 200 keV electrons of a few materials is reported in *Table 2*.

Material	Au	Ag	Pt	Si_3N_4	SiO_2	Al_2O_3	a-C
Theoretical (nm) [166]	76.148	88.3	76.43	135.33	133.62	135.67	106
Experimental (nm) [167]	84 (120)	100 (125)	82 (120)		155	140	160

Table 2: Theoretical and experimental mean free path of 200 keV electrons in different materials, The experimental values report the total inelastic mean free path, comprehensive of also single electron excitations like inner-shell ionization edges, while the terms in parentheses represent just the collective valence electron component or plasmon contribution [166].

It must be further noticed that the construction of pure amplitude holograms, where absorptive material is alternated to vacuum, is quite complicated when scaling down to small sizes because of the high probability that the long and thin parts of the hologram may collapse or join together during the fabrication process or under the beam illumination.

4.1.2.2 Phase holograms

In optics, one way to implement phase modulation is by etching grooves of the desired structure on a (transparent or reflective) surface, making the optical path inside (or upon reflection from) the material varying from one ray to another, locally changing the phase of the outcoming wavefunction. In electron optics, phase modulation is achieved by exploiting the relationship between the scalar and vector potentials. The phase shift of the electron wave function [36], is given by,

$$\Delta\varphi(\vec{\rho}) = C_E \int_{-\infty}^{+\infty} V(\vec{\rho}, z)dz - \frac{e}{\hbar}\int_{-\infty}^{+\infty} A_z(\vec{\rho}, z)dz \tag{133}$$

where

$$C_E = \frac{2\pi}{\lambda}\frac{e}{E}\frac{E_0 + E}{2E_0 + E} \tag{134}$$

here, $V(\vec{\rho}, z)$ and $A_z(\vec{\rho}, z)$ are the scalar potential and the z-component of the vector potential respectively, e the absolute value of the electron charge, $\hbar = h/2\pi$ is the reduced Planck constant, λ is the beam wavelength, E_0 is the electron energy at rest and E is the energy of the moving electrons. Typical electron energy values in a TEM are $200\,keV$ or $300\,keV$, yielding the corresponding C_E values $C_{E\_200keV} = 7.3 \times 10^{-3}\frac{rad}{V\cdot nm}$ and $C_{E\_300keV} = 6.6 \times 10^{-3}\frac{rad}{V\cdot nm}$.

In non-magnetic materials, only the scalar potential contributes to the phase shift and it can be approximated by using the mean inner potential, V_{mip}. In fact, V_{mip} produces a local acceleration on the traveling electrons for the whole thickness of the hologram [28], modifying the electro-optical path.

The phase variation due to V_{mip} and the local thickness $t(\vec{\rho})$ is given by a modified version of Eq.133:

$$\Delta\varphi(\vec{\rho}) = C_E \int_0^{t(\vec{\rho})} V_{mip} dz = C_E V_{mip}\, t(\vec{\rho}) \tag{135}$$

Another important parameter in Eq.135 is the magnitude of the material-dependent quantity V_{mip}. In choosing the material for the synthetic hologram, one must take into account the robustness, electrical conductivity and V_{mip}. Most of the phase S-CGH are now made using Si_3N_4. The choice in this specific material resides in the fact that amorphous Si_3N_4 can be considered a nearly pure-phase mask since it is almost transparent to an incoming electron beam. S-CGH are then obtained by "carving" grooves on top of a free-standing Si_3N_4 membrane. The calculated V_{mip} of a Si_3N_4 membrane was estimated by *Grillo et al.* [168] to be $\approx 15\,V$. The same value has been reported also by Harvey *et al.* [169] and slightly higher by Bhattacharyya *et al.* [170], while it is worth mentioning the value $V_{mip} \approx 10V$ found by Shiloh *et al.* [109]. It is reasonable to assume that the specific SiN preparation process and tension strongly affect this value. By considering $V_{mip} \approx 15\,V$ the required thickness values to induce a 2π phase shift at various commonly used electron energies are shown in Table *3*.

$E\ (keV)$	60	80	120	200	300
$t\ (nm)$	36.9	41.5	48.5	57.4	64.2

Table 3: *Si_3N_4 thickness required for inducing a full 2π phase shift of the electron wavefunction at different electron energy values.*

In recent years, new novel materials for phase S-CGHs production are being explored. For instance, promising results have been shown with the employment of amorphous Carbon [171].

The mean inner potential for the materials collected in Table 2 are reported below in Table *4*.

Material	Au	Ag	Pt	Si_3N_4	SiO_2	Al_2O_3	a-C
Theoretical (V)	$25 \div 31$	$18.74 \div 23$	$20 \div 27$	$11.3 \div 17.6$	$\sim 15.1\,V$	$15.7 \div 16.7$ [172]	$10.1 \div 11.3$ [173]
Experimental (V)	$21 \div 30$	$17 \div 23$	$\sim 25\,V$	$\sim 15\,V$ [168]	$\sim 17\,V$	16.9 ± 0.36 [172]	$9.09 \div 10.7$ [174] [175]

Table 4: *Theoretical and experimental mean inner potentials of the materials reported in table 1, most of the values are taken from* [176] *and* [177], *for Amorphous Carbon the mip is highly dependent on the chosen material's density.*

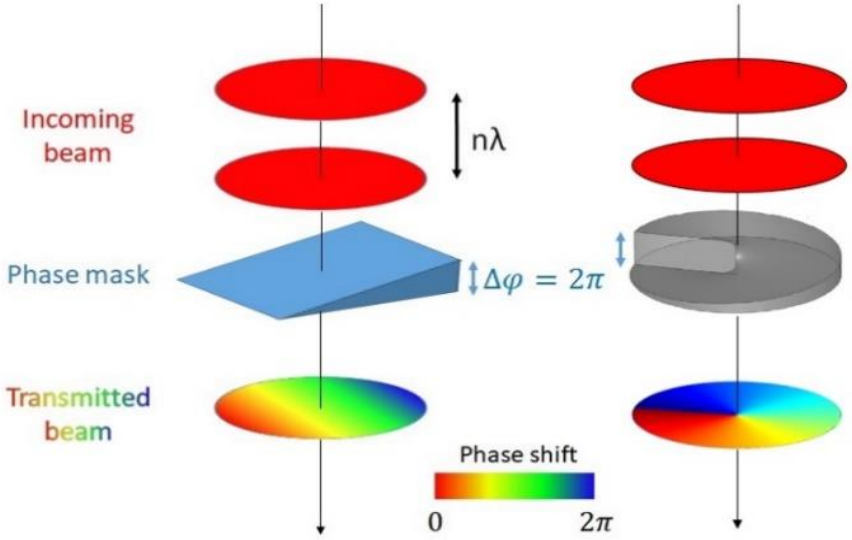

Figure 23: Schematic view of the phase shifting effect of a ramp phase mask (left) and a spiral phase mask (right) on an impinging electron plane wave. Adapted from [10].

An intuitive schematic picture of how simple phase masks work is given in Figure 23. Here, an incoming plane wave whose wavefront, shown in red circles, impinges on a phase mask. The ramp structure, thanks to the phase shifting effect of the material's thickness, will induce a linear phase shift (along a specific direction or the azimuthal angle) onto the plane wave. The thinner region will not alter the beam wavefront, while the thicker region has been chosen such that it introduces a 2π phase shift onto the incoming beams. The resulting linearly phase-shifted wavefront is visualized as a circle with a transverse linear colour gradient, causing the beams to either deflect (left image) or carry OAM (right image). The two-phase masks shown in Figure 23 are so-called *in-line* S-CGH, the simplest type of S-CGH which can be designed - having no transverse grating. From a manufacturing point of view, this type of hologram requires great manufacturing precision, as the phase shift of the outgoing wave is entirely encoded in the pointwise thickness of the hologram. *In-line* holograms, *in-line* S-CGH, do not need a reference beam, because the entire of the desired beam information (phase) is directly imprinted onto the substrate. However, one may use a spherical wave as a reference, and consequently, the generated interest beam is diffracted at a different longitudinal point (along the incident beam propagation direction).

Off-axis S-CGHs are instead obtained by using a plane wave as a reference just as *Off-axis* holograms are. The main advantage that *Off-axis* S-CGH offers with respect to on-axis ones is that the encoded beam and the hologram diffraction pattern are decoupled from one another. This means that off-axis S-CGHs are quite flexible and allow for small imperfections during the fabrication. In fact, the phase shift of the outcoming wave is encoded in the diffraction pattern, and imperfections, if present, will only affect the efficiency and the beam will still be properly generated.

4.1.2.3 Amplitude-phase holograms

This is the most general class of holograms, where the material encodes a deliberate change of both amplitude and phase of the interest beam. In a strict sense, all phase holograms are amplitude-phase holograms since any variation of the thickness affects both amplitude and phase of the diffracted phase upon propagation [178]. Moreover, given the practical difficulties in fabricating complex amplitude gratings that stay mechanically stable, the amplitude gratings can often be built on a continuous thin Si_3N_4 membrane, as for example done by McMorran *et al.* [22]. In such a case, depending on how transparent the thick parts of the grating are, all the levels between amplitude to phase gratings can be obtained. It is worth mentioning that smart use of the two materials could be used to cancel completely the amplitude effects in a phase hologram or rather to add an amplitude envelope to the phase grating. This would allow for a joint amplitude and phase encoding of the wavefunction, but this has not been explored in detail so far. Moreover, even the aperture effect that encloses phase holograms (and truncates the beam) is still a sort of amplitude filtering. In general, amplitude and phase modulations have slightly different effects and the two are superimposed. It is, therefore, more difficult to control both the amplitude and phase simultaneously with different holograms.

4.1.3 Calculating the hologram

4.1.3.1 Encoding the wave phase in a phase hologram and amplitude holograms.

In this section, it will be shown how to calculate the actual hologram profile given the target wavefunction in the off-axis case, i.e., when the target wavefunction $\Psi(k)$ is reproduced in the first order of diffraction. Here we assume that the reference wave is a plane wave with in-plane wavevector $\vec{g}$, determined by the inclination of the reference wave.

In "imaging holography" the condition for a good reconstruction is a very high fringe spacing and an object with a narrow frequency band in order to be able to isolate and demodulate the phase properly. Similarly, for the synthetic holography, to work the desired function $\Psi(k)$ has to be a function with a compact support so that its extension in the Fourier space is smaller than the reference frequency $\vec{g}$. The desired phase modulation $\Delta\varphi(\vec{\rho})$ at the level of the hologram needs to be calculated based on the aimed diffraction shape. The hypothesis is that one can control the phase $\alpha(\vec{\rho})$ of the aimed diffracted beam at the plane of the hologram. This should be simply the phase of the inverse FT of the object $\Psi(k)$ with the addition of a phase gradient affecting the off-axis tilt. So one can numerically calculate $\alpha(\vec{\rho})$ as $\alpha(\vec{\rho}) = arg\{FT^{-1}(\Psi(k))\} + \vec{g} \cdot \vec{\rho}$. For example, in the case of a vortex beam $\alpha(\vec{\rho}) = \ell\theta + \vec{g} \cdot \vec{\rho}$ where ℓ in the desired winding number of the vortex and θ is the azimuth of the $\vec{\rho}$ coordinate in the hologram plane.

If the function $FT^{-1}(\Psi(k))$ had a constant amplitude, then $T_g = exp\left(i\alpha(\vec{\rho})\right)$ would be the exact transmission function T_g of the desired phase plate. Its Fourier transform would be $\Psi(k)$ except a tilt of $\vec{g}$.

The limitation here is that only the phase at the exit of the hologram is encoded. Later methods to generalize this approach will be discussed. The actual phase in the phase hologram case can be any function $\Delta\varphi = f(\alpha)$ with the periodicity condition that $f\big(\alpha(\vec{\rho})\big) = f(\alpha(\vec{\rho}) + 2n\pi)$. For example, a sinusoidal grating generating vortex beams would be then $\Delta\varphi = \varphi_0 sin\,(\ell\theta + \vec{g}\cdot\vec{\rho})$. Since the function $\alpha(\vec{\rho})$ is by definition Fourier bandlimited, the transmission function of the full hologram is $T = exp\,(i\,f(\,\alpha(\vec{\rho})))$ and has approximate periodicity $\vec{g}$. This means it has a diffraction pattern given by many well-separated beams centered at $n\vec{g}$ where $n \in \mathbb{Z}$ is the diffraction order. Each diffracted beam can be spatially separated then and for the first order beam the hologram would act as the desired transmission function $T_g = exp\,(i\alpha(\vec{\rho}))$. What changes as a function of f is the distribution between the diffraction orders: for any form of f, the 1st order will only be affected by a phase effect $T_g = exp\,(i\alpha(\vec{\rho}))$.

In the case of amplitude holograms, it is possible to assume a simplified form of the interference where one retains only the cross term in Eq.83. The simplest form of the interference would be therefore just such cross term $T = cos\,(\alpha(\vec{\rho}))$ which is clearly an amplitude modulation. However, as in the case of phase holograms, one can use any function in the form of $T = f(\alpha(\vec{\rho}))$.

So analogously to the phase hologram a sinusoidal amplitude grating generating a vortex would be in the form of $T = \frac{A_0}{2}(1 + sin\,(\ell\theta + \vec{g}\cdot\vec{\rho}))$ where positivity of the amplitude hologram has been enforced. Even in this case, for any form of f, the 1st order beam will only be affected by a phase effect $T_g = exp(i\alpha(\vec{\rho}))$.

Therefore, at a level of single diffracted beams, amplitude and phase holograms are exactly the same. However, the phasing and amplitude ratio between the diffraction order is different. For example, in amplitude holograms, the first diffraction order is in phase with the zero-order, while in phase hologram there is generally a dephase close to $\pi/2$. The amplitude effects are discussed in section 1.4.3. and 1.4.4.

A special mention is deserved for another category of holograms, named Fresnel holograms. In this case the desired intensity is not reached in the Fraunhofer plane, but instead reached an intermediate one, a Fresnel one hence the name. The Fourier transform must be then substituted with the Fresnel integral, but the concept is identical:

$$\alpha(\vec{\rho}) = arg\{\Psi(\mathrm{k}) \otimes \mathrm{P}(-\Delta\mathrm{z})\} + \vec{g}\cdot\overrightarrow{\rho} \tag{136}$$

where $\otimes$ indicates the convolution integral and $P(\Delta z) = \frac{1}{i\lambda\Delta z} exp\,(i\frac{\pi}{\lambda\Delta z}(x^2 + y^2))$.

4.1.3.2 Diffraction efficiency and groove profile

One of the key parameters of an off-axis S-CGH is its diffraction efficiency. Diffraction efficiency can be defined as the ratio between the measured diffracted intensity of a specific diffraction order and either the intensity of the incoming beam or the total transmitted intensity. According to the first definition, the efficiency is given by:

$$\eta_n^{(i)} = \frac{I_n}{I_{inc}} \tag{137}$$

where I_n is the intensity of the diffraction order n, and I_{inc} is the intensity of the incident beam, and $\eta_n^{(i)}$ is then known as the absolute diffraction efficiency. The total transmitted efficiency is defined as:

$$I_{trans} = \sum_n I_n \tag{138}$$

From this the second definition of diffraction efficiency is given by:

$$\eta_n^{(t)} = \frac{I_n}{I_{trans}} \tag{139}$$

It is worth mentioning that in both amplitude and phase S-CGH, given that the beam has to pass through the material of choice, the total transmitted intensity will be significantly reduced, with typical values up to 50% and 30-40% less, respectively. This is due to absorption and other inelastic processes that happen in the material, and this is valid even for high-transmittance materials such as Si_3N_4. To mitigate this problem, it has been suggested to use high-brightness electron sources as the excessive absorption can make the detection of the target beam quite challenging. From now on, we use η in order to refer to $\eta_n^{(t)}$, the so-called transmitted efficiency. In case a proper distinction between the two is necessary, the appropriate symbol will be used. As it was mentioned previously, the efficiency of a S-CGH depends on its type, i.e., if it is a phase modulated S-CGH or an amplitude modulated one. However, efficiency also depends on the groove profile/thickness pattern of the hologram. To establish the relationship between the groove profile and the efficiency, it is worth explaining how an incoming wavefunction is transformed, after its interaction with an S-CGH, first. This interaction depends on the groove pattern.

As previously introduced in Eq. (8), the transfer function, $T_H(\vec{\rho})$, governs the amplitude and the phase of a beam which has passed through a diffraction grating. An alternative representation of the transfer function specifically for the phase S-CGH is given by:

$$T_H(\vec{\rho}) = e^{i\tilde{V}t(\vec{\rho})} \tag{140}$$

where $\tilde{V} = C_E V_{mip} + i\gamma$ is the complex index of refraction of the material, γ is the absorption coefficient defined as $\gamma = \frac{1}{\lambda_{mfp}}$, λ_{mfp} is the mean free path of an electron inside the material of choice, and $t(\vec{\rho})$ describes the thickness profile of the grating. Since the transfer function is independent from the incident wavefunction $\Psi_{inc}(\vec{\rho})$, then the transmitted wavefunction can be written as:

$$\Psi_t(\vec{\rho}) = \Psi_{inc}(\vec{\rho})T(\vec{\rho}) = \Psi_{inc}(\vec{\rho})e^{i\tilde{V}t(\vec{\rho})} \tag{141}$$

In most case studies, the incoming wave is assumed to be a plane wave, thus it can be ignored as it has a flat-phase wavefront and the expression of the transmitted wavefunction, and its FT will mainly depend on the

transfer function. A generic diffraction grating is characterized by a periodic wavefunction $f(\alpha)$ which describes the groove pattern of the grating. Usually, this function is dimensionless, normalized from zero to one and its period is equal to 2π, so that $f(\alpha + 2\pi) = f(\alpha)$. The function can be expanded into a Fourier series as:

$$f(\alpha) = \sum_{n=-\infty}^{\infty} c_n e^{in\alpha} \tag{142}$$

where c_n is the n^{th} Fourier coefficient, with $n \in \mathbb{Z}$ and it is defined as:

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha) e^{-in\alpha} d\alpha \tag{143}$$

Each n value represents one single diffraction order. If $f(\alpha)$ is real-valued, then $c_n = c_{-n}^*$, where the asterisk denotes the complex conjugate, and c_0 is real.

The Fourier power spectrum of $f(\alpha)$ is given by:

$$S = \sum_n |c_n|^2 \tag{144}$$

For a bi-dimensional grating with a single type of groove profile the periodic function $f(\alpha)$ has

$$\alpha = \alpha(\vec{\rho}) \tag{145}$$

where $|\vec{\rho}|$ and θ, the azimuthal angle, are polar coordinates which define the grating and $|\vec{\rho}|$ is measured in units of the grating spatial period Λ along the $\theta = 0$ direction. For most of the functions $\alpha(\vec{\rho})$ and consequently $f(\alpha(\vec{\rho}))$ that are used for diffraction gratings, the Jacobian, with the exception for the zeroth order, is the same for each diffraction order, so it can be disregarded in the calculations.

For an amplitude S-CGH the field transmission Fresnel function $T(x, y)$ is proportional to the grating function:

$$T(\vec{\rho}) = b\, f\big(\alpha(\vec{\rho})\big) \tag{146}$$

where b is a constant and $0 \leq b \leq 1$.

We denote the wavefunction impinging onto the S-CGH as $\Psi_{in}(\vec{\rho})$. The output wavefunction consequently will be determined as:

$$\Psi_{out}(\vec{\rho}) = T(\vec{\rho})\Psi_{in}(\vec{\rho}) = b\, f\big(\alpha(\vec{\rho})\big)\Psi_{in}(\vec{\rho}) \tag{147}$$

While for phase S-CGH, the Fresnel transmission function is given by

$$T(\vec{\rho}) = e^{i\tilde{a} f(\alpha(\vec{\rho}))} \tag{148}$$

where $\tilde{a}$ is a complex number defined as $\tilde{a} = a_1 + ia_2$. In this relation $a_1 = C_E V_{mip} t_M$ and $a_2 = \gamma t_M$. Here t_M represents the maximum value of the thickness difference between a peak and a valley in the grating. Mathematically, it can be shown that depending on the beam accelerating voltage, for Si_3N_4 $a_2 \sim 7 \div 8\%\ a_1$.

Furthermore, the product of t_M and $f(\alpha(\vec{\rho}))$ yields the local thickness of the grating $t(\vec{\rho})$. Therefore, for this type of holograms, the output wavefunction is given by:

$$\Psi_{out}(\vec{\rho}) = \Psi_{in}(\vec{\rho}) e^{i\tilde{a} f(\alpha(\vec{\rho}))} \tag{149}$$

The efficiency of the S-CGH can be estimated/calculated from the power transmission spectrum, given by the sum of the Fourier coefficients of the transmission function Fourier series expansion:

$$\mathcal{T}(\vec{\rho}) = \sum_n |\tau_n|^2 \tag{150}$$

where the index n yields the diffraction order and the n-th diffraction order intensity is modulated by the transmission coefficient $|\tau_n|^2$. Furthermore, for amplitude S-CGHs the diffraction angle of the n-th order diffracted beam will be $\theta_n = \frac{n\lambda}{\Lambda}$. Lastly, it is easy to observe/notice that the power transmittance for pure phase gratings is unity, so ideally they do not absorb and can support a very high power peak.

4.1.3.3 Comparison between different geometry profiles

In this section, a series of example grating profiles regarding S-CGHs are presented, for both amplitude and phase ones. For each of them the profile function and the Fourier coefficients of the transmission function, τ_n, are given.

Sinusoidal/cosinusoidal profile

The simplest profile that can be studied is a sinusoidal/cosinusoidal. The sinusoidal and cosinusoidal cases have the same characteristics, therefore only one of them will be analyzed here. For the amplitude S-CGH with a cosinusoidal profile, the profile function is $f(\alpha) = \frac{1}{2}(1 + cos(\alpha(\vec{\rho})))$. So that the transmission function is:

$$T(\vec{\rho}) = \frac{b}{2}(1 + cos(\alpha(\vec{\rho}))) \tag{151}$$

where $0 \leq b \leq 1$ is a constant. Instead, for a phase S-CGH, the transmission function can be written as:

$$T(\vec{\rho}) = e^{i\frac{\tilde{a}}{2}(1+cos(\alpha(\vec{\rho})))} = e^{i\frac{a_1}{2}cos(\alpha(\vec{\rho}))} e^{-\frac{a_2}{2}cos(\alpha(\vec{\rho}))} e^{i\frac{\tilde{a}}{2}} = e^{ia_1' cos(\alpha(\vec{\rho}))} e^{-a_2' cos(\alpha(\vec{\rho}))} e^{i\tilde{a}'} \tag{152}$$

where the primed variables are used to simplify the calculations and comprise the $\frac{1}{2}$ factor, and for both types of holograms $f(\alpha)$ is normalized between zero and one so that for the amplitude S-CGH the power transmittance locally changes between 0 for full absorption, and 1 in case of no absorption. Conversely, for the ideal phase S-CGH the power transmittance is always unitary and it is possible to estimate the optimal phase shift introduced by the local thickness to maximize one of the diffraction orders intensity, usually the n = ±1. The square modulus of the Fourier coefficients of the transmission function for the nth order are for an amplitude S-CGH:

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} \frac{1}{4}b^2 \ \ for\ n = 0 \\ \frac{1}{16}b^2\ for\ n = \pm 1 \\ 0\ for\ the\ other\ orders \end{cases} \tag{153}$$

while for a phase S-CGH we can derive it from rewriting $T(\vec{\rho})$ using the known Jacobi-Anger expansion:

$$T(\vec{\rho}) = e^{i\tilde{a}\prime} \sum_{n=-\infty}^{\infty} \left(i^n J_n(\tilde{a}')\right) e^{in\alpha(\vec{\rho})}\,. \tag{154}$$

The Fourier series coefficients of $T(\vec{\rho})$ are:

$$\tau_n = i^n J_n(\tilde{a}') e^{i\tilde{a}\prime} \tag{155}$$

It is possible to plot $|\tau_n|^2$ with the complex argument Bessel Function, while an analytical approximation can be derived restarting from Eq.152, in particular the last step. The first term of $T(\vec{\rho})$ can be rewritten using the known Jacobi-Anger expansion, while the second we expand it by approximating $a_2' \sim 0$:

$$T(\rho) \approx \exp(ia)\exp(ia_1\cos(\alpha))\,(1 - a_2\cos(\alpha)) \tag{156}$$

This means

$$T(\rho) \approx \exp(i\tilde{a}')\left[\sum_n (-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right](1 - a'_2\cos(\alpha)) \tag{157}$$

The Fourier coefficient is defined

$$\tau_m = \int T(\rho)\exp(-im\alpha)\,d\alpha \tag{158}$$

We can use the property of convolution

$$\tau_m = \exp(i\tilde{a}')\left\{\int\left[\sum_n (-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right]\exp(-im\alpha)\,d\alpha\right\} \tag{159}$$
$$*\left\{\int (1 - a'_2\cos(\alpha))\exp(-im\alpha)\,d\alpha\right\}$$

We can develop the 2 terms

$$\int\left[\sum_n (-i)^n \mathrm{J}_n(a'_1)\exp(in\alpha)\right]\exp(-im\alpha)\,d\alpha = \sum_n \delta_{m,n}\mathrm{J}_n(a'_1)(-i)^n \tag{160}$$

$$\int(1 - a_2\cos(\alpha))\exp(-im\alpha)\,d\alpha = \int\left(1 - \frac{a'_2}{2}(\exp(i\alpha) + \exp(-i\alpha))\right)\exp(-im\alpha)\,d\alpha = \tag{161}$$
$$= \delta_{m,0} - \frac{a'_2}{2}\left(\delta_{m,1} + \delta_{m,-1}\right)$$

So the final coefficient is

$$\tau_m = (-i)^m\left(\mathrm{J}_m(a'_1)\right) * \left[\delta_{m,0} - \frac{a'^2_2}{2}\left(\delta_{m,1} + \delta_{m,-1}\right)\right]\exp(i\tilde{a}') \tag{162}$$

If we now calculate the discrete convolution then:

$$\tau_m = \exp(i\tilde{a}')\sum_k (-i)^k\left(\mathrm{J}_k(a'_1)\right)\left[\delta_{k,m-0} - \frac{a'^2_2}{2}\left(\delta_{k,m-1} + \delta_{k,m+1}\right)\right] \tag{163}$$

So it results

$$\tau_m = (-i)^m \mathrm{J}_m(a_1) - \frac{a_2}{2}\left((-i)^{m+1}\mathrm{J}_{m+1}(a'_1) + (-i)^{m-1}\mathrm{J}_{m-1}(a_1)\right]\exp(i\tilde{a}') \tag{164}$$

$$\tau_m = (-i)^m\{\mathrm{J}_m(a_1) + \frac{ia_2}{2}\left[\mathrm{J}_{m-1}(a_1) - \mathrm{J}_{m+1}(a_1)\right]\}\exp(i\tilde{a}')$$

The efficiency of the n^{th} diffraction order is proportional to $|\tau_n|^2$, where:

$$|\tau_n(\vec{\rho})|^2 = J_n^2(\tilde{a}')\, e^{-2a_2'} \approx \left[J_n{}^2(a'_1) - \frac{a'^2_2}{4}\left(J_{n-1}(a'_1) - J_{n+1}(a'_1)\right)^2\right] e^{-2a'_2} \tag{165}$$

Figure 24a shows the characteristic efficiency of the first diffracted order for a phase S-CGH as a function of the two parameters a'_1 and a'_2. It is possible to observe how the efficiency reaches a maximum when a'_1 is equal to $\approx 1.84\ rad$. In the "real phase S-CGH" trend the contribution of the absorption is appreciable, the main contribution to it originates from the exponential term $e^{-2a_2'}$, while it's negligible when considering the modified Bessel function I_n. The ideal value of a'_1, which is the one that maximizes the $|\tau_1(\vec{\rho})|^2$ of an ideal phase S-CGH that has been just estimated, corresponds to a peak to valley phase difference of approximately 3.68 radians ($1.17\ \pi$ in units of *pi)* corresponding to an approximate value of $t_M \approx 33.6\ nm$ for 200 keV electrons.

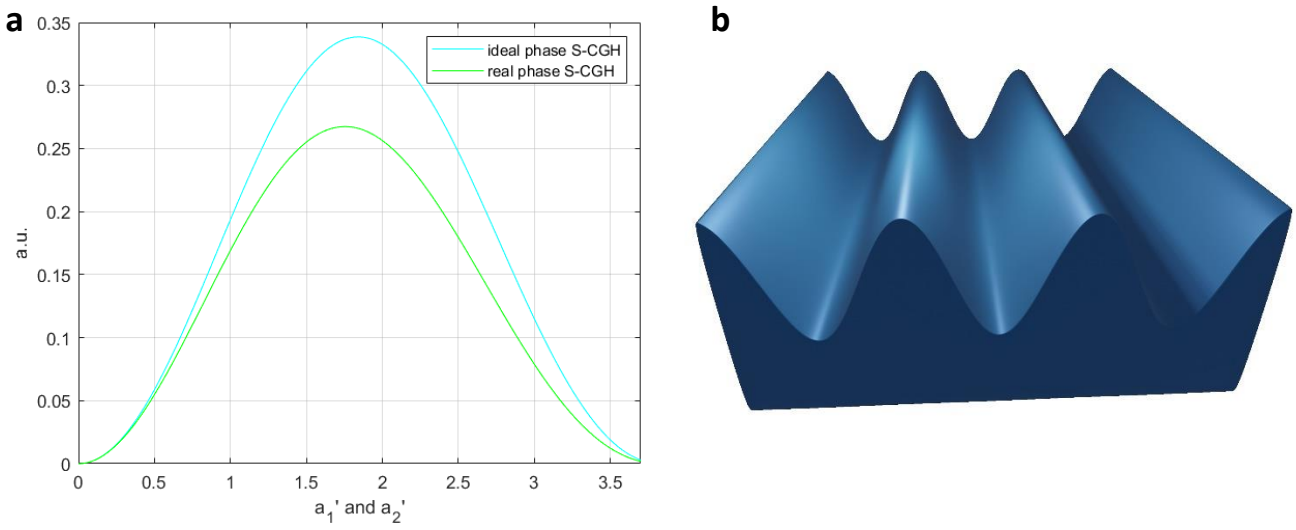

Figure 24: a) Plot of the profile of $|\tau_1|^2$ of a phase S-CGH as a function of a'_1. b) 3-D rendering of a cosine profile

Squared profile

The second profile that is studied here is the squared profile. In this case, the periodic function represented by $f(\alpha) = \frac{1}{2}(1 + Sign\left(sin\left(\alpha(\vec{\rho})\right)\right))$, so that for an amplitude S-CGH the transmission function is:

$$T(\vec{\rho}) = \frac{b}{2}\left(1 + Sign\left(sin\left(\alpha(\vec{\rho})\right)\right)\right) \tag{166}$$

while, for a phase one the transmission function is written as:

$$T(\vec{\rho}) = e^{i\frac{\tilde{a}}{2}(1+Sign(sin(\alpha(\vec{\rho}))))} = e^{i\tilde{a}\prime Sign(sin(\alpha(\vec{\rho})))}e^{i\tilde{a}\prime} =$$

$$= e^{ia\prime_1 Sign(sin(\alpha(\vec{\rho})))} \cdot e^{-a\prime_2 Sign(sin(\alpha(\vec{\rho})))}e^{i\tilde{a}\prime} \tag{167}$$

where the primed variables are used to simplify the calculations and comprise the $\frac{1}{2}$ factor. The square modulus of the Fourier coefficients of the transmission function for the two type of gratings are:

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} \frac{1}{4}b^2 \ \ for\, n = 0 \\ \frac{1}{n^2\pi^2}b^2 \ for\, n = odd \\ 0 \ for\, n = even \end{cases} \tag{168}$$

for the amplitude grating, while, as before, for the phase one we start from Fourier coefficients of $T(\vec{\rho})$ which can be calculated from:

$$\tau_n = \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}\prime Sign(sin(\alpha(\vec{\rho})))}e^{-in\alpha(\vec{\rho})}d\alpha \tag{169}$$

Due to the properties of the *Sign* function, Eq.169 can be split in:

$$\tau_n = \frac{1}{2\pi}\left[\int_0^{\pi} e^{i\tilde{a}\prime}e^{-in\alpha(\vec{\rho})}d\alpha + \int_{\pi}^{2\pi} e^{-i\tilde{a}\prime}e^{-in\alpha(\vec{\rho})}d\alpha\right] = \begin{cases} \cos(\tilde{a}^{'}) \ \ \text{for } n = 0 \\ 0 \text{ for } n \text{ even} \\ \frac{2\sin(\tilde{a}')}{n\pi} \text{ for } n \text{ odd} \end{cases} \tag{170}$$

So that, for example:

$$\tau_1 = \frac{2\sin(\tilde{a}')}{\pi} = \frac{2}{\pi}[\sin(a'_1)\cosh(a'_2) + icos(a'_1)\sinh(a'_2)] \tag{171}$$

And the efficiency of the first diffracted order is proportional to:

$$|\tau_1|^2 = \frac{4}{\pi^2}[\sin^2(a'_1)\cosh^2(a'_2) + cos^2(a'_1)\sinh^2(a'_2)] \tag{172}$$

is:

$$|\tau_n(\vec{\rho})|^2 = \begin{cases} [\cos^2(a^{'}_1)\cosh^2(a^{'}_2) + sin^2(a^{'}_1)\sinh^2(a^{'}_2)]e^{-2a_2\prime} \ for\, n = 0 \\ \frac{4}{n^2\pi^2}\left[\sin^2(a^{'}_1)\cosh^2(a^{'}_2) + cos^2(a^{'}_1)\sinh^2(a^{'}_2)\right]e^{-2a_2\prime} \ for\, n = odd \\ 0 \ for\, n = even \end{cases} \tag{173}$$

Figure 25a reports the efficiency profile of the first diffraction order for the phase S-CGH, the maximum is reached when $a'_1 \approx 1.57\, rad$, so that the optimal phase difference between peek and valley is $\Delta\varphi \approx \pi$, in an ideal phase S-CGH.

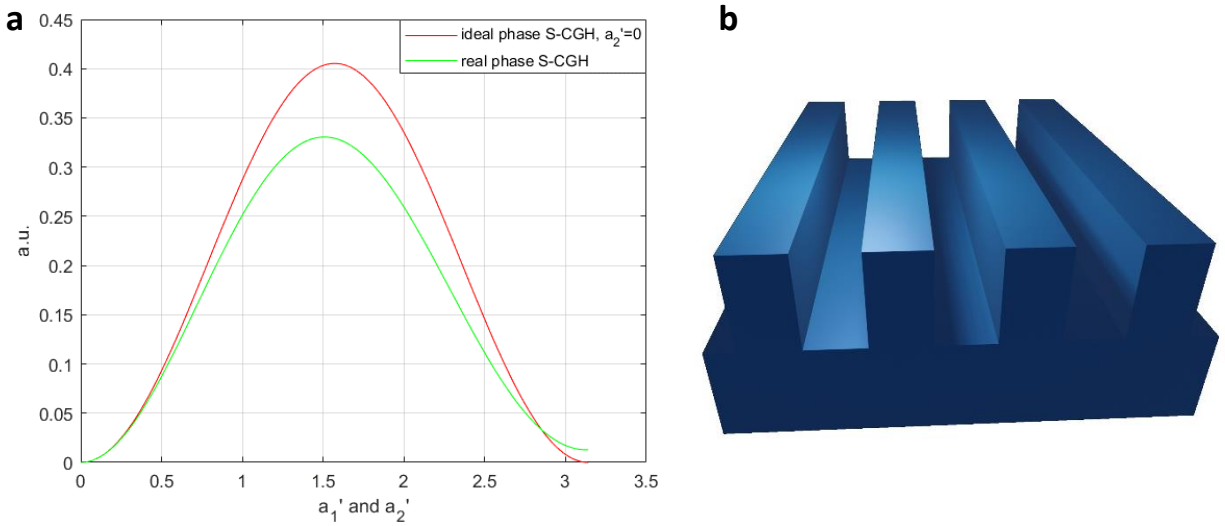

Figure 25: a) Plot of the profile of $|\tau_1|^2$ as a function of a'_1, the amplitude of the sign function, corresponding to half of the peak-to-valley distance expressed in radians. In Red without absorption, in green with absorption. b) 3-D rendering of a squared profile.

A special square profile for amplitude S-CGHs that is worth reporting is the square profile with variable duty cycle. For this profile the profile function between 0 and 2π is:

$$f(\alpha) = \begin{cases} 1 & for\ 0 < \alpha \leq D \cdot 2\pi \\ 0 & for\ D \cdot 2\pi < \alpha \leq 2\pi \end{cases} \tag{174}$$

Where D is the duty cycle, it is constant and $0 < D < 1$. The fundamental frequency associated to $f(\alpha)$ is $\omega_0 = 1$ since the period is 2π. It is now possible to calculate the Fourier coefficients:

$$\tau_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha)d\alpha = \frac{1}{2\pi}\int_0^{D*2\pi} 1\, d\alpha = D \tag{175}$$

$$\tau_{n\neq 0} = \frac{1}{2\pi}\int_0^{2\pi} f(\alpha)e^{in\omega_0\alpha}d\alpha = \frac{1}{2\pi}\int_0^{D*2\pi} 1e^{in\alpha}d\alpha = \frac{1}{2\pi}\frac{1}{in}\left[e^{inD2\pi} - 1\right] = \\ = \frac{1}{2}\frac{e^{inD\pi}}{in\pi}\left[e^{inD\pi} - e^{-inD\pi}\right] = \frac{e^{inD\pi}}{n\pi}sin(nD\pi) = De^{inD\pi}sinc(nD\pi) \tag{176}$$

So that:

$$|\tau_n|^2 = \begin{cases} D^2 & for\ n = 0 \\ D^2 sinc^2(nD\pi) & for\ n \neq 0 \end{cases} \tag{177}$$

This, naturally, can be re-elaborated to be readapted to a phase S-CGH.

Triangular profile

The third case we discuss here is the case of a triangular shaped profile, specifically an isosceles triangle. This particular shape can be described by the profile function $f(\alpha) = \frac{1}{\pi}(Sign\left(sin\left(\alpha(\vec{\rho})\right)\right))((\pi - Mod(\alpha(\vec{\rho}), 2\pi))$. In the case of the amplitude S-CGH the transmission function is defined as:

$$T(\vec{\rho}) = b\frac{1}{\pi}\left(Sign\left(sin\left(\alpha(\vec{\rho})\right)\right)\right)\left(\pi - Mod(\alpha(\vec{\rho}), 2\pi)\right) \tag{178}$$

while for the phase one:

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{\pi}\left(Sign\left(sin\left(\alpha(\vec{\rho})\right)\right)\right)\left(\pi - Mod(\alpha(\vec{\rho}),2\pi)\right)} \tag{179}$$

where $Mod(p,q)$ represents the remainder obtained when dividing p by q.

It is then possible to derive that the generic efficiency of the n^{th} diffracted order is proportional to:

$$|\tau_n|^2 = \begin{cases} \frac{1}{4}b^2 \ for\ n = 0 \\ \frac{4}{n^4\pi^4}b^2 \ for\ n = odd \\ 0 \ for\ n = even \end{cases} \tag{180}$$

for the amplitude case. In the case of a phase S-CGH (with triangular modulation) we calculate first the Fourier coefficients from the integral:

$$\tau_n = \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}\frac{1}{\pi}\left(Sign\left(sin\left(\alpha(\vec{\rho})\right)\right)\right)\left(\pi - Mod(\alpha(\vec{\rho}),2\pi)\right)} e^{-in\alpha(\vec{\rho})} d\alpha \tag{181}$$

Which results in:

$$\tau_n = \frac{-i(a_1 + ia_2)\left[(-1)^{n+1} + e^{i(a_1+ia_2)}\right]}{(a_1^2 + 2ia_1a_2 - a_2^2 - n^2\pi^2)} \tag{182}$$

From which, the generic efficiency of the n^{th} diffracted order is proportional to:

$$|\tau_n|^2 = \frac{(a_1^2 + a_2^2)[1 + 2(-1)^{n+1}\, e^{-a_2}(\cos(a_1)) + e^{-2a_2}]}{[a_1^4 + 2a_1^2a_2^2 + a_2^4 + (n\pi)^4 - 2n^2a_1^2\pi^2 + 2n^2a_2^2\pi^2]} \tag{183}$$

In Figure 26a it is possible to observe how the efficiency changes as different parameters are varied. If a_2 is different from zero, i.e., the contribution to the efficiency of the absorption is considered, the efficiency is

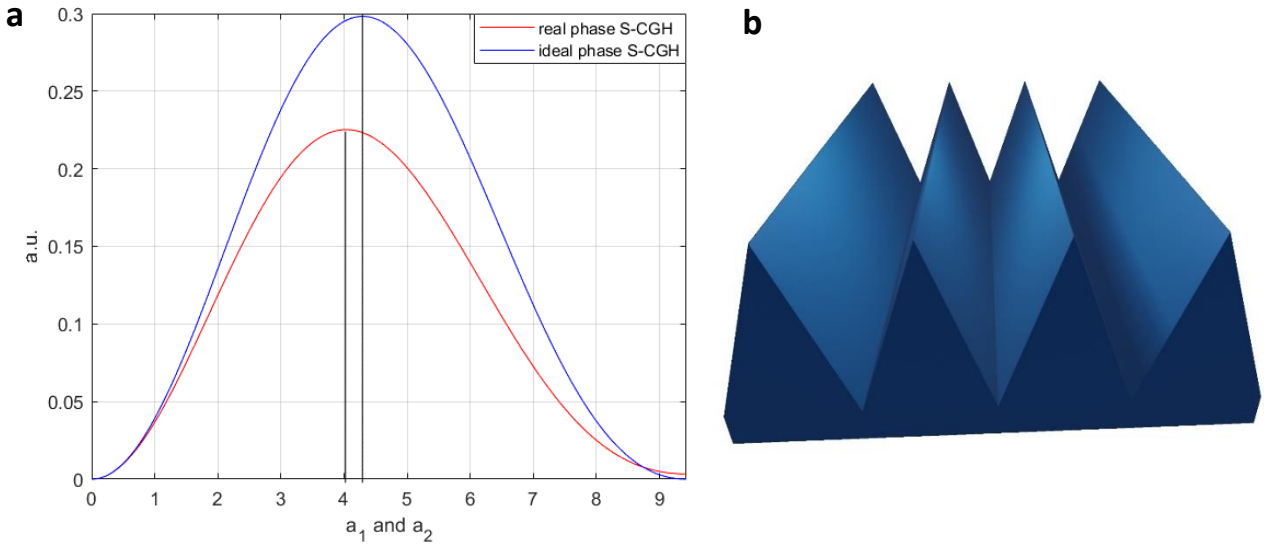

Figure 26:a) Plot of $|\tau_1|^2$ for a phase S-CGH as a function of a_1 and a_2 . b) 3-D rendering of a triangular profile.

reduced and the peak position slightly changes, moving to lower values of a_1. In fact, while for an ideal phase S-CGH $|\tau_1|^2$ has a maximum at $a_1 \approx 4.31\ rad$. For a real S-CGH the peak is reached when $a_1 \approx 4\ rad$.

Blazed profile

A particular and very interesting case of a triangular profile is the blazed profile. In this case, the profile is similar to the one found in a sawtooth blade and can be described by the function: $f(\alpha) = \frac{1}{2\pi}\big(Mod(\alpha(\vec{\rho}), 2\pi)\big)$. Same as the previous profiles, the transmittance function for the amplitude S-CGH is then written as:

$$T(\vec{\rho}) = b\frac{1}{2\pi}\big(Mod(\alpha(\vec{\rho}), 2\pi)\big) \tag{184}$$

while for the phase one the transmittance function is:

$$T(\vec{\rho}) = e^{i\tilde{a}\frac{1}{2\pi}\left(Mod(\alpha(\vec{\rho}),2\pi)\right)} \tag{185}$$

So that for the amplitude synthetic hologram, the efficiency of n-th order of diffraction is given by:

$$|\tau_n|^2 = \begin{cases} \frac{1}{4}\ for\ n = 0 \\ \frac{1}{4\pi^2 n^2}\ for\ n \neq 0 \end{cases} \tag{186}$$

For a phase S-CGH, the efficiency can again be calculated from the Fourier coefficients of Eq. 185:

$$\tau_n = \frac{1}{2\pi}\int_0^{2\pi} e^{i\tilde{a}\frac{1}{2\pi}\left(Mod(\alpha(\vec{\rho}),2\pi)\right)} e^{in\alpha(\vec{\rho})} d\alpha = \frac{-i\left(-1 + e^{i\tilde{a}}\right)}{(\tilde{a} + 2\pi n)} \tag{187}$$

So that:

$$|\tau_n|^2 = \frac{(1 + e^{-2a_2} - 2\cos(a_1)e^{-a_2})}{[(a_1 + 2\pi n)^2 + a_2^2]} \tag{188}$$

Figure 27a shows two key features of the blazed profile in phase S-CGH: first, in the ideal case the maximum efficiency is reached when the peak-to-valley distance is equivalent to a phase difference of 2π; second, by well tuning the shape it is possible to reach an efficiency of almost 100% on one of the two first diffraction orders. This is better represented in Figure 27b.This type of profile theoretically is the only one that allows for the whole transmitted wave to be directed to one of the two first diffraction orders during measurement. In the case that the blazed profile is not perfect and the shape of the grooves is more similar to a series of scalene triangles instead of a series of rectangular triangles, then the transmitted intensity is no longer concentrated on one of the first diffracted orders, but is spread over the others. This can be simulated but it is usually clearly visible when looking at the diffraction pattern of a real blazed hologram, which is rarely ideal, due to fabrication limitations and the profile approximates the calculated one.

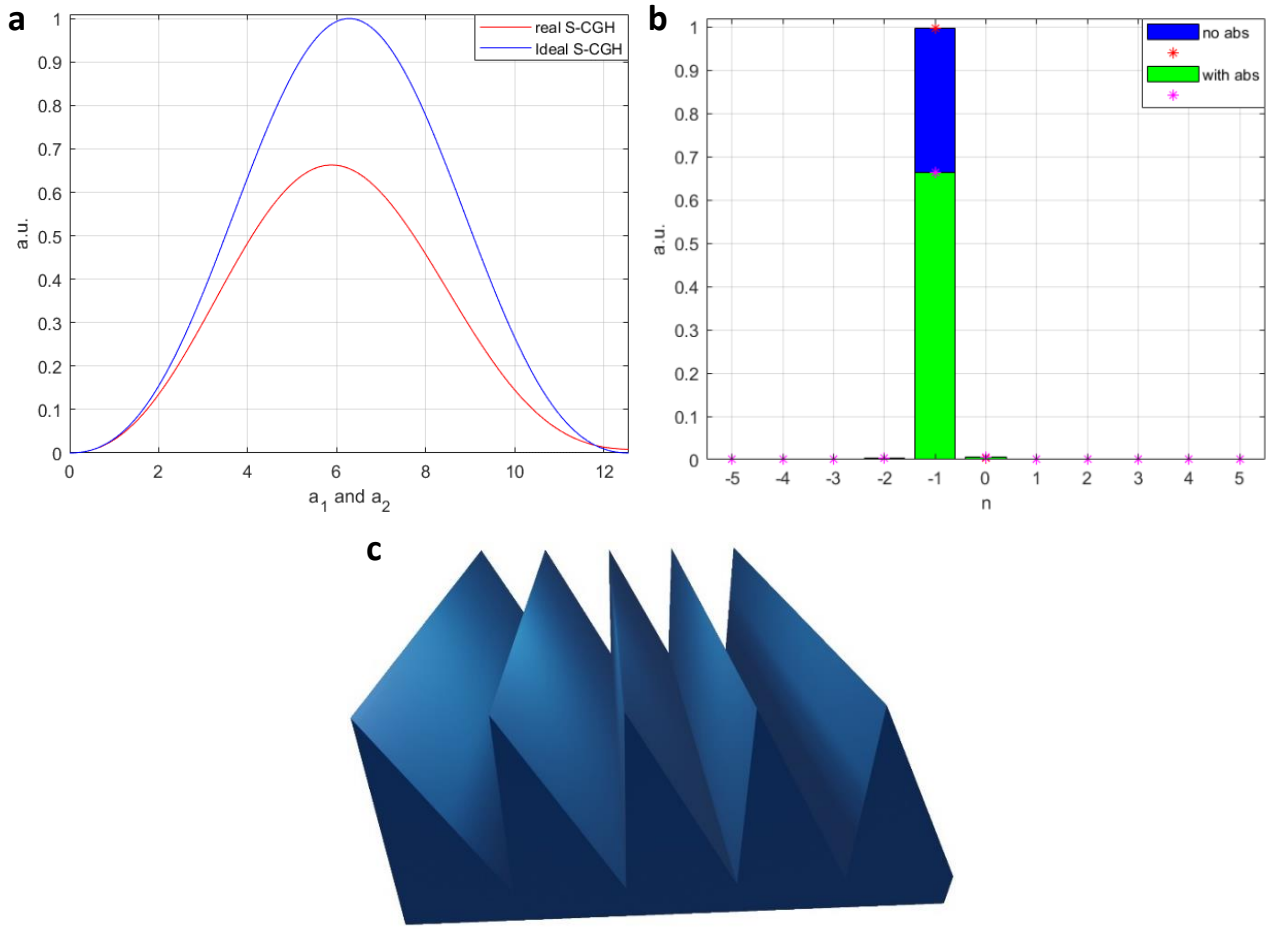

Figure 27: a) Simulated relative transmitted efficiency for the -1 diffracted order with and without absorption. In this case the maximum of the relative efficiency is reached for $a_1 \sim 2\pi$, meaning that in the end when no absorption is considered, the ideal efficiency is obtained when the phase difference between peak and valley due to the thickness is equal to 2π. b) Comparison between the values of $|\tau_n|^2$ forthe different diffraction orders for the ideal case of a blazed phase S-CGH and the "real case" where absorption is taken into account. It is possible to notice how in the ideal case only the minus one order survives, and all the others should ideally have zero intensity, while for the case in which absorption is considered, $\mathcal{T}$ is not anymore unitary. c) 3-D rendering of a blazed profile.

4.1.3.4 Efficiency of the different profiles

In this final subsection the efficiency of each of the profiles that have just been described will be compared, distinguishing between amplitude and phase S-CGH. As mentioned before, efficiency is one of the critical parameters to consider during the designing process of a synthetic hologram. Here, a series of histograms is reported, one for each type of groove pattern, where the diffraction order is reported on the horizontal axis, and on the vertical axis shows the transmitted efficiency, $\eta_n^{(t)}$. For ease of visualization only the diffraction orders between -5 and +5 are presented.

Amplitude S-CGH

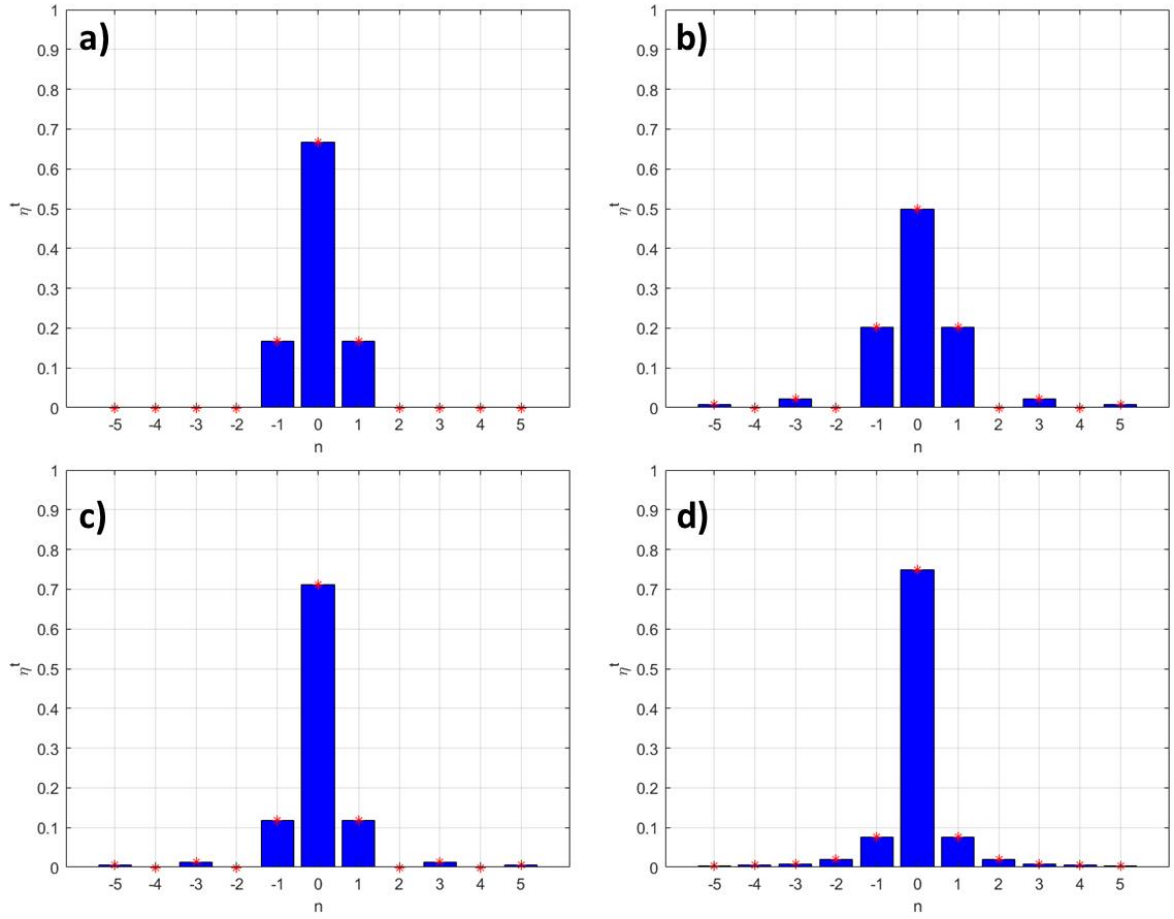

Figure 28: transmitted efficiency of the various diffraction orders, $n \in [-5,5]$, for an amplitude S-CGH: a) sinusoidal profile, b) squared profile, c) triangular profile and d) blazed profile.

Figure 28 shows the intensity distribution between the diffraction orders for different kinds of profile shapes (sinusoidal, squared, triangular and blazed) for the amplitude S-CGH. One common feature is that the central peak or the zeroth order has the highest diffraction efficiency. The total transmitted intensity for amplitude S-CGH is never 100% since the hologram absorbs some of the incoming electrons. As a matter of fact, in the first approximation, if only the absorption from the opaque part of the hologram is considered and the one from the supporting layer is neglected, the best performing shape is the squared one, where 50% of the incoming intensity is transmitted. In comparison the worst one is blazed where only 33% goes through. The key values for the efficiencies are reported in Table *5*.

Ideal Phase S-CGH

The intensity distributions between the various diffraction orders for the possible profiles of an ideal phase S-CGH are shown in Figure 29. In this case, the calculations have been carried out so that the phase difference maximizes the intensity on one of the two first diffraction orders. Therefore, the zeroth order is always less

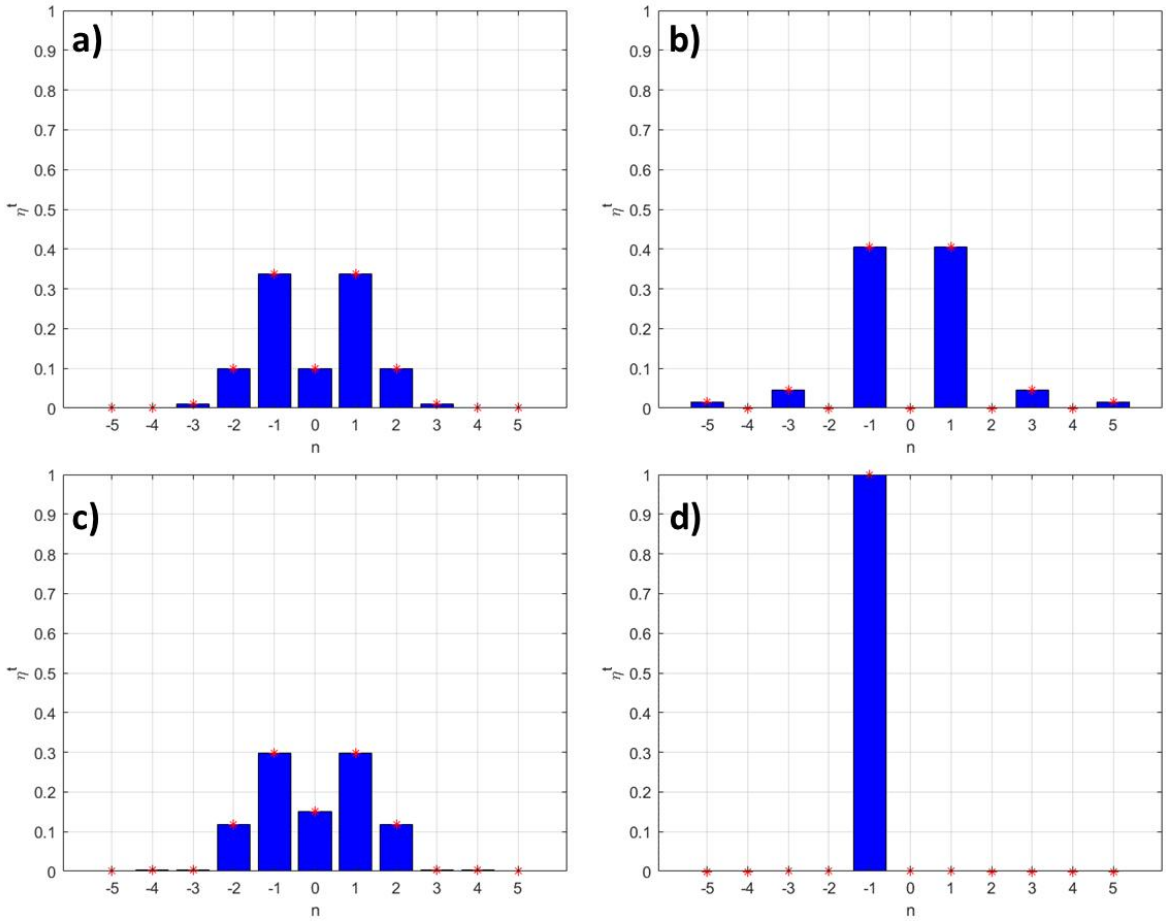

Figure 29: transmitted efficiency of the various diffraction orders, $n \in [-5,5]$, for an ideal phase S-CGH: a) sinusoidal profile, b) squared profile, c) triangular profile and d) blazed profile.

intense than the first orders. While this is easily achievable in phase S-CGHs by properly tuning the phase difference between peak and valley, in amplitude S-CGHs this is not possible and usually the zeroth diffraction order is always the most intense one. Moreover, as it can be expected from ideal phase S-CGHs, where the absorption contribution is omitted, the total transmitted intensity is almost 100% for all the various shapes considered.

The transmitted efficiency for the various shapes and S-CGH types are summarized in Table *5*:

Profile shape	$\mathcal{T}_{amp}$	$\lvert\tau_{\pm1}\rvert^2$	$\frac{\lvert\tau_{\pm1}\rvert^2}{\mathcal{T}_{amp}}$	$\mathcal{T}_{phase}$	$\lvert\tau_{\pm1}\rvert^2$	$\frac{\lvert\tau_{\pm1}\rvert^2}{\mathcal{T}_{phase}}$
Cosinusoidal	*37.5%*	*6.25%*	*16.67%*	*100%*	*33.86%*	*33.86%*
Squared	*50%*	*10.13%*	*20.26%*	*100%*	*40.53%*	*40.53%*
Triangular	*35.13%*	*4.11%*	*11.69%*	*100%*	*29.82%*	*29.82%*
Blazed	*33.3%*	*2.53%*	*7.60%*	*100%*	*100%*	*100%*

Table 5: Summarizes the transmission power function, the n=±1 square modulus of Fourier coefficients and the transmitted efficiency for the two kinds of S-CGH, and all previously studied groove profiles.

4.1.4 Encoding both amplitude and phase in a synthetic hologram

4.1.4.1 Encoding amplitude and phase in a phase hologram

This family of holograms allows the encoding and generation of both the amplitude and the phase of an arbitrary wavefunction using a single phase-only S-CGH. Unlike the other S-CGHs presented so far, which aimed at generating the desired wavefunction in all the nonzero diffraction orders (apart from a multiplicative factor for the angular momentum), mixed holograms generate the desired wavefunction only at a specific diffraction order. The method is based on the tuning of the peak-to-valley phase difference in each region of the S-CGH profile. This yields a local change in efficiency, which changes the wavefront phase at the exit of the hologram, resulting in a change of intensity in the produced beam. Being $A(\vec{\rho})$ and $\varphi(\vec{\rho})$ the amplitude and phase of the desired wavefunction, $B(\vec{\rho})$ a normalized bounded positive function of amplitude and $C(\vec{\rho})$ an analytical function of the amplitude and phase profiles of the desired field and Λ the period of the diffraction grating, the profile to be fabricated has the following form [165]:

$$T_{Mix}(\vec{\rho}) = exp\left[iB(\vec{\rho})Mod\left(C(\vec{\rho}) + \frac{2\pi\rho_{\theta=0}}{\Lambda}, 2\pi\right)\right] \tag{189}$$

where

$$B(\vec{\rho}) = 1 + \pi^{-1}sinc^{-1}\big(A(\vec{\rho})\big) \tag{190}$$

and

$$C(\vec{\rho}) = \varphi(\vec{\rho}) - \pi B(\vec{\rho}) \tag{191}$$

here, $sinc^{-1}()$ is the inverse of sinc function in the interval of $[-\pi, 0]$. As a result, the areas characterized by a full 2π phase shift contribute to the amplitude of the +1st diffraction order, whereas the other areas spread the intensity over the other orders, limiting the intensity of the 1st order. The beam of interest is therefore generated in the 1st diffraction order only, with the correct phase and amplitude information.

4.1.4.2 Encoding amplitude and phase in an amplitude hologram

The method explained in section 4.1.4.1 for phase-only holograms is based on the modulation of the peak-valley value. This modulation locally varies the efficiency of the grating and therefore the amplitude encoding. This same method can be used to create an amplitude encoding in an amplitude hologram.

For sake of definiteness and simplicity one can start by considering a binary mask, so a rectangular profile, and first encode the phase only then later add a modulation to the width of the groove that is directly related to the local efficiency of the hologram, as explained in eq. (36).

For sake of definiteness and simplicity one can start by considering a binary mask, so a rectangular profile and encode the phase only and later add a modulation to the width of the groove that as explained in eq. 167 is directly related to the local efficiency of the hologram.

In simple terms, the center of each groove is related to the phase modulation while the width is related to the amplitude of the wave. Of course, instead of the rectangular groove one can choose any groove profile. Now, if the phase only modulation is chosen so that $f(\alpha) \propto cos\ (kx + \alpha(\vec{\rho}))$, then the center of the fringes of such phase only correspond to the condition $cos\ \left(kx + \alpha(\vec{\rho})\right) = 1$. The idea of the amplitude modulation is to substitute the 1 with a "bias" function $cos\ (q(\vec{\rho}))$, where q(r) is a function of the local desired efficiency, so that the relation $cos\left(kx + \alpha(\vec{\rho})\right) = cos(q(\vec{\rho}))$ can be used to find the clipping points at the side of the groove. Based on this simple principle many different mathematical approaches can be used and will not be here further mentioned but the reader can refer for example to [179].

Of course, the conceptual scheme is somehow more general and can be stretched to include the ideas for the phase hologram encoding of amplitude and phase as seen above. Furthermore, one can notice that the approach above is more exact as it accounts for the amplitude modulation effect on phase shift and phase effect on the amplitude.

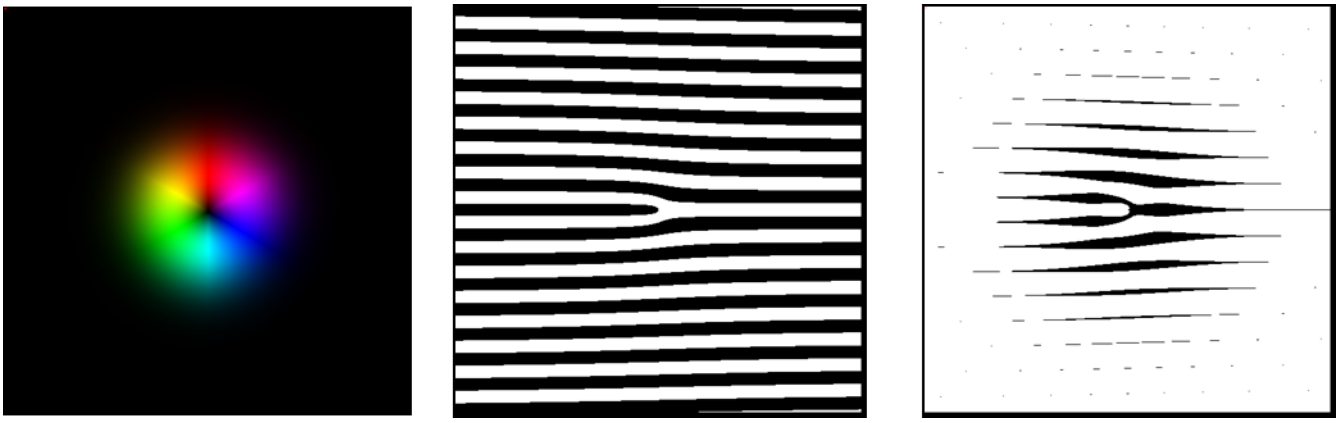

Figure 30: example of encoding of amplitude and phase of a Laguerre Gauss beam (ℓ =1 p=0) into an amplitude hologram. The phase encoding gives the pitchfork but the change in the groove width gives the amplitude envelope.

4.1.5 Sampling Effect and the choice of the groove shape

When deciding on the design of the hologram and particularly on the groove shape, the practical problem of the limitation in the number of addressable or calculated pixels should be taken into account. The calculation so far makes use of a limited number of pixels, typically a hologram uses a square of pixels with size between 1000 to 4000. Beyond 8000 pixels it is both computationally and experimentally demanding to build a S-CGH. This means that a typical groove is sampled with $n_{pg} = 5$ to 20 pixels. A naive argument makes us conclude that the resolution in the groove positioning is given by $1/n_{pg}$ and therefore the phase is defined to be within $2\pi/n_{pg}$. Normally, such a significant phase error determines problems in the actual phase shaping. Actually, this problem exists only with the rectangular groove and to some extent in the blazed groove. In fact, the implied discontinuity is always defined by the size of the pixel. A sinusoidal groove instead has the advantage that each pixel intensity defines the phase with no discontinuity. A more explicit argument could be the following: even if the center of the groove is not defined by a single pixel it can be easily calculated with sub-

pixel precision by a weighted position average while in the rectangular groove the phase is defined only on a discrete grid. The typical feature of the falsely encoded phase can be observed in the hologram's Fourier transform as a set of strike intensity between the diffraction order. In particular conditions, it is possible to recognize many (actually n_{pg}) copies of the same beam that become fainter with increasing n_{pg}. The mathematics of this effect, which can be connected to the Talbot effect, has been calculated by Lee [180]. A second point to be taken into account is the bandwidth of the function to be encoded. The carrier frequency $|\vec{g}|$ must be much larger than the bandwidth of the signal. Moreover, for a sinusoidal pattern at least 4 pixels are needed per period. So, if the bandwidth is B, then $|\vec{g}| >> 2B$ and

$$n_p = K_{max} \gg 8B \tag{192}$$

For example, in the case of a vortex beam with top hat amplitude cutoff, it is possible to find that roughly $B \approx a\,\ell$ with $a \approx 1/\pi$, so for a beam with $\ell = 1000$ we need roughly 4000 pixels. It is obvious that a different groove shape could make the difference in the maximum winding number of the vortex that can be generated. Finally, one should consider that the groove shape also depends on the fabrication approach. For example, with EBL, it is much more difficult to reproduce a groove different from rectangular. More details on both vortex beams generation and fabrication techniques will be provided in the sections ahead.

4.2 - Production of holograms: the EBL and FIB techniques

The final step of S-CGHs production is the fabrication of the designed pattern on a designated substrate. As previously said, the most common substrate of choice is silicon nitride (Si_3N_4). The two typical fabrication techniques used to make S-CGHs are Focused Ion Beam Milling (FIB-Milling) and Electron Beam Lithography (EBL).

4.2.1 FIB Milling process

FIB-Milling is a powerful tool to fabricate designed patterns. The FIB instrument generates a focused high energy beam of accelerated ions directed towards the sample surface to remove material by sputtering. Ga ions are the most widely used for this purpose however, Au, Ir, Ar, He, Xe, O, N and Si ions are available as well. A higher-Z element provides a higher milling yield, whereas a lower Z element offers a higher accuracy in reproducing the desired pattern. FIB-milling exploits the so-called *Knock-on sputtering* phenomenon. This is one of the main interaction phenomena between accelerated ions and a target substrate. It must be noted that, for this to happen, the ion's kinetic energy needs to be accelerated by a potential in the 1-50 kV range [47]. During the FIB-milling process, an incoming ion hits a surface atom, transfers part of its kinetic energy to it, so that the hit atom is displaced from its equilibrium lattice position and eventually collides with the neighboring atoms, which can result in their release from the substrate. The incoming ion after several impacts loses almost entirely its primary energy and can be trapped in the target substrate, leading to ion implantation and a change of the properties of the target substrate. In addition, FIB ions can also be exploited for imaging purposes; however, many FIB instruments, called *dual beam*, comprise a coupled SEM column that allows non-destructive electron imaging. While every FIB machine substantially works the same way, except for the generation of the ions, depending on the manufacturer there are differences in the procedure for the fabrication of a S-CGH. These differences lie mainly in the electronics and the software that manages the beam scanning or patterning. The authors' experience is mainly with FEI, now Thermo Fisher Scientific, instruments, so most of the discussion regarding the procedure that follows will mainly refer to such instruments.

As previously explained, the S-CGHs are first designed with the aid of a computer and dedicated software. The end result is usually an image of the pattern that needs to be reproduced. This image in some cases can be directly fed to the patterning software that comes with the dual beam machine, or it needs to be converted in a file format that can be read by the software. In the first case, the most common image file formats are .bmp or .png, in the second case vectorial (.dxf or .gdsII) or stream files (.str) are used.

Generally speaking, any file format fed to the software will be used to tell the FIB controller where to position the beam and how long to stay at a certain position. A pixel position in the image will be converted to a position in the coordinate system of the beam controller. In contrast, the pixel intensity is proportional to

the time the milling beam spends on that particular position, i.e., the *dwell time*. This last parameter is what one can use to select between these formats. Most of the afore mentioned image files are 8-bit ones, meaning that the vertical resolution in the milling is limited to a maximum of 256 levels of intensity. Therefore, if higher fidelity in the profile shape is needed, a different file format is required. This usually translates into the need to use vectorial file formats (.dxf and .gdsII) or the direct coordinate and milling time format (.str) where the resolution in z dimension is no longer a limiting factor.

An additional distinction arises between image, vectorial and stream files, is the order in which the pattern points are scanned. In the case of a picture or vectorial formats, the FIB pattern handling software allows the choice of scanning direction, e.g., line by line or column by column in different directions or spiraling. All software packages allow for the choice of the number of passes across the sample. The total milling time can, therefore, be subdivided into longer dwell time for fewer passes or shorter dwell time for a higher number of passes. These aspects will be analyzed in detail later in this chapter.

Optional procedure: Au coating

In a general observation of a hologram in LAD, a central spot and some lateral spots can be identified. The central spot is the center of the Fraunhofer plane which is also referred to as the 0^{th} diffraction order. The lateral spots are the actual nonzero diffraction orders, which arise from the periodicities on the sample. The portion of the electron wavefunction impinging on the S-CGH area, which contains the patterning periodicities, will be diffracted and will contribute to the intensity of the diffraction spots, i.e., the electron holograms encoding the wavefunction of interest. All the portions of the wavefunction impinging on the unpatterned areas in the surroundings of the S-CGH, along with the unscattered electrons and a contribution from non-ideal S-CGH fabrication, will contribute to the intensity of the central spot. In order to minimize the intensity contribution from the surrounding unpatterned areas, which can be detrimental during LAD as it may overlap and suppress the contribution from the S-CGH, it is possible to carry out the Au coating procedure. This procedure consists of first depositing a thick (≈150 nm) layer of Au by sputtering or evaporation followed by FIB removal of this Au layer only in the area where the S-CGH will be fabricated. The procedure is quite quick and easy with a downside which is an increase of the SiN surface roughness. This is mainly due to the roughness of the Au coverage which is subsequently projected onto the SiN surface after the Au removal by FIB milling. An alternative Au coating procedure can be carried out using EBL, which is less straightforward and more time-consuming. Yet, it allows to preserve the initial SiN surface roughness but may leave residues. The Au layer will then block most of the electrons impinging on the membrane at the unpatterned areas.

4.2.2 FIB-Milling Calibration

After the proper file format is chosen, the milling process requires calibration to be able to mill reproducible patterns with well-defined groove depth, and in the case of phase S-CGHs, to obtain the desired phase change. Multiple factors play a role in the milling yield, mainly the beam current, the dwell time, and the number of passes. The calibration process requires to produce a series of very simple patterns, controlled by one of the source files described before, constructed to have each the same size (i.e., number of pixels), but different milling time and different real pixel size. For example, it is possible to use a square-shaped pattern

Figure 31: Calibration pattern image. The white colour corresponds to the pixel of maximum intensity, i.e., the longest dwell time, while black is the "zero" intensity, so no milling will happen in these points. The grey area corresponds to an intermediate intensity (dwell time).

where half of the side is milled, e.g., twice as much of the other half. A pattern image with these features would correspond to that reported in Figure 31.

The next step is relatively straightforward: try to reproduce the pattern on a membrane of known thickness and let the FIB mill until holes start appearing at the white side of the pattern on the membrane until it completely breaks. By knowing the total pattern size, the beam current, the milling time and the physical size corresponding to a pixel, it is possible to evaluate the dose and therefore the depth milling rate for that specific pixel size and current. An example of this process is shown in the SEM images of Figure 32. Here, the total milling time is increased from (a) to (d) by increasing the number of repetitions and keeping the pixel dwell time equal to $10^{-4}s$. In the bottom part, it can be observed a slight bending (b), local milling through (c) and severe milling through (d). The milling depth in the four conditions can be measured, and an estimation of the milling rate can be calculated.

Another critical parameter is the pixel size. Each image or stream file loaded in the milling software is ultimately made of pixels. The pixel size is the area every pixel from the image will occupy on the substrate, and it is equivalent to the distance between neighboring pixels squared. The pixel area can be varied in many ways. For instance, instruments controlled by a Raith scan and control unit typically allow the user to choose the pixel size once the pattern image is loaded, to modify the pattern and to impose custom sampling conditions and milling mode on the designed pattern. TFS instruments, on the other hand, do not allow this

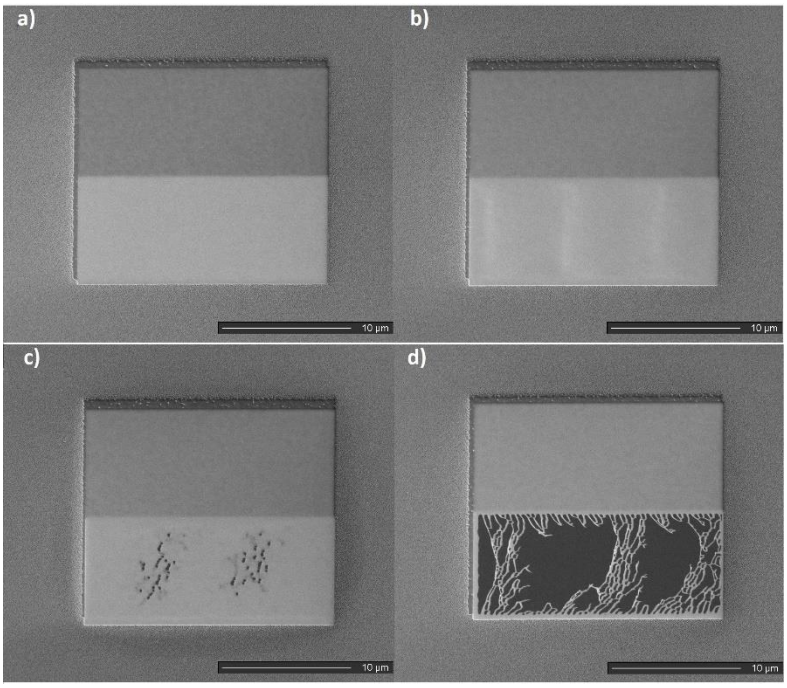

Figure 32: SEM images of the reproduction of the pattern in Fig.6 on a Si_3N_4 membrane. From a) to d) the total milling time and number repetitions have been linearly increased. The pixel size was ~ 30 nm, the ion current ~ 104 pA and the accelerating voltage 30kV. By repeating this process many times, it is possible to estimate the milling ratio of a machine. A low electron energy was used to enhance the surface sensitivity during imaging.

operation. Instead, they define the pixel position in the imaging reference frame. Therefore, for the same pattern, the choice of different magnification leads to different pixel sizes.

Once information is collected, it can be used as a starting point for actual pattern milling aimed at the desired phase shift. To be more accurate in the calibration, the suggestion of our group is to repeat the same procedure at different pixel sizes and ion beam current, in order to find the milling rate parameters for each setup. In theory, once calibrated, this should be valid up until the ion beam aperture, which defines the beam current and spot size, are healthy.

This method is very effective at determining the milling rate, and by increasing the number of tests, it is possible to statistically decrease the error. The rule of thumb in this case is repeating 4 to 6 times the procedure reported in Figure 32 for each of the ion currents that will mainly be used for patterning.

Apart from ordinary surface profilometry methods, such as Atomic Force Microscopy mapping, there are complementary TEM measurements that help further improve the fabrication depth accuracy. This method can be employed as a way to meticulously examine if the fabricated S-CGH works appropriately or not. These methods are LAD, Energy Filtered-TEM (EFTEM), and Low Mag Off-Axis Holography. While LAD is commonly available in modern TEMs, with camera lengths that can easily reach 1.4 km, EFTEM and Low Mag Off-Axis Holography are less commonly used. The first requires an additional energy filter. The second requires a biprism and a free lens control to work in a non-standard configuration that, if done carelessly, can permanently damage or break the biprism.

Using LAD, by fabricating a series of diffraction grating patterns identical to each other, where the only changing parameter is the overall milling time, it is possible to estimate the proper milling time by comparing the diffraction intensity in LAD. For example, suppose a S-CGH with a sinusoidal modulation is chosen. In that case, the pattern which will either suppress the central spot or have the highest intensity on the first order diffraction between all will be defined as the one properly milled. It is a good practice to start with big variations in the milling time used to be able to probe a wide range of possible values. Subsequently once the best fabricated sample is found, the process should be refined with a smaller range of variations. As a rule of thumb, this needs two to three iterations to find the best milling time. A clear problem with this approach is the fact that it is time consuming. Before a good calibration is achieved, at least seven to ten patterns need to be optimized, where the pixel size or the milling current must be changed from one run to another. Furthermore, care should be taken to avoid the 2π ambiguity when a very broad range of thickness is explored.

EFTEM mapping is a complementary technique, which can provide the real space three-dimensional image of the pattern. This technique exploits the inelastic interaction between the electrons and the sample: the scattered electrons lose a small amount of energy and this can be measured using the energy filter. The number of electrons that have undergone inelastic scattering compared to the number of electrons that have either undergone elastic scattering or any scattering at all, can yield a value proportional to the local thickness, using the log-ratio method [166]. This value, multiplied by the electron mean free path of the sample, provides the thickness value, which can be compared to the thickness aimed at in the milling process. In this way, it is possible to reconstruct an x-y map with additional thickness information. The thickness determines the phase shift. Therefore, it is possible to use the obtained image for computer simulations of electron beam propagation for knowing how the hologram's phase and amplitude information are expected in LAD.

As previously mentioned, another method to validate the quality of a S-CGH and calibrate the FIB machine is to use Low Mag Off-Axis Holography. This method requires setting up the TEM in a non-standard configuration and because of this it may not always be possible to perform. This technique allows us to directly measure the phase and amplitude of a large region of interest of a sample. However, the region of interest is fairly limited (usually not more than 30µm ×30 µm), it requires to mill a large window near the S-CGH to allow the unperturbed wave to pass through. Almost always, there will be a linear gradient in the phase image, which needs to be carefully removed in the post-processing. Even in this case measurements and analysis will be quite time consuming as for the previous two techniques. The calibration process needs to be repeated every time the substrate material on which the S-CGH needs to be produced is changed. If more complex patterns are to be milled, these methods can provide valuable information also for their

fabrication. It is advisable to carry out a new calibration process for every new pattern at a given experimental condition if high accuracy is needed.

Figure 33 summarizes the fabrication process of an S-CGH by FIB milling (Figure 33a-d). It also provides an intuitive algorithm/recipe for the calibration process of the FIB machine (Figure 33e). For simplicity and clearness, only the main steps are presented.

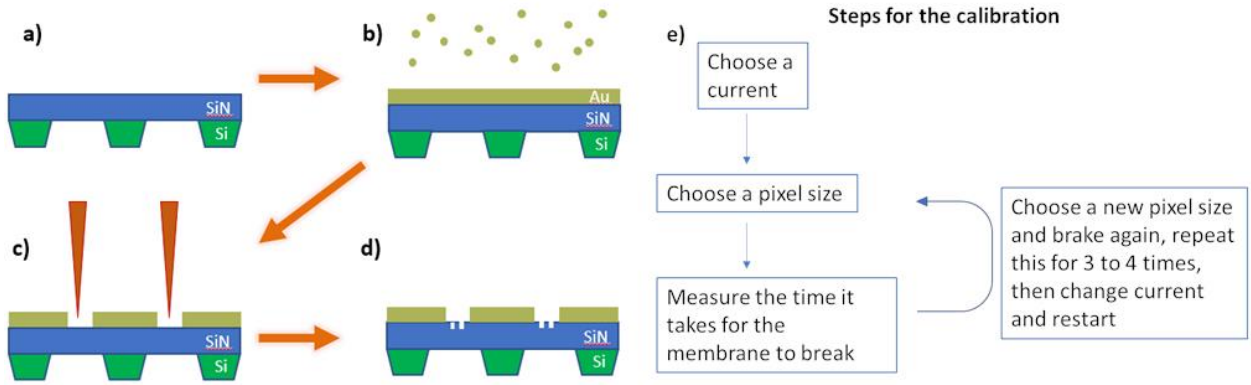

Figure 33: Schematic diagrams showing the typical step in the fabrication process of an S-CGH by FIB milling: (a) a fresh device; (b) Au evaporation; (c) Au removal and FIB patterning; (d) grooves in the membranes. A simple and intuitive algorithm for the calibration process is provided in (e).

4.2.3 FIB Milling pattern reproducibility optimization

Once the milling process is calibrated, it is possible to start the S-CGH fabrication. The calibration process focuses on trying to estimate the milling rate of the FIB instrument as best as possible, while the optimization process finely tunes the parameters for optimum result. Parameters that can be optimized include the beam current, pixel size, distance and dwell time, number of passes or repetitions of the pattern and the scanning strategy. Even the pre-S-CGHs milling membrane thickness will influence the result. In the next few paragraphs, a few tips and tricks will be described.

Optimization of the ion current

Choosing the right current for a particular fabrication is a crucial parameter in the process. Selecting the proper value for the current is closely related to the ion probe size, which ultimately defines the hologram resolution. The main parameters that need to be considered during the choice are the total milling time, estimated by the patterning software, and the pixel size. The pixel size is important because it is related to the intrinsic resolution of the S-CGH, and the finer the details present in the profile, the smaller the pixel size needs to be. In general, a higher pattern resolution is desirable; however, there is a limit on how small the pixel size can be since a too high-resolution or too big hologram area can result in a large patterning file that cannot be handled by the patterning software. A pattern resolution higher than the milling resolution will not be properly reproduced in the S-CGH.

Ga^{2+} sources one can have from high-end instruments, at the lowest current, have a spot size of approximately 5 nm or less. This size scales as the square root of the current, but the actual patterning

resolution depends on other factors such as the local milling time or instabilities, resulting in a larger effective spot size. For a higher resolution a lower current is needed, at the cost of longer patterning time, as the sputtering rate depends on the current. However, long continuous patterning times may incur in stage drift and beam defocus, except in cases of machines with interferometric stages and high beam stability. Therefore, normally a trade-off between ion current and the total patterning time must be found. As a rule of thumb, patterning times longer than two hours are not recommended. For these reasons, before starting the fabrication, it is worth carefully choosing the proper current to have the best resolution and a reasonable patterning time.

Optimizing the dwell time, repetition number, pixel distance and scan direction

The local milling time, or dwell time, is one of the parameters that can be optimized along with pixel distance (if available), number of repetitions and patterning strategy or scan direction. The dwell time influences the end shape of a milled pattern. Figure 34 shows a box pattern. One can appreciate the difference between short dwell times with many repetitions and long dwell time with few repetitions, keeping the total dose constant. The first case results in a rectangular box profile with mild redeposition on the sidewalls (Figure 34a). On the other hand, if an extended dwell time and few repetitions are chosen, the end result will be a sloped profile (Figure 34b) with dramatic redeposition effects along the horizontal direction of the serpentine scan [47].

On the other hand, having too many repetitions can be detrimental too. In fact, during the homing phase at the end of a repetition, though rarely, a small drift of few nanometers can occur, leading to a smearing of the end result. Again, a trade-off between number of repetitions and dwell time must be chosen, long dwell times and a high number of repetitions must be avoided.

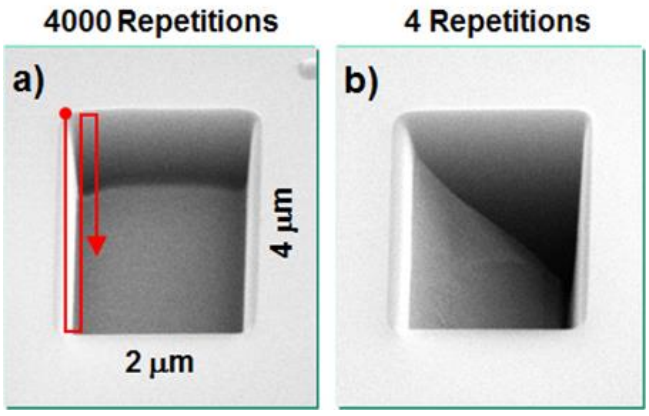

Figure 34: SEM image (tilted view) of a box pattern milled with a) a short dwell time and many repetitions, b) long dwell time and few repetitions. The serpentine beam scan is shown (red line).

It is also worth studying the effect of the pixel-to-pixel distance, how the end result might change if two pixels overlap of a different amount. It is evident that very high pixel-to-pixel distance, i.e., highly negative overlap, is detrimental as the end result of the S-CGH would be a dotted pattern. On the other hand, a too-short

distance could increase the patterning time and the pattern file size without improving the S-CGH final resolution as other factors (mainly ion current, i.e., probe size) more prominently affect it.

The last parameter that may be worth looking into is the "scanning strategy", or the path the beam is following to reproduce the pattern. The most common cases are the so-called zig-zag scanning, like the ones shown in Figure 35(a) or alternatively in Figure 35(b). An alternative to the more conventional "zig-zag" is the spiral patterning method, which is shown in Figure 35(c) [181]. It is important to try to use the best possible scanning strategy because scanning direction and path contribute to determining where the redeposited material will mainly go. In the conventional zig-zag scanning patterning the redeposition will be mainly found on the opposite side to the scanning direction, as shown in Figure 34b. For example, when long rows are to be patterned, it is suggested to scan the beam along the rows instead of perpendicular. The unconventional spiral strategy, with its continuous "back and forth" motion should allow for a "cleaner" result. However, few examples are presented in the literature [181].

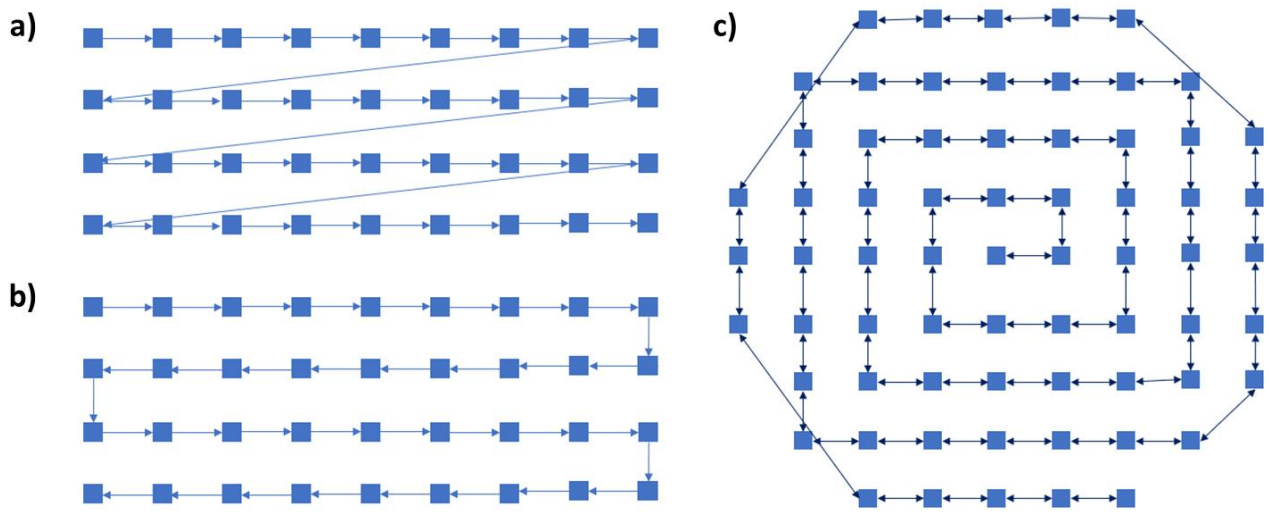

Figure 35: Patterning strategy examples, it is important to mention that the distance between the symbolic pixels has been increased to ease the visualisation.

4.2.4 EBL for S-CGH

Unlike FIB, EBL usually requires a series of steps and controlled processes to achieve the final result. With this technique it is possible to reproduce features as small as a few nanometers and it opens the possibility to mass produce S-CGHs. The usual workflow for the production of a S-CGH is illustrated in Figure 36. In this case, a negative tone resist is used. It is also possible to use a positive resist paired with Reactive Ion Etching to transfer the pattern, however, it usually leads to poorer results.

A complete set of calibration procedures is also needed for EBL. However, in this case, it is less time consuming when one wants to use a negative tone resist as HSQ that polymerizes in a SiO_x layer when illuminated by the e-beam and possesses a V_{mip} similar to Si_3N_4. Following the standard procedure, to select the best dose one can do a dose matrix of a small feature of the pattern one wants to reproduce. Moreover, no milling rate is required in this case since the resist thickness dictates the peak-to-valley height.

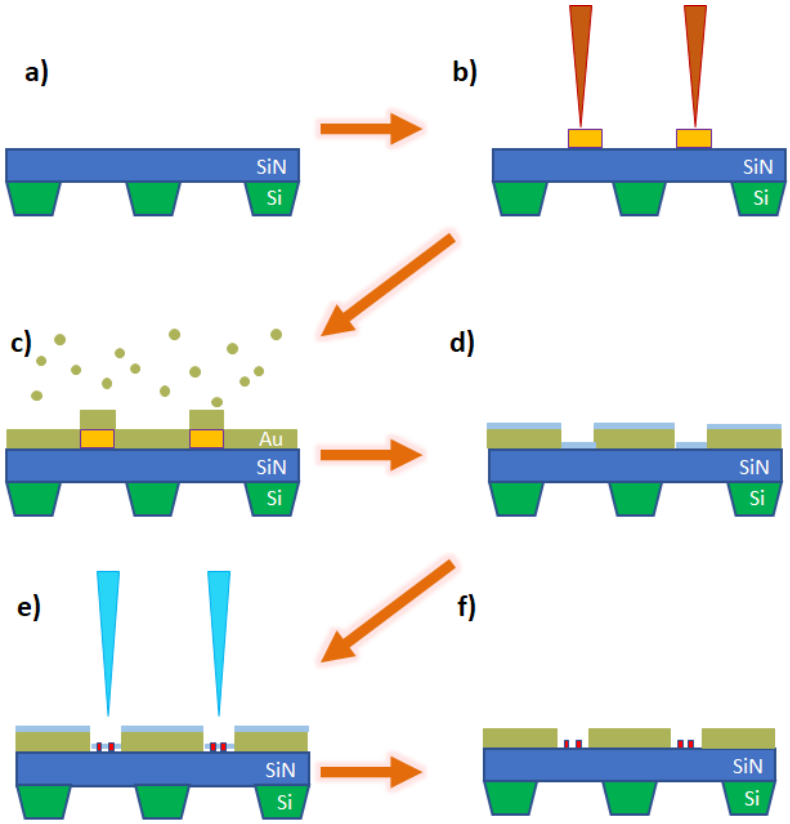

Figure 36: Schematic diagrams showing (a) a fresh device; (b) EBL and developed resist (a negative one); (c) Au evaporation; (d) lift-off and hydrogen silsesquioxane (HSQ) spin coating; (e) EBL; and (f) developing the HSQ.

The steps for the S-CGHs preparations via EBL can be summarized as follows. First, a layer of negative tone resist is spin-coated on the TEM membrane, patterned in the desired S-CGH enclosure shape and then developed. Then, a layer of gold or any other metal (with a large atomic number) is evaporated on the device and even in this case the metallic layer is used to block off a portion of the incoming beam. It is highly suggested to use an adhesion promoter layer of either Cr or Ti and then deposit the metal of choice to ensure that the coating will stick to the surface. The device is then immersed in a resist remover to lift-off the metallic layers that were sitting on the previously developed resist, so that the canvas for the S-CGH is finally ready. HSQ or the resist of choice can now be spin coated up to the desired thickness, patterned and then developed. At this point, a few nanometers thick layer of metal or amorphous carbon can be flash evaporated on top of the developed pattern to balance the generation of secondary electrons during the TEM use. More details on the fabrication process and all the steps and exact parameters can be found in the paper by Mafakheri et al. [113] and many others related to the EBL technique.

Some of the EBL technique's limitations are evident: the thickness is fixed so one needs to fine-tune the spin coating process to achieve the required thickness consistency for the phase shift. Secondarily, the pattern profile will either be squared or sinusoidal. Usually, it is difficult to achieve a blazed profile. Most importantly, as explained above there are multiple steps required to complete the process and the devices are small and fragile, which means they have to be carefully handled during processing. The advantages of the EBL however are manifold and can outshine the disadvantages. The Si_3N_4 membrane thickness can be reduced up to even

15 nm. While normally in FIB milling the membrane thickness should be at least 75 to 100 nm before patterning, in this case, the process consists in reproducing the pattern on top, meaning that the Si_3N_4 is just a support layer. Having a thinner membrane leads to a lower number of inelastically scattered electrons and subsequently a reduction in background noise and absorption. Moreover, a smaller thickness allows for a lower electron dose during patterning and fewer secondary electrons will be generated in the resist-supporting substrate. Fewer secondary electrons open the possibility to sub 10 nm features reproduction and consistency if the process is well optimized.

As just suggested, even for the EBL made S-CGHs it is possible to tinker with the basic recipe to obtain finer details. Due to the large number of required steps, fabrication procedure is a tedious process of trials and errors. Some possible improvements can be made in the following procedures:

- changing the pre-patterning bake temperature or adding a post pattern bake.
- precisely search for the proper dose and using proximity correction for improved results.
- experiment with the development temperature and time. Some resists lead to a higher contrast in the end result when developed at a colder temperature and slightly longer times than room temperature development [182], while others behave in a similar manner when developed at higher temperatures [183].
- understand very well the chemistry of the resist to find the optimal developer, which might be different from the one suggested by the producer.
- test different support layer thicknesses, nowadays there are a significant number of companies producing Silicon Nitride supporting layers for usage inside TEM.

Some of the inherent limitations of the EBL, and even FIB milling, have been recently overcome by employing a thermal Scanning Probe as the patterning probe, instead of an electron probe. This improvement resulted in higher accuracy and control in the patterning depth and morphology [184].

4.2.5 Experimental limiting factors for the use of the Synthetic Holograms in microscopy

The use of S-CGHs can be very effective in the realization of complicated phase patterns as it is so far the only tool that provides proper and arbitrary wavefront control. On the other hand, the drawback of synthetic holograms is that they are static and even the exchange with a different one in an aperture plane of the microscope usually requires to break the vacuum of the electronic column. The alternatives such as a “hijack” of the multipoles of spherical aberrations correctors [185], electrostatic field [139], or different forms of programmable phase plate [186,187] are far from reaching the same level of arbitrary shaping with the same number of pixels. Thus, thin synthetic holograms are preferred for experiments where a well-known effect is sought for. However, inserting material along the electron path has its downsides, especially for phase S-CGHs. Major issues which one must consider include 1) introduction of inelastic effects and decoherence

along the beam propagation 2) reduction of the beam intensity 3) contamination, damaging and aging of the device under beam exposure and 4) charging of the device during operation.

As for the first point, one must bear in mind that the thin membranes used as patterning media for the S-CGHs, no matter how good the quality and the fabrication are, have local thickness variation at the scale of a few nm. Therefore, a frosted glass effect exists that somehow is similar to the effect of light crossing a turbulent or inhomogeneous medium. So, even the elastically scattered part of the beam will have a lateral spread of momentum distribution due to the membrane. Furthermore, different forms of inelastic scattering reduce the beam current and increase the lateral distribution.

The beam on the long term produces an alteration of the groove profile that can alter the desired phase profile. The effect is more prominent if the synthetic hologram is in the condenser plane where the current is higher. Experimentally, we found that the quality of the synthetic hologram may deteriorate very quickly due to contamination (local carbon deposits can already form in a couple of days). In contrast, actual damage from for example knock-on effects and irradiation is generally slower (minor profile alterations were visible after one week of intensive use). It is very important to take care of the vacuum quality in the TEM column and to be very careful during operations such as the sample exchange that can increase the probability of contamination. It is also quite important not to concentrate the beam on a spot on the S-CGHs in any phase of the operation. The most serious problem is potentially the charging issue. SiN is a very insulating material and it can be difficult for the charge generated by the impinging beam to dissipate over the typical length of the synthetic hologram in a short time. As previously mentioned, most of the membrane on which the S-CGH is patterned is covered with a relatively thick gold layer, and the electrons should only pass through the S-CGH transparent area. The gold layer is quite efficient in removing the charge, but the problem remains (to a lower extent) in the uncovered area. Experimentally we found that, at steady state, the hologram gets a charge density such that the projected potential is parabolic. In this condition a focusing effect is superimposed to the hologram phase. It is possible to compensate for such effects using the microscope lenses. Unfortunately, this compensation depends on the electron dose, i.e., the higher the dose, the greater the effect. Finally, with large synthetic holograms and unfavorable materials such as HSQ, it has been noted that the steady state can never be reached, and the beam's phase continuously oscillates. The use of a hologram can, therefore, be complicated.

Possible solutions to this specific problem are the use of more conductive membranes like carbon ones or coating both surfaces of the S-CGH with a thin metal or carbon layer. Using smaller synthetic holograms is also helpful. Finally, an alternative route is to return to the amplitude S-CGHs approach where virtually all the elements are conductive [100]. In this case the presence of thin material bridges exposed to vacuum makes the structure mechanically unstable and difficult to fabricate. A common approach is to substitute the separate lines with a cross grating. This has the effect of dispersing the diffraction orders in two directions

with an overall reduction of the efficiency in the order of interest and greater difficulty in isolating it. This can clearly be appreciated in Figure 37.

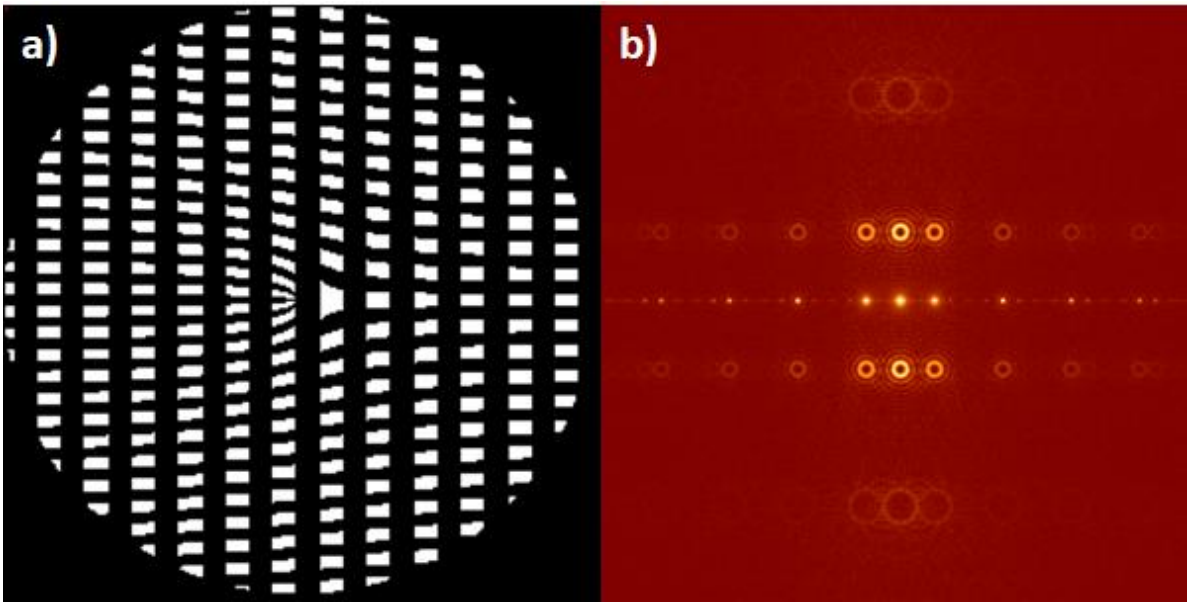

Figure 37: a) Example of an amplitude S-CGH with a grid-like structure for improved mechanical resistance and b) the resulting diffraction pattern that forms a 2-D array of beams. This idea is explained in greater details in [100].

4.3 - Examples

4.3.1 Design of a Phase S-CGH for the generation of Electron Beam Vortices carrying Orbital Angular Momentum.

The possibility to generate Electron Vortex Beams (EVBs) was first discovered between 2009 and 2010 by three different research groups: by using a "spiral phase plate" constructed by using thin films of graphite [20], or by using S-CGHs with a pitchfork design [21,22]. In the first experiments EVBs were generated by using amplitude S-CGHs or similar structures. Since then, most research groups try to use phase S-CGHs or mixed amplitude-phase ones as they inherently have a higher efficiency. New methods for the generation of EVBs have been presented in [115,120,121]and nowadays the topic has matured enough that most of the efforts are directed toward ways to measure the OAM value and increase the number of novel applications in the field of plasmonics, magnetic materials and in general chiral structures such as proteins.
In a circular symmetric reference system, an EVB has an angular dependent helical phase term described by the mathematical expression:

$$\varphi(l,\theta) = \ell\theta \tag{193}$$

where $\hbar\ell$ is the OAM eigenvalue of the Schrödinger equation solved in cylindrical coordinate and ℓ it is known as the topological charge or OAM quantum number, and θ is the angular coordinate.
So that the wave function of a generic EVB is given by:

$$\Psi_{helical} = A_0 e^{i\ell\theta} \tag{194}$$

Different strategies can be used to create EVBs and here are illustrated some of the most prominent ones. Further details regarding the topic of EVBs and Vortex beams in general can be found in various reviews that have been written by different groups involved in the field, for example [67,81,100–102] .

4.3.1.1 Spiral design

The simplest way to generate an EVB using a S-CGH is to design an inline [109,110] phase S-CGH with a spiraling/helical form, similar to the one in Figure 1, where thanks to the smoothly varying thickness it is possible to continuously tune the phase shift imprinted to the wavefront of the outgoing beam.
This design, in its simplest form, is an inline S-CGH and the realization requires a good control of the fabrication process for the reasons explained in Chapter 1 and 2. However, it is quite straightforward once designed and fed to a well-calibrated machine. In fact, to generate an EVB with topological charge ℓ for a spiral phase plate it is necessary that the total phase shift along a complete revolution is $\Delta\varphi = \ell \cdot 2\pi$.
A typical design of an EVB with a spiral phase is reported in Figure 38(a). It is possible to notice six angular sections: there are six phase ramps where the phase shift goes from 0 to 2π, which means that the outgoing

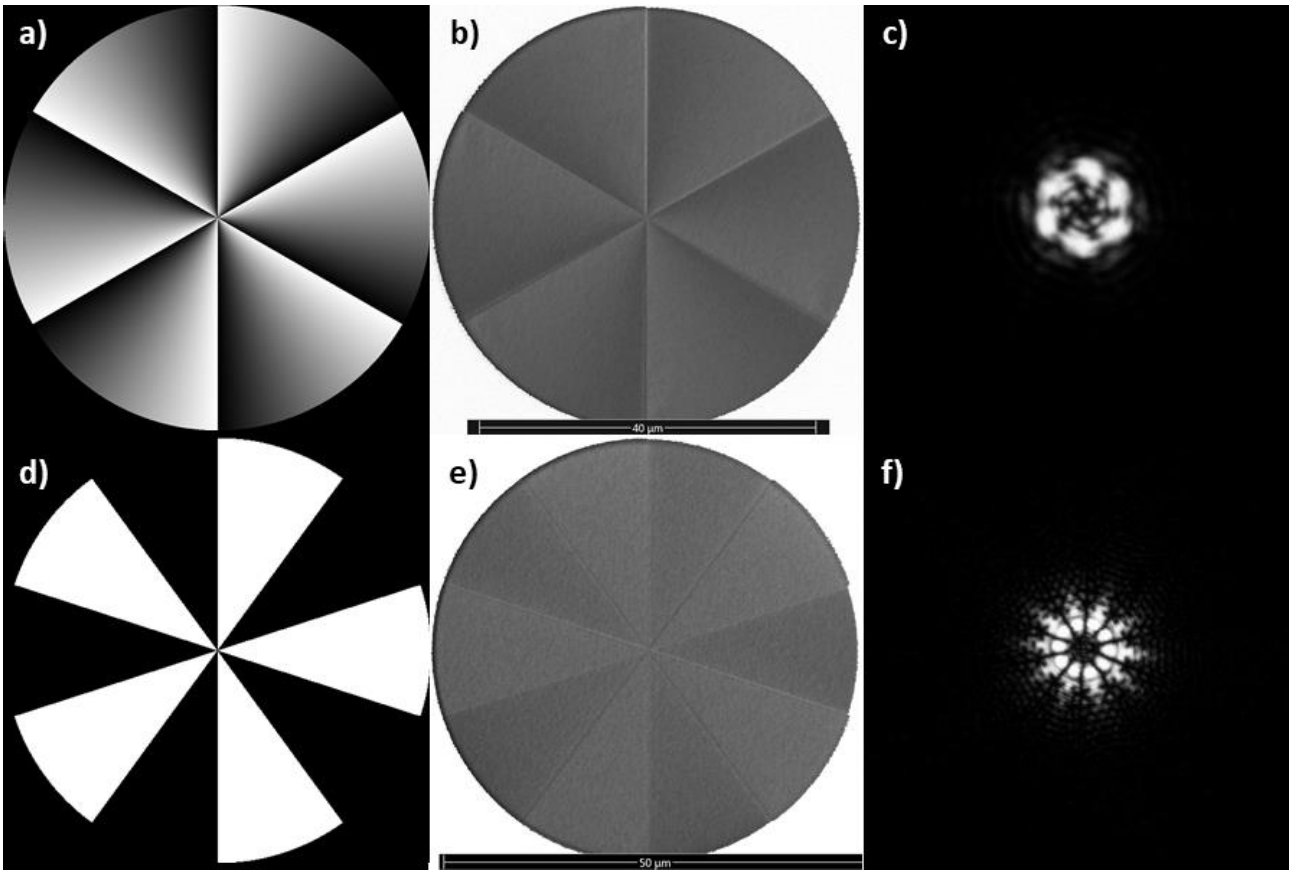

Figure 38: a) phase of an l = 6ħ EVB, used to fabricate a Phase-S-CGH with spiral/helical design for EVB generation. Here the phase smoothly varies from 0 (black) to 2π (white), so that the total phase after a complete round goes from 0 to 12π. (b) SEM image of the phase S-CGH resulting from a) and c) the experimental EVB generated in the Fraunhofer plane. d) and e) are respectively phase and corresponding phase S-CGH for an EVB with $\ell = \pm 5\hbar$, while f) shows the experimental EVB.

EVB will carry an OAM value equal to ℓ = 6. Furthermore, this design allows to generate a superposition of EVBs. When the two superimposed EVB have opposite L the beam has actually no azimuthal current and is called “petal beam”.

For example, it is possible to design a hologram to generate electron beams being in coherent superposition of ℓ=-5 and ℓ =$+5$. This can be achieved by first summing the wave functions of the two EVBs, one with $\ell = 5$ and one with $\ell = -5$, and then by calculating the phase of the resulting wave function, it is possible to fabricate a phase S-CGH. Mathematically the phase is just $\Delta\varphi = arg\,(sin\,(l\,\theta))$, so it is assuming alternatively the values 0 and π. The phase of the beam carrying the OAM superposition ℓ=± 5, as shown in Figure 38(d), where in this case the white colour corresponds to a phase shift of π with respect to the black areas. For a generic EVBs generator with the spiral design the enclosure is a circle, similarly to a conventional TEM aperture and the physical dimensions of this type of S-CGH range from $10\ \mu m$ up to $50\ \mu m$.

4.3.1.2 Pitchfork design

The pitchfork design is the simplest one along with the spiral one, and it can be used in both amplitude S-CGH [21,22] and phase S-CGH [113]. A pitchfork S-CGH falls in the category of off-axis S-CGH, meaning that the EVBs will be generated in the nth diffraction order, where n can vary from 1 to infinity.

The origin of this design can be found by calculating the interference pattern between a plane wave, $\Psi = A_0 e^{i(k_x x + k_z z)}$, and an helical one at the $z = 0$ plane:

$$I = 2|A_0|^2(1 + cos(k_x x - \ell\theta)) \tag{195}$$

from which it is possible to find the phase term of the interference wave and design the pitchfork S-CGH by borrowing the mathematical description for the different kinds of S-CGH. As anticipated in section 4.1.4.1, in order to generate a pitchfork S-CGH the argument of the profile function is

$$\alpha(x, y) = \ell\xi + 2\pi x \tag{196}$$

where x is one of the two in-plane coordinates, $\xi = ArcTan\left(\frac{y}{x}\right)$ and ℓ is the topological charge. The planar Cartesian coordinates x and y are expressed in units of the grating spatial period Λ.

In Figure 39 the 2-D representations of the profile function $f(\alpha)$ for a pitchfork design with $\ell = 2$ are reported. Each pattern has been obtained by combining the generic profile functions described in section 4.1.4.3 and Eq.196, so that:

- Figure 39a: $f_{sqrd}(\alpha) = \frac{1}{2}(1 + Sign(sin(\ell\xi + 2\pi x)))$;
- Figure 39b: $f_{cos}(\alpha) = \frac{1}{2}(1 + cos(\ell\xi + 2\pi x))$;
- Figure 39c: $f_{trian}(\alpha) = \frac{1}{\pi}(Sign(sin(\ell\xi + 2\pi x)))(\pi - Mod(\ell\xi + 2\pi x, 2\pi))$;
- Figure 39d: $f_{blzd}(\alpha) = \frac{1}{2\pi}\big(Mod(\ell\xi + 2\pi x, 2\pi)\big)$.

This design is versatile as it can be used to generate both low OAM EVBs and high OAM ones. However, in the latter case, the center part has features so small (in some cases even smaller than a pixel) that are almost impossible to reproduce by either one of the two fabrication techniques discussed before. Hence, a common strategy is to mask out the center part up to a certain radius. The circular mask will affect the transmitted efficiency by consistently reducing it. In any case an EVB with the correct OAM value will be generated as reported in [113].

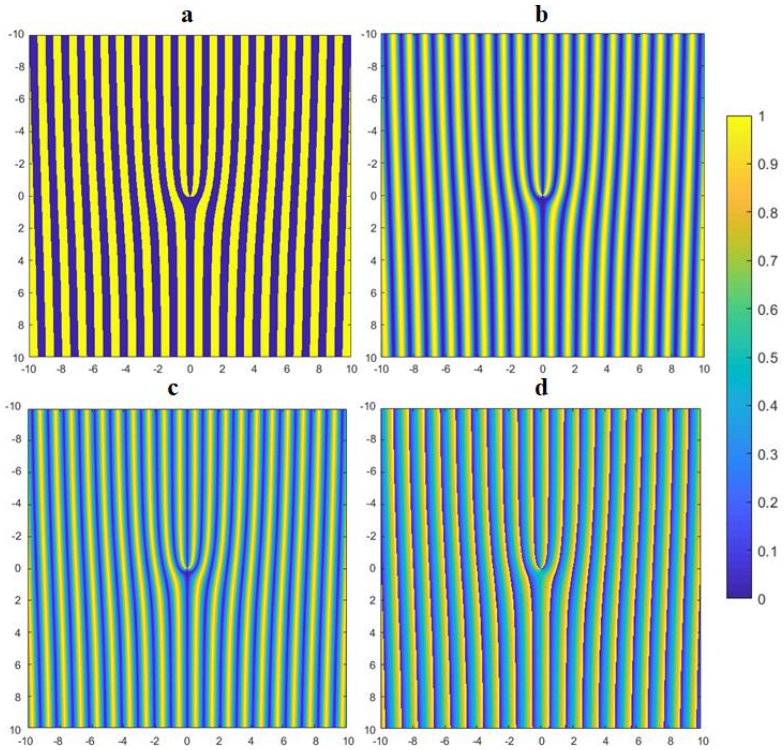

Figure 39: the different designs of a pitchfork S-CGH with $\ell = 2$: a) squared design, b) cosinusoidal, c) triangular, d) blazed. The colorbar is common between all of them and it represent the value on the pixel of $f(\alpha)$.

4.3.1.3 Case study: optimizing and understanding a real blazed phase S-CGH with a pitchfork design.

In recent years, our research group delved into optimizing the fabrication process of a blazed phase S-CGH via FIB, and the design of choice has been the pitchfork with $\ell = 1$ [188]. We aimed at reaching the highest diffraction efficiency on one of the two first diffraction orders (100% in this case, see Table 5) converging most of the intensity in the beam carrying the desired amount of OAM.

To reduce the number of possible variables most of the parameters were kept constant and only the number of passes and maximum dwell time were changed in order to tailor the phase shift and approach 2π. First, the number of passes was varied for a rough optimization, then the maximum dwell time for a finer one.

The parameters that were kept constant and their values are reported in Table 6.

S-CGH diameter	Ion beam current	CGH resolution	Step Size	Magnification
$20\ \mu m$	$\sim 260\ pA$	$1024 \times 1024\ px$	2	10400 X

Table 6: Patterning parameters that were kept constant during optimization

Figure 40 shows both the CGH and one of the many fabricated S-CGHs, in this case, the best performing one. The patterning parameters which led to the best performance, other than the ones reported in Table 6, are:

- Number of Repetitions: 8 passes.
- Maximum Dwell Time: $91.6\ \mu s$.

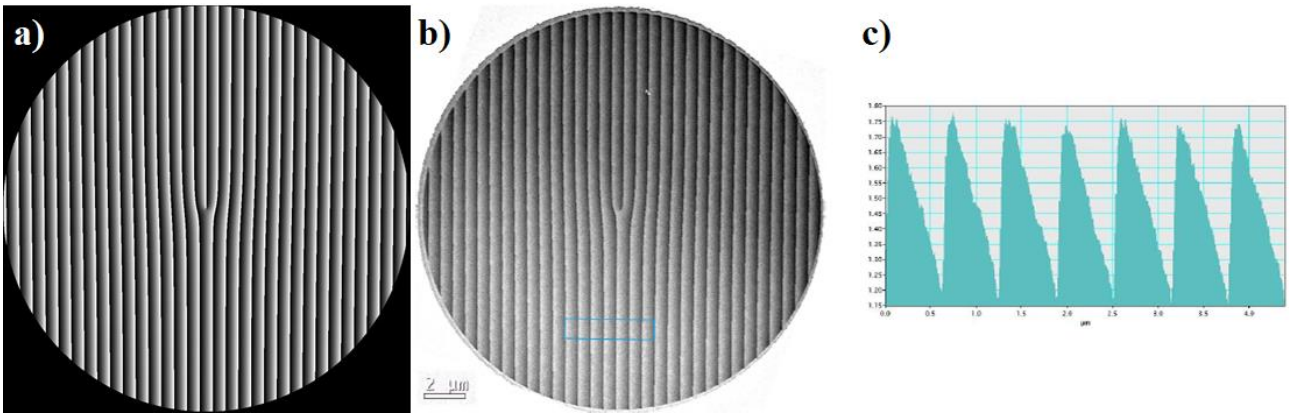

Figure 40: a) CGH of the blazed $\ell = 1$ pitchfork, b) EFTEM thickness map image of the fabricated S-CGH and c) profile of the region evidenced by the light blue rectangle.

These numbers can slightly vary from one FIB machine to another, but also from one fabrication session to another, as factors such as lab environment, vacuum quality and machine characteristics also influence the fabrication process.

From the EFTEM image (Figure 40b) and the profile of the region in the rectangle (Figure 40c) it is possible to understand that the pattern in-plane periodicity is roughly $\sim 600\ nm$ and the distance between peak and valley is $\sim 70\ nm$. This is slightly bigger than the required one, which for 300 keV electrons is $\sim 64\ nm$ as reported in Table *3*. Furthermore, it is possible to notice that as far as the shape of the peaks is concerned they well approximate the ideal shape of a blazed profile, but differ one from the other, the troughs are sharp enough, but the peaks are almost all blunt.

All these effects show some of the FIB limitations and are connected to the reduction in the diffraction efficiency that we measured. In fact, the best performing sample was able to achieve 66.22% of the transmitted intensity on the +1st diffraction order, as reported in Figure 41.

From Figure 41 it is possible to observe that the experimental diffraction intensity distributes between the various orders in a different way from the one found in Figure 27b or Figure 29d.

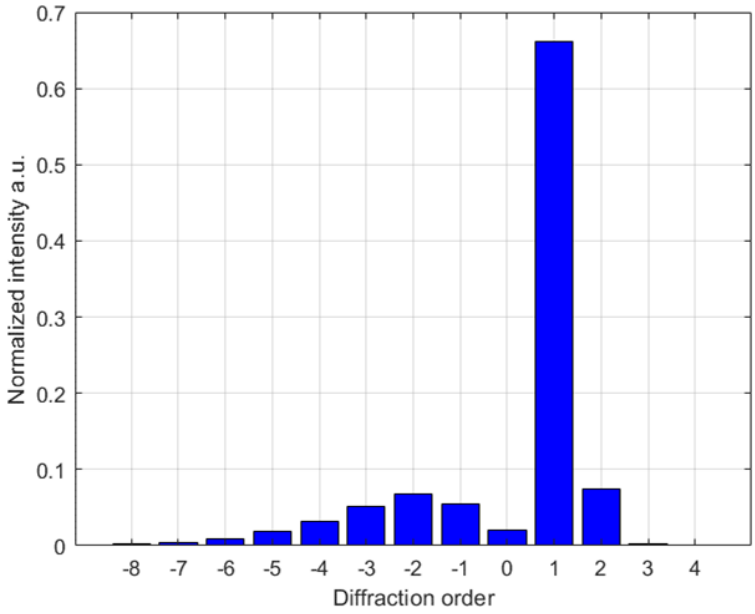

Figure 41: Experimental diffraction intensities distribution of the best-working sample, the total intensity has been normalized to 1.

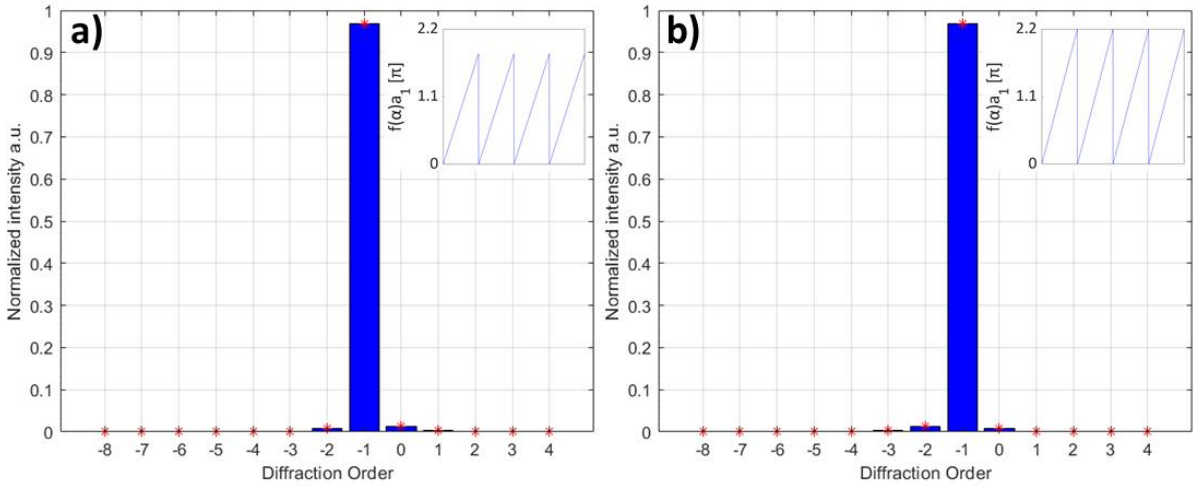

Figure 42: Diffraction intensity distribution in case of a $\pm 10\%$ phase mismatch, a) -10%, b) $+10\%$. The insets show the groove profile for each case.

Through simulations, we investigated the possible cause of this behavior: first, we looked at the effect of a non-ideal peak-to-valley phase difference by recalculating the intensity distribution in case of a $\pm 10\%$ phase mismatch for an ideal phase S-CGH. The effect of the mismatch is shown in Figure 42. It is noticeable that even a 10% mismatch influences almost negligibly the diffraction intensity distribution. This means that the intensity distribution measured in the experiments arises from something else. Also, absorption definitely plays a role in the diffraction intensity. However, as it is possible to evince from the end of section 4.1.4.3 and specifically from Figure 27b, absorption mainly decreases the total transmitted intensity and slightly to redistribute the intensity evenly between the orders.

Even by considering the effect of both the absorption and the phase mismatch it is still impossible to reproduce the same intensity distribution. We then focused on the profile shape of the fabricated S-CGH. As previously mentioned, by carefully looking at Figure 40(c), it is possible to notice that the actual shape is more similar to a scalene triangular profile, not of a blazed one. The scalene triangular profile function is:

$$g(\alpha) = \begin{cases} Mod\left(\frac{1}{s}\alpha(\vec{\rho}), 2\pi\right) \ \ for \ \ \alpha(\vec{\rho}) < s \\ 1 + \frac{s}{(2\pi - s)} - Mod\left(\frac{\alpha(\vec{\rho})}{2\pi - s}, 2\pi\right) \ \ for \ s \leq \alpha(\vec{\rho}) < 2\pi \end{cases} \tag{197}$$

Like the rest, this profile function is normalized between 0 and 1, and has its maximum for $\alpha(\vec{\rho}) = s$. This tells us that, the further s is from 0, the further we are from an ideal blazed profile.

Figure 43 reports the intensity distribution for $s = 1.1$. It is evident that, even though we purposely accentuated the shape difference by choosing such a high value of s, in the end, it is mainly due to the imperfections of profile shape that leads to the experimentally observed intensity distribution. Furthermore, even the small differences between the shape of one "tooth" and another lead to the spreading of the diffraction intensity between the various orders, but this effect is quite difficult to simulate.

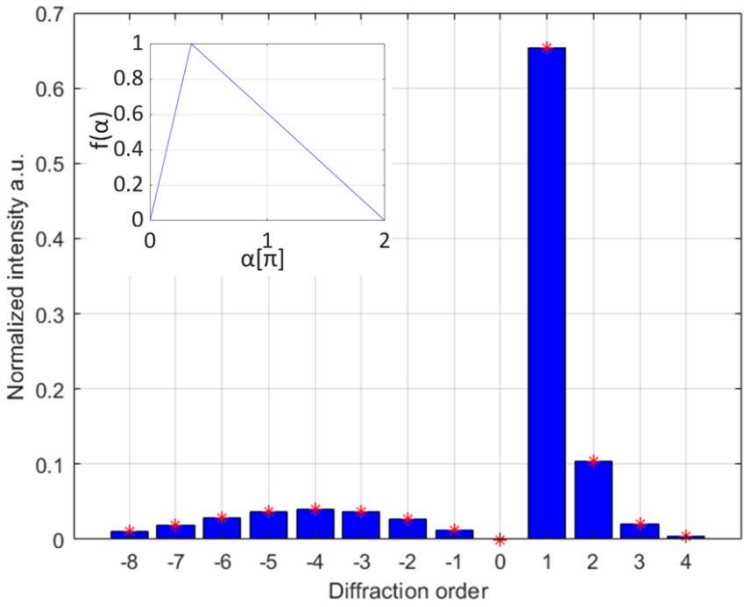

Figure 43: Diffraction intensity distribution for the scalene triangular profile, with s=1.1. In the inset it is possible to appreciate the groove profile over 1 period.

In summary, from dwelling in the optimization process of a real blazed phase S-CGH, we have been able to model and discover most of the main factors that limit the result. Unfortunately, most of them are related to the inherent limitations of the fabrication process. In fact, the imperfections of the shape and the phase mismatch are usually due to the instabilities of the FIB machine during operation. It is clear that at this stage optimizing the lateral resolution of the FIB machine can bring a major improvement compared to optimizing the phase shift. Moreover, even though rare, a stage shift (thermal or mechanical) of even a few nanometers in the middle of the fabrication procedure can compromise the result. It is possible to reduce or even remove some of the limitations by either using a FIB machine designed ad-hoc for S-CGH production or on a small scale by rethinking the fabrication steps, perhaps by increasing them. For example, gas assisted FIB milling has shown that it is possible to increase the reproduction fidelity and patterning speed of blazed profiles [189]. On the flip side, some effects that arise from inelastic scattering will always be present no matter the workarounds. For example, as already seen, absorption reduces the total transmitted intensity and modifies the intensity distribution. Even diffuse scattering by superimposing background noise blurs the diffraction pattern and further deteriorates the efficiency.

4.3.1.4 Generation of EVBs using Gaussian beams

The vortex beam generators defined above are characterized by a hard cut aperture in the hologram plane. This family of beams are sometimes referred to as “Hypergeometric beams” [190]. In optics, a more suitable vortex beam class has been derived starting from a member of the Gaussian beams family: the so-called Laguerre-Gaussian (LG) beams. Exact Gaussian beams are characterized by a flat phase wavefront at $z = 0$ and a definite amplitude structure, i.e., planes perpendicular to the optical axis are equiphase surfaces. An in-depth mathematical description can be found in the book by Robert Guenther on optics [191], and an

important parameter is the Gouy phase term, which is closely related to the transverse confinement of the beam and induces an anomalous behaviour of the phase of the beam when it passes through focus [94,97,192,193]. In a TEM, an exact Gaussian beam or coherent Gaussian beam cannot be straightforwardly obtained. It requires specific arrangements. However, Gaussian-like beams that only reproduce the intensity of an exact one can be readily obtained by converging the beam.

As previously introduced in Section 4.3.2.2 Laguerre-Gaussian beams are of greater interest with respect to the simpler Gaussian beams as they are the solutions of the paraxial Helmholtz equation in cylindrical coordinates. LG beams are of interest to scientists involved in both fields of magnetic materials and structured waves. For example, the LG wavefunction is functionally similar to the Landau states wavefunction, the eigenstates of the propagation of electron beams inside a constant magnetic field [98]. By properly tuning the LG beam waist it has been experimentally demonstrated that it is possible to couple them to Landau states [99]. Furthermore, an LG beam was used as an example to demonstrate that it is possible to use paired S-CGHs for almost direct phase retrieval of EVBs, and structured beams in general, in the Fraunhofer plane [194] Pure LG beams can ideally be generated by means of mixed S-CGH [195], the principle of which has been described previously in section 4.1.5.1. The design and fabricated mixed S-CGHs are reported in Figure 44. Here two experimental examples of LG beams with different characteristics are shown. While the first (Figure 44(a-d)) is a pure LG_0^{10} mode that presents a simple circular structure. The second (Figure 44(e-g)) is a more complicated case where two states with different OAM and p quantum numbers have been coherently summed together. This gives a superposition of LG modes having different radial and azimuthal indices, which can be considered as an interesting beam. The phase in Figure 44(f) is the theoretical phase obtained by Fourier transforming the thickness profile of the hologram. It shows all the complexity of the beam. This can be therefore considered a proof of the power of the technique of amplitude and phase encoding in a single S-CGH for Laguerre-Gauss beams generation, and more in general EVBs generation. Furthermore, in Figure 44(d) it is noticeable how there are no intensity ripples neither inside or outside, that are normally present in the EVB generated with the spiral design described in section 3.1.1. LG beams generation has also been reported in the literature with slightly different techniques [196].

Another interesting fact is that LG beams are a solution of the paraxial Helmholtz equation in cylindrical coordinates, while another type of Gaussian beams, the Hermite-Gaussian (HG) beams, are the solutions of the same equation in cartesian coordinates [64]. HG do not carry an OAM, but the first vortex beams generated by Allen et al. in 1992 [64] were obtained by using a cylindrical lens to transform high-order HG modes into LG modes. In a TEM, it is possible to obtain the same effect of a cylindrical lens in optics by increasing the lens's astigmatism. This property has been exploited by Schattschneider *et al.* [134] as a mean to measure the OAM of an EVB and as a method to measure the azimuthal (and radial) state for exact LG states.

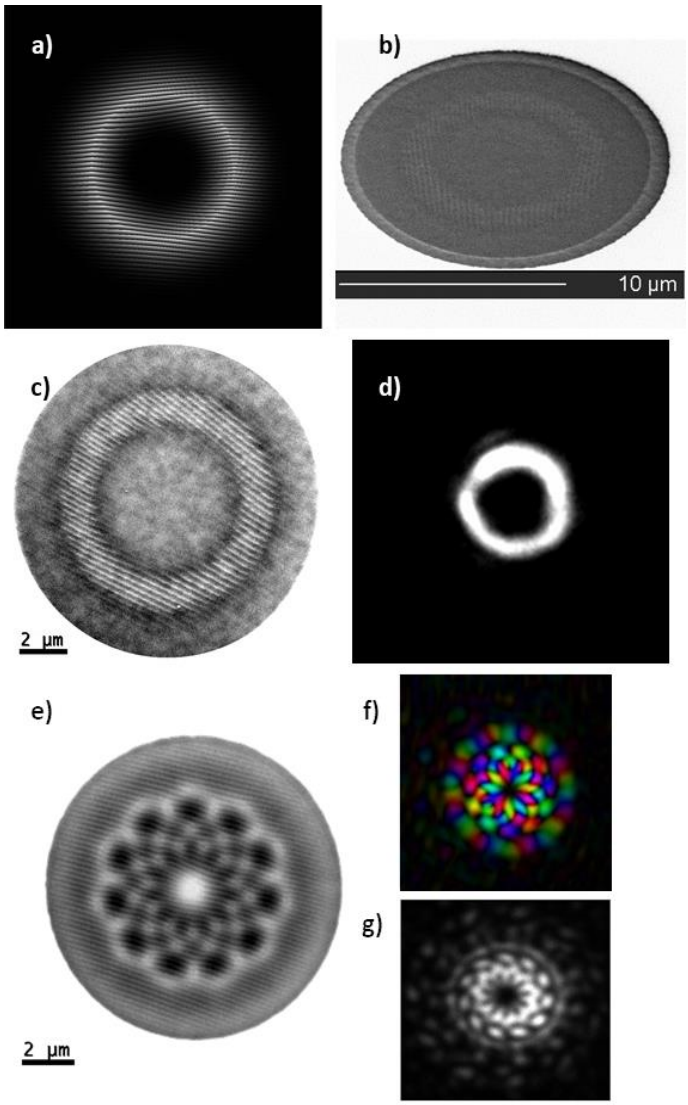

Figure 44: Here reported the steps for the fabrication and convalidation of two LG beams: a) phase and amplitude utilized for the fabrication of an , b) tilted SEM image of the resulting mixed S-CGH (patterning parameters: ion current: 300pA, repetitions: 192, magnification: 3900X), c) EFTEM map and d) diffraction image of the S-CGH where it is possible to see the "donut-like" shape of the generated EVB; e) EFTEM map of a superimposed LG beams where , f) simulated amplitude and phase and g) diffraction intensity of the 1st diffraction order.

4.3.2 Design and realization of a holographic OAM sorter

An interesting application of electron holograms is developing a device that allows the measurement of the OAM spectrum of an electron beam; this is often referred to as an OAM sorter, details can be found in the paper by Grillo et al. [23]. The device is mainly composed of two S-CGHs: a "unwrapper" S-CGH, which unwraps the impinging OAM carrying electron beam, and a "corrector" S-CGH, which corrects the phase distortion caused by the first S-CGH. The electron beam impinging on the sorter has to contain the phase information of interest, after having interacted with a sample; the most straightforward example of an OAM-generating sample is the type of in-line S-CGH presented in section 4.3.1.1. A schematic representation of the setup and the transformations involved is presented in Figure 45. It schematically shows the process of OAM generation, unwrapping, correction and detection. An electron beam impinges on the generator S-CGH and is then endowed with a spiraling phase shift, with OAM = 1 in this case, as represented by the 2π phase

shift along one complete azimuthal path. The generator is an in-line S-CGH, the use of in-line hologram simplifies the process of alignment of the beam on the sorter and excludes the effect of tilt (and off-axis aberration).

The beam carrying OAM is then directed onto the unwrapper S-CGH, which operates as a conformal transformation from log polar to cartesian coordinates. Basically, the phase information is unwrapped from an azimuthally varying arrangement to a linear arrangement, with the phase information aligned along one cartesian coordinate. The first S-CGH, or Sorter 1 element, is a diffractive hologram. Therefore, the resulting pattern will be found in the reciprocal plane. This unwrapping operation introduces a strong phase gradient to induce the transformation. However, after the transformation has occurred, the additional transformation phase must be removed. Therefore, an additional S-CGH is needed: the aforementioned corrector. This S-

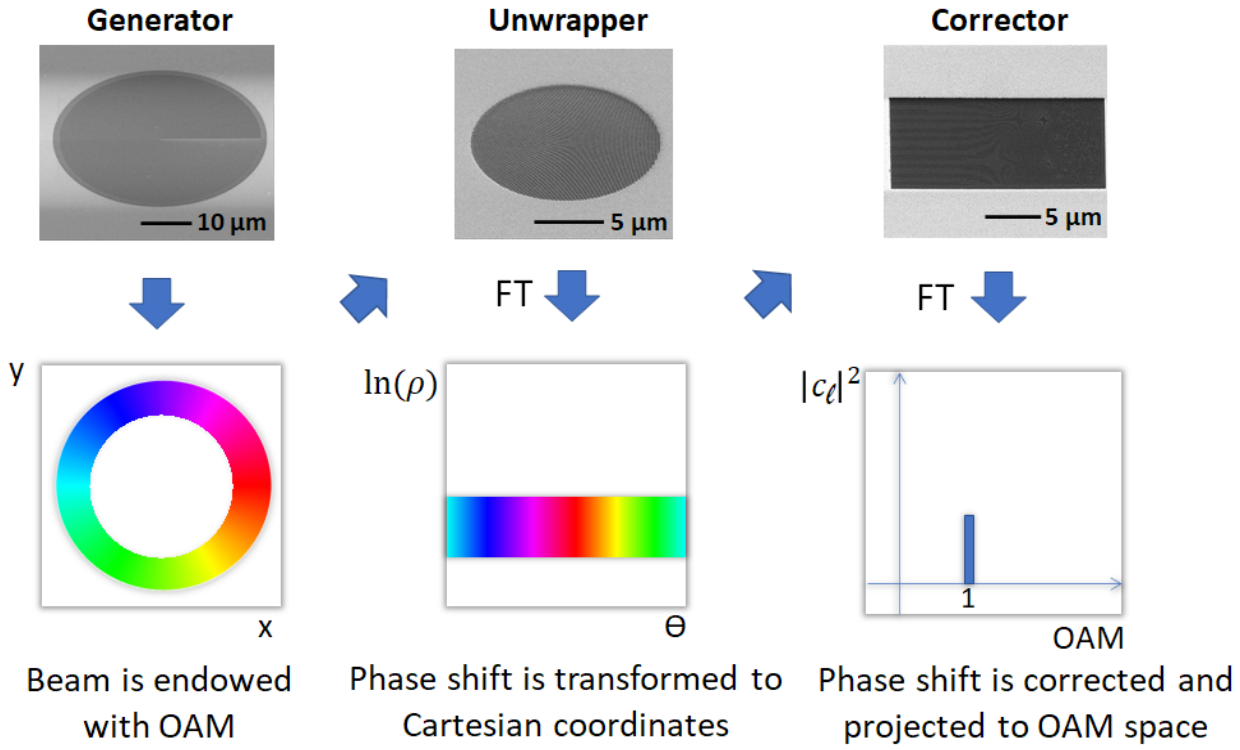

Figure 45: Schematic view of the holographic masks and the transformations carried out by the sorter device.

CGH is again an off-axis version, and the final result, the OAM spectrum, will be found in the reciprocal space. After this correction operation, the OAM value will be readily found as an intensity spot: the farther its position from the center of the first diffraction order in the reciprocal space, the higher the OAM value. A calibration procedure with different OAM-generating S-CGH of known OAM value needs to be used to calibrate the device for a particular electro-optical configuration. Changes in the electro-optical configuration can induce rotations and magnification, requiring a new calibration procedure. Once the device is calibrated, a real sample, i.e., an object endowing the beam with an unknown amount of OAM, can be used instead of the generator S-CGH and the full OAM spectrum of the beam that has interacted with it can be collected and studied. It is important to note that it is possible to measure the full OAM spectrum of a beam in one acquisition. One interesting application is the measurement of the magnetic moment of a dipole, as

presented in [23]. Further possible applications are foreseen in EMCD [72,75] and plasmon characterisation [197].

Regarding the S-CGH fabrication, here the phase functions encoded in the unwrapper S-CGH (Λ_1) and in the corrector S-CGH (Λ_2) are reported. The phase corresponding to the first element of the sorter is:

$$\Lambda_1 = \varphi_0 \, sign\left(sin\left(2\pi a \left| y \arctan\left(\frac{y}{x}\right) + x \, ln\left(\frac{\sqrt{x^2+y^2}}{b} \right) + x \right| \right) \right) \tag{198}$$

where *a* and *b* are two parameters scaled to optimise the experimental efficiency while sign denotes the sign function. The phase associated with the second hologram is:

$$\Lambda_2 = \varphi_0 \, sign\left(sin\left(2\pi ab \, exp\left(-2\pi \frac{u}{a}\right) cos\left(2\pi \frac{v}{a}\right)\right) + 2\pi cv \right) \tag{199}$$

where $u = -a \, ln\left(\sqrt{x^2+y^2}/b\right)$, $v = a \, arctan\left(\frac{y}{x}\right)$, and c consists of an additional scaling parameter. For the unwrapper presented in Figure 45, the parameters $a = 2$, $b = 0.01$, and $c = 0.6$ were used. These parameters can be tuned for matching the relative S-CGH and holographic beam sizes in a particular electro-optical configuration. The peak-to-trough depth should be tuned in order to maximize the diffraction efficiencies. This device is adaptable to any TEM, however, the electro-optical configuration is quite challenging and extra features are recommended, such as free lens control module, additional sets of lenses, and apertures.

4.3.3 Improvement to a holographic OAM sorter: design and realisation of a Fan-Out holographic OAM sorter

The OAM sorter realisation that has been reported in the previous section (4.3.2) which depicts the design details of the system used in Ref. [23], demonstrated that it is possible to record the OAM spectrum with a reasonable OAM sorting resolution, $2.5\hbar$. Unfortunately, it has a series of inherent limitations that are partially related to the design itself but are mainly due to the holographic approach itself. The main limitations are the low transmitted efficiency, the subsequential low signal to noise ratio and the theoretical limit of the approach itself (as introduced in section 3.3.2). By optimizing the fabrication process and changing the design it should be possible to improve the sorting resolution and reduce some these limitations. In fact, by taking inspiration to Mirhosseini *el al.* work [137] we designed and realized a "Fan-Out" holographic electron OAM sorter. Mirhosseini *el al.* [137] were able to produce more copies of the same unwrapped beam by adding to the sorting system two further optical elements: one which effectively generates copies of the unwrapped beam (which they referred as Fan-Out), while the second was another phase corrector to compensate unwanted artifacts introduced by the first elements. Due to limitations in the number of accessible aperture planes in the TEM (that as we have previously seen are only 4 in conventional TEMs) to achieve the fan-out effect we joined together the fan-out element and the unwrapper in a single new

eleemnt. Therefore, we modified the phase profile of the unwrapper element of the sorter, and at the same time we also modified the phase profile of the corrector element to correct for errors in the phase of the beam coming from both the unwrapping and beam copying effects.

The phase profile function of the "Fan-Out" unwrapper is given by the formula:

$$S_1 = mod\left(\frac{2\pi a}{\lambda f}\left[y\,atan\left(\frac{y}{x}\right) - xln\left(\frac{\sqrt{x^2+y^2}}{b}\right) + x\right], 2\pi\right) + F\,sin\,(g\,y + p) \tag{200}$$

From this expression it is possible to notice that the phase profile is composed of two terms: the first corresponds to the profile function of a phase S-CGH with blazed profile along the y direction, where the argument of the profile function is the phase of the ideal OAM sorter (Eq.). As previously saw, the blazed profile has the advantage to direct most of the transmitted intensity on one of the first diffraction orders, which should lessen the limitation related to the transmitted intensity. The second term is an added sinusoidal grating producing the copies of the beam in the y direction (here horizonthal) and therefore orthogonal to the main sorter carrier frequency with F a constant chosen to distribute most of the beam intensity in the $+1, -1$ and 0 diffraction orders, g is the carrier frequency and $g = \frac{2\pi}{\Lambda}$, where Λ is the grating spatial period, while p is a phase term that can be used for reciprocal alignment with the sorter itself. Since

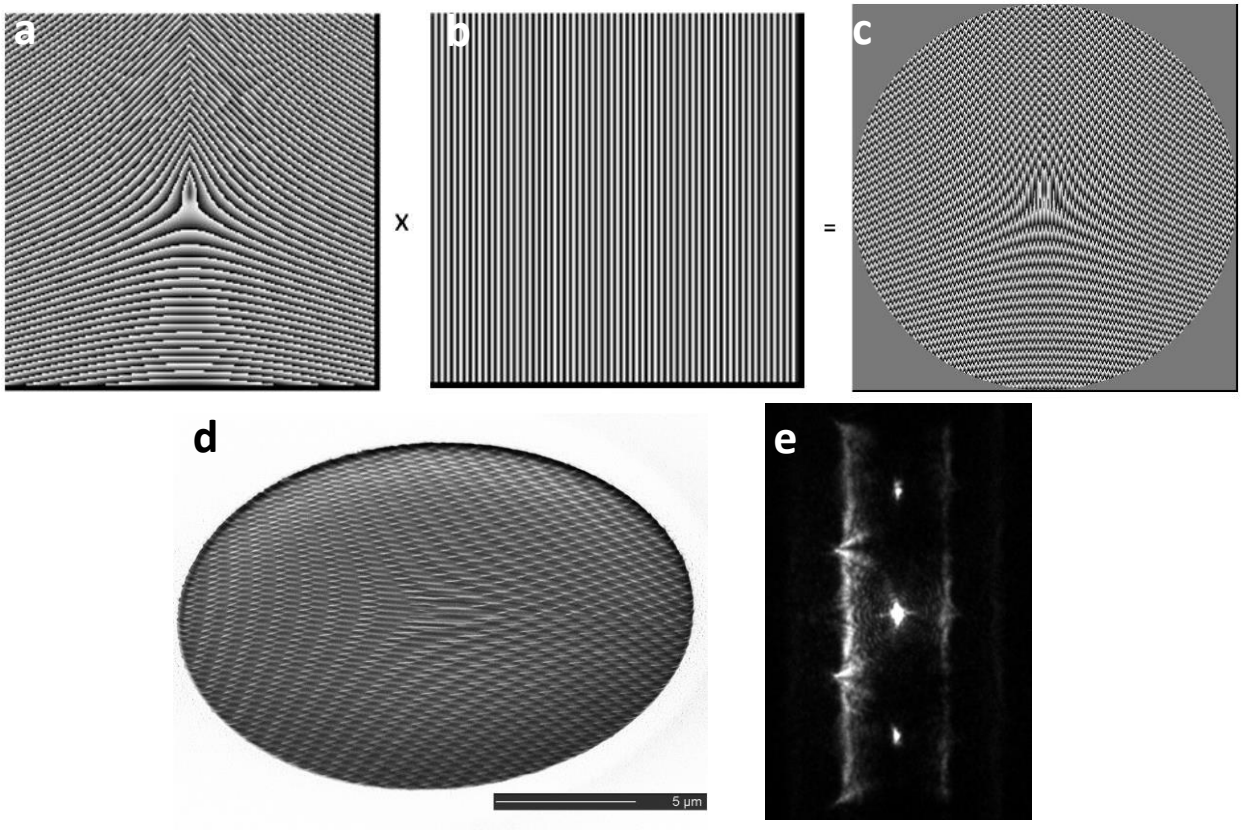

Figure 46: The steps for the creation of the phase profile of the Fan-Out unwrapper are reported: in a) it is possible to observe the phase profile of the OAM sorter with a blazed profile (white is 2π black is 0), we then convolve it with a sinusoidal profile (b) to obtain the final profile (c) which is described by Eq.200. In d) it is reported the fabricated S-CGH corresponding to c). In e) is displayed the diffraction pattern obtained after a plane wave impinges on the S-CGH in d).

the intensity of the m-th order of a sinusoidal grating is proportional to the Bessel function $J_m(F)$, the condition $J_0(F) = J_1(F)$ defines the optimal value of F at $F \approx 1.4347$. It should be also noticed that the sinusoidal phase grating can only produce at best three copies of similar intensity of the beam. In Figure 46(a-c) are reported the graphical steps on how the Fan-Out unwrapper has been designed. In Figure 46d if shown the tilted image of the fabricated phase S-CGH, and in Figure 46e is shown its diffraction pattern where it is possible to appreciate the effect of the combination of the blazed sorter profile and sinusoidal diffraction grating. Unfortunately, even after optimisation the phase of sorter 1 was not perfectly reproduced on the S-CGH so that the diffraction pattern still presented spurious diffraction orders. Each diffraction order is replicated in the vertical direction by the fan-out grating. Still, we will neglect their contributions. The shape of the sorter 2 must now keep into account the three copies of the unwrapped beam and should extend over three periods. The exact form of the sorter 2 phase is

$$S_2 = sin\left(2\pi \frac{ab}{\lambda f} exp\left(-\frac{u}{a}\right) cos\left(\frac{v}{a}\right) + cu + \phi(u)\right) \tag{201}$$

where

$$\phi(u) = \begin{cases} 0 \; for \; |u| < \dfrac{a}{4\pi} \\ \dfrac{\pi}{2} \;\; otherwise \end{cases} \tag{202}$$

In Eq.201 the in-plane variables u and v are defined as below Eq.199. Comparing the two it is possible to notice that the two expression are similar, except that in Eq.201 is present the extra term $\phi(u)$. This term is necessary because the two copies of the unwrapped beam corresponding to the two first diffraction orders are $\frac{\pi}{2}$ out of phase to the zero-order. In Figure 47a is shown the bi-dimensional phase profile corresponding to Eq.201 with $a = 10\mu$m. Figure 47a has been used to fabricate the S-CGH shown in Figure 47b, and in Figure 47c can be found the diffraction pattern obtained when a plane wave impinges on Figure 47b.

The effective implementation of the holographic fan-out has been tested in a microscope, the results of which can be appreciated in Figure 48. Unfortunately, the results are not that satisfactory. After we carried

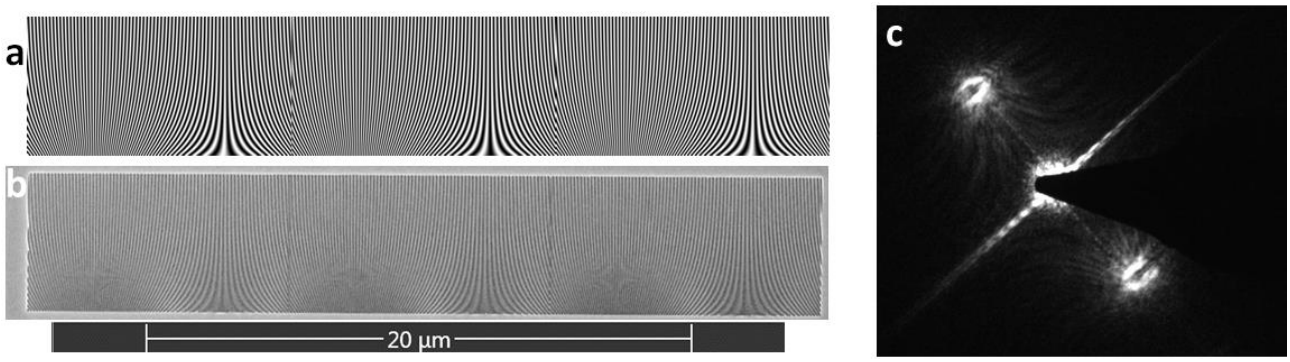

Figure 47: Here displayed in a) the computer-generated phase profile of the second element of the Fan-Out OAM sorter that is used to fabricated the S-CGH in b). In c) is reported the experimental diffraction pattern obtained by illuminating b) with a plane wave.

the experiments with the electrostatic OAM sorter (chapter 5) we now understand that the main factor broadening the OAM resolution are the imperfect size and the rotation mismatch between sorter2 and the diffraction of sorter1. These factors are very difficult to control in the holographic version. So, all the improvements due to the fan-out approach are mainly masked by the overall imperfection of the system. Furthermore, the various imperfections of the sorter holograms themselves, including the superimposed amplitude effect, make the OAM spectrum more complicated to interpret with a lot of background intensity about each OAM peak. Still, the peak of the experimental Fan-Out version is slightly smaller that that of an ideal standard holographic sorter.

This has been an activity to which I dedicated a part of PhD and I fabricated and characterized both phase S-CGHs.

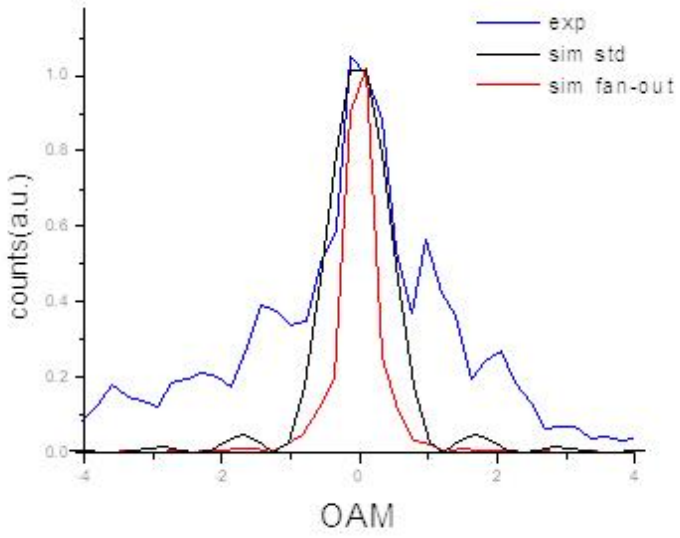

Figure 48: Comparison between the simulated and experimentsl OAM spectrum for a beam with $\ell = 0$ (normal beam with gaussian wavefront): black ideal sorter, red ideal fan-out sorter and blue experimental fan-out sorter.

4.3.4 Bessel Beam

The third example of possible applications of S-CGHs is the generation of a kind of non-diffractive beam, i.e., Bessel Beam. I already introduced them in Section 3.2 as one of the most commonly used solutions of the Schrödinger equation in cylindrical coordinates that represents a Vortex beam.

In the light optics field, the generation of Bessel beams, or more precisely quasi-Bessel beams, has been achieved in a number of ways. The crudest and simplest one is to use an annular slit or ring aperture [85]. This method works since the Fourier transform of a Bessel beam is a ring. A more efficient generation method is to use axicon lenses [198–201], which remove the on-axis intensity oscillation, resulting in a smooth intensity variation along the beam propagation direction. Other efficient methods comprehend S-CGHs [202], SLMs [203,204] and cavities [205,206].

In recent years, by taking inspiration from the light field, different methods for the generation of electron quasi-Bessel beams have arose: in 2014 Grillo *et al.* reported the generation of non-diffractive quasi-Bessel

beams by means of a S-CGH. The produced beams were able to propagate for 0.6 m without noticeable spreading of the central maximum and could reconstruct [90,207]. Taking inspiration from the first experiments by Durnin, Saitoh and colleagues have used annular slits to generate quasi-Bessel beams [208]. Interestingly, in 2017 Zheng and colleagues used magnetic vortices with circular magnetic moment distributions which are naturally present in a soft-magnetic thin film as axicon lenses [209].

The phase S-CGH that was used by Grillo *et al.* [90] in imprints a phase modulation to the transmitted beam given by the formula:

$$\varphi(\rho,\theta) = \varphi_0 sgn\left[cos\left(k_\rho \rho + \ell\theta + g\rho\, cos\theta\right)\right] \tag{203}$$

The generated hologram is an off-axis Fresnel hologram with carrier frequency $g = \frac{2\pi}{\Lambda}$, where Λ is the grating spatial period. The relation between the linear phase and the Bessel beam has been demonstrated in [90,207].

From this formula, it is possible to notice that the chosen profile shape was the squared one, and that its argument is $\alpha(\rho,\theta) = k_\rho \rho + \ell\theta + g\rho cos\theta$. It is possible to notice how the $\alpha(\rho,\theta)$ that was used, is actually quite similar to the one found in Eq.196, i.e., the one of the pitchfork design. As a matter of fact, the quasi-Bessel beam generated was an OAM carrying one. Figure 49 represents a fabricated phase S-CGH for quasi-Bessel beams generation, the CGH used to produce it and an experimental diffraction image where it is

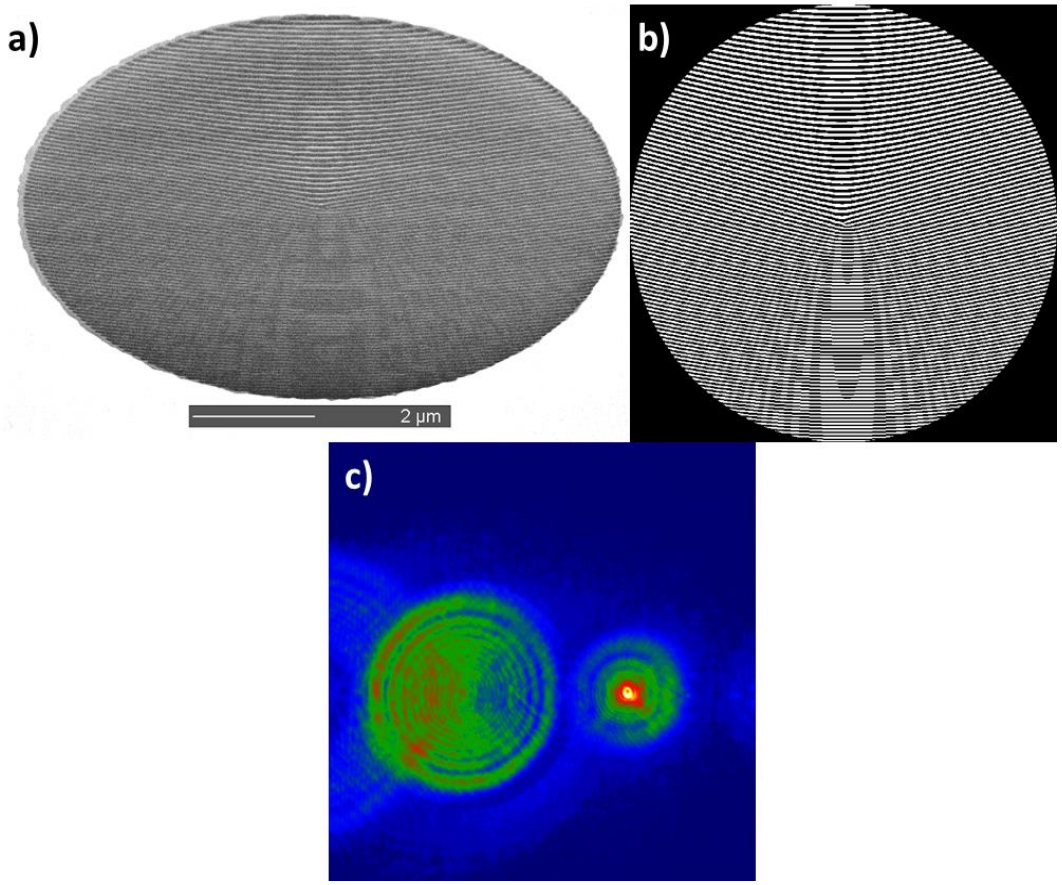

Figure 49: Here reported a) SEM image of a fabricated phase S-CGH used for the generation of quasi-Bessel beams with l=2, the stage is tilted to highlight the 3-D features, b) the 2-D image of the profile function fed to the FIB machine used to obtain a) and c) shows the experimental diffraction image where on the right it is possible to observe the first diffraction order with the quasi-Bessel beam. To recreate b) the same parameters used in [90] *were used (and , aperture diameter).*

possible to observe the generated quasi-Bessel beam. At the very center of Figure 49b, it is possible to notice the typical dislocation of the pitchfork design.

In a later paper [207], by switching to the cosinusoidal profile and by optimizing the fabrication process through a systematic procedure, the research group led by Grillo was able to increase the transmission efficiency by up to $37 \pm 3\%$. They pointed out the possible application fields of quasi-Bessel beams generated through S-CGHs depending on the aperture radius: small radii are best suited for STEM, while larger radii for interferometry techniques.

Possible applications of this type of structured beams in electron microscopy range from classical techniques like tomography [210] or strain mapping [211], up to conventional STEM, low dose STEM, and HR-STEM imaging as proposed in [73,90,207,212].

4.3.5 Cs corrector

On an ending note, it is worthwhile devoting a few words to another example. This example is a bit peculiar with respect to beam shaping in the sense that we cannot apply the scheme of section 4.1.4. In fact, it is well known that the problem that afflicted electron microscopy for a long time has been the inevitable presence of spherical (Cs) aberration for any lens maintaining the cylindrical symmetry. While the solution has now been found using a complicated set of multipoles [34,35], it would be nice to see if one can produce a S-CGH that is able to compensate the spherical aberration by introducing, in the condenser aperture, an equal phase but with opposite sign to that of the Cs aberration. The peculiarity here is that the aimed phase α is not known in the diffraction plane but directly in the S-CGH plane. Therefore, the aim is to correct the STEM probe for the aberration working with an aperture in the condenser plane. As it often happens, different groups [213–215] have produced this hologram more or less at the same time but with slightly different recipes.

The general formula for the phase is therefore:

$$\alpha(\rho) = \frac{2\pi}{\lambda}\left(-\Delta f \rho^2 + \frac{1}{4} C_s \rho^4\right) + g\rho\, cos(\theta) \tag{204}$$

It is worth mentioning here that the inline approach is recovered when g=0. It is easy to see that for the correction to be extended over a sufficiently wide field (to extend beyond the standard STEM probe) it is necessary to have a phase ranging over $4 - 6\pi$. One could use a continuous slope where the thickness $t = \alpha$ or a modified discontinuous slope $t = Mod(\alpha, 2\pi)$ introducing few 2π wrappings. The first approach would produce a very thick membrane and significant absorption effects, while the second requires a precise tuning of the discontinuity.

The completely inline version with a large value of Δf can be used to create many beams that go in focus in different values of the z-coordinate [216]. In this case, any kind of groove like sinusoidal could be used but this approach has never been used so far (see Figure 50(b)). Here instead it is reported the case of the off-

axis approach that ensures potentially an excellent control of the phase employing a relatively thin membrane at the cost of spurious diffraction order. Grillo *et al.* [213] explained how to remove this spurious order by means of smartly using optics' components. Figure 50(a) shows the typical aberration function in the presence of defocus (Cs=0.5 mm and $\Delta f = 40\ nm$), while the equivalent off-axis phase S-CGH pattern and a realization are shown in Figure 50(c), and Figure 50(d), respectively.

In the actual realization of the hologram, it is preferable to use a sinusoidal or a blazed groove shape as this ensures a smoother variation of the phase as described in section 1.6. Also, for practical reasons, the carrier frequency must be quite large so that isolating the specific beam (with the exact positive Cs value) in the diffraction plane is easier. These two constraints almost naturally lead to very large holograms, one of the main challenges for fabrication and durability of the hologram, as previously mentioned in section 2.5.

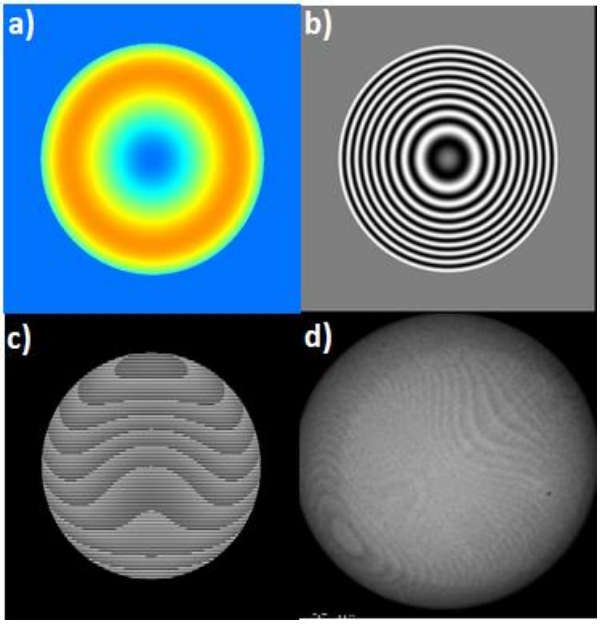

Figure 50: Design of an holographic Cs corrector. The figures show a) the aimed phase plate, b) the inline S-CGH using a sinusoidal groove and Δf=400 nm, c) the off-axis S-CGH and d) a SIN realization.

4.4 – Conclusions

In this chapter, we have introduced the concept of “imaging” holography and then delved in synthetic holography. In particular, for synthetic holography, we showed how the concept was born, starting from the idea of CGHs to simulate the interference patterns up to the discovery of the possibility of engineering the wavefunction, which paved the way for new techniques in materials science. We have provided the mathematical description for the most commonly used groove profiles in S-CGHs, differentiating for each profile between amplitude and phase ones, and explaining what the most efficient profiles are and that it is possible to design mixed-phase and amplitude S-CGHs to take care of the negative aspects of the two kinds. We then introduced two fabrication techniques that are widely spread that can be utilized to manufacture S-CGHs. For each technique, we provided a guide on how to optimize the production process and what are the pros and cons of each technique. As a completion we introduced a few examples of possible uses of S-CGHs in the field of Electron Vortex Beams. As a case study, we showed that real phase S-CGHs are quite sensitive to the imperfections inherently introduced by the fabrication. On the other hand, they are robust enough to be able to generate the desired wave functions with losses mainly in the efficiency.

Chapter 5

This chapter is dedicated to the original research work regarding the realization of the electrostatic quantum sorter for Electron Vortex Beams. However, I deem necessary to first introduce the theory behind the device that was first reported by McMorran *et al* [138] and later revised by Pozzi *et al* [143]. The starting point of the calculations reported by McMorran *et al* [138] is Eq.92. If the needle now lies parallel to the x-axis, with one tip at $x = 0$ and the other at $x = -L$, while the constant charge density can be written as $\sigma = Q/L$. The needle is oriented perpendicularly to a conducting plate that lies parallel to the $y - z$ plane at $x = h$ (which is used so that the total charge of the system in study is zero, i.e, as a charge neutralizer). Then, Eq.92 can be readapted into

$$\varphi_U(x,y) = -\frac{QC_E}{4\pi\varepsilon_0 L}\left[|y|sin^{-1}\left(\frac{-x-L}{\sqrt{(x+L)^2+y^2}}\right) - |y|sin^{-1}\left(\frac{-x}{\sqrt{x^2+y^2}}\right)\right.$$
$$+|y|sin^{-1}\left(\frac{-x+2h-L}{\sqrt{(x-2h-L)^2+y^2}}\right) - |y|sin^{-1}\left(\frac{-x+2h}{\sqrt{(x-2h)^2+y^2}}\right)$$
$$-xln\left(\frac{\sqrt{(x+L)^2+y^2}}{\sqrt{x^2+y^2}}\right) - xln\left(\frac{\sqrt{(x-2h-L)^2+y^2}}{\sqrt{(x-2h)^2+y^2}}\right)$$
$$\left.-L\,ln\left(\frac{\sqrt{(x+L)^2+y^2}}{\sqrt{(x-2h-L)^2+y^2}}\right) + 2h\,ln\left(\frac{\sqrt{(x-2h-L)^2+y^2}}{\sqrt{(x-2h)^2+y^2}}\right)\right. \quad (205)$$

This expression can be simplified by making a series of assumptions. First, we can consider the case in which the electron beam is confined only to the region immediately adjacent to the tip of the needle nearest to the plate. Second, we can take the distance h to be much larger that the region of interest, so that $h \gg \sqrt{x^2+y^2}$. With these two assumptions we have that the third and fourth terms cancel each other out, the sixth goes to zero and the last two terms go to a constant phase shift that depends only on L and h. It is then possible to write the unwrapper phase as

$$\varphi_U(x,y) = -\frac{QC_E}{4\pi\varepsilon_0 L}\left[|y|sin^{-1}\left(\frac{-x-L}{\sqrt{(x+L)^2+y^2}}\right) - |y|sin^{-1}\left(\frac{-x}{\sqrt{x^2+y^2}}\right)\right. \quad (206)$$
$$\left.-xln\left(\frac{\sqrt{(x+L)^2+y^2}}{\sqrt{x^2+y^2}}\right)\right] + \varphi_0$$

where φ_0 is a constant phase term that does not affect the sorter mechanism. Now by considering that $sin^{-1} = \frac{\pi}{2} - cos^{-1}$ and that $cos^{-1}(-\alpha) = \pi - cos^{-1}(\alpha)$, it is possible to rewrite Eq.206 as

$$\varphi_U(x,y) = -\frac{QC_E}{4\pi\varepsilon_0 L}\left[|y|cos^{-1}\left(\frac{x+L}{\sqrt{(x+L)^2+y^2}}\right) - |y|cos^{-1}\left(\frac{x}{\sqrt{x^2+y^2}}\right) - xln\left(\frac{\sqrt{(x+L)^2+y^2}}{\sqrt{x^2+y^2}}\right)\right] + \varphi_0 \quad (207)$$

By considering the length of the needle to be much larger than the region of interest, i.e., $L \gg \sqrt{x^2+y^2}$, it is possible to notice that the first term in Eq.207 tends to 0, since $cos^{-1}(1) = 0$. We can then write the unwrapper phase close to the needle tip as

$$\varphi_U(x,y) = -\frac{QC_E}{4\pi\varepsilon_0 L}\left[-|y|cos^{-1}\left(\frac{x}{\sqrt{x^2+y^2}}\right) + xln\left(\frac{\sqrt{x^2+y^2}}{L}\right)\right] + \varphi_0 \quad (208)$$

And finally by using the fact $cos^{-1}\left(\frac{x}{\sqrt{x^2+y^2}}\right) = tan^{-1}\left(\frac{|y|}{x}\right)$ we obtain

$$\varphi_U(x,y) = \frac{QC_E}{4\pi\varepsilon_0 L}\left[|y|tan^{-1}\left(\frac{|y|}{x}\right) - xln\left(\frac{\sqrt{x^2+y^2}}{L}\right)\right] + \varphi_0 \quad (209)$$

Which represents the desired phase of the unwrapper (Eq.124) for $\frac{2\pi a}{\lambda f} = \frac{QC_E}{4\pi\varepsilon_0 L}$ and minus the linear phase term. Since the term $\frac{Q}{4\pi\varepsilon_0 L}$ from a point of view of units is an electrostatic potential, it is possible to substitute to the charged line a polarized needle with an applied equivalent potential $V_{needle} = \frac{Q}{4\pi\varepsilon_0 L}$. In the paper by McMorran *et al*. [138] they compared the phase between the ideal sorter and the electrostatic one by calculating the deviation of the phase profiles. Here the deviation is defined as the difference between the simulated phase (of the electrostatic one) and the ideal phase, divided by the difference between the maximum and minimum ideal phase in a certain region. The phase profile of the electrostatic sorter was calculated using a grounded plate to respect the Dirichlet boundary conditions. They noticed an astigmatic $y^2 - x^2$ phase in the deviation and proposed that it was possible to correct it with the commonly present quadrupolar stigmators. In a more recent paper by Pozzi *et al*. [143], owing to previous calculation on caustic patterns generated in the out-of-focus images of a charged needle [122], they reported a more complete formula of the phase of a charged needle (which can be also substituted by a needle kept at a specific electrostatic potential), which slightly deviated from the formula of the ideal sorter. The analytical formula of the phase at $z = 0$ of a charged needle that they obtained is [122,143]:

$$\varphi_U(x,y) = \frac{QC_E}{4\pi\varepsilon_0 L}\left[-(L+y)ln\left(\frac{(x+L)^2+y^2}{\delta^2}\right) + y\ ln\left(\frac{x^2+y^2}{\delta^2}\right) + 2L + \right. \quad (210)$$
$$\left. +2y\ arctan\left(\frac{x}{y}\right) - 2y\ arctan\left(\frac{L+x}{y}\right) + L\ ln\left(\frac{(x+x_D)^2+(y+y_D)^2}{\delta^2}\right)\right]$$

where (x_D, y_D) is the position of the charge neutralizer conducting plate and δ^2 is a scaling factor with the dimension of a length. Even in this case they compare the deviation of the two phase profiles (ideal vs

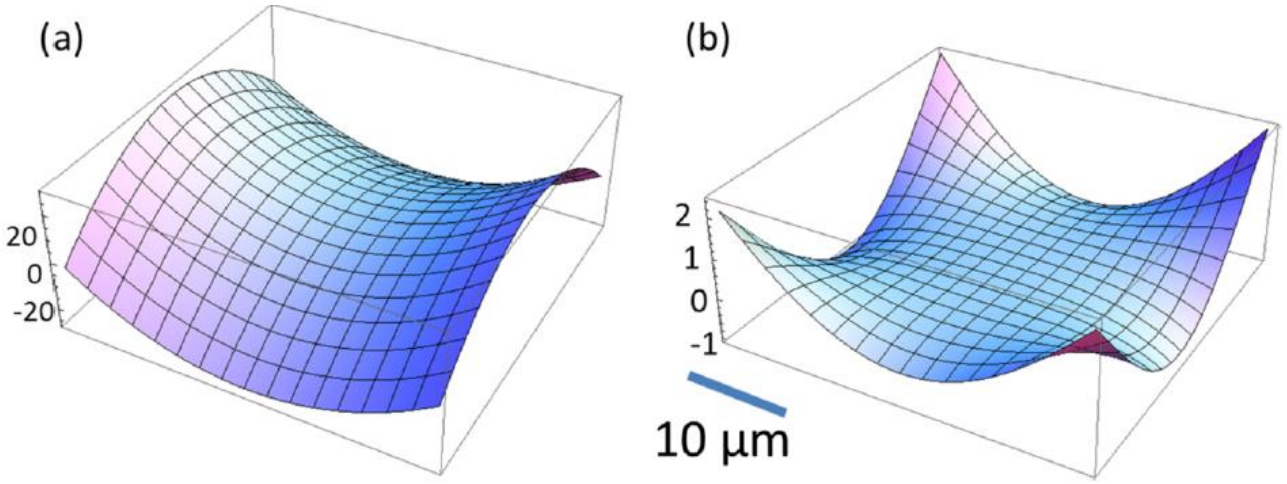

Figure 51: three dimensional shape of the phase deviation between the ideal sorter and the electrostatic sorter phase profiles: a) with the previously proposed "one-needle" configuration, b) with the new "three-needles" configuration to compensate for the astigmatism. The hue in the images has no meaning and it is artificially introduced by the software used to plot them. Taken from [143].

electrostatic) and recovered that the electrostatic sorter phase had an addition astigmatic phase. However, they also proposed a design improvement to correct, at least partially, for the inherent astigmatism. In their new design two other charged line are introduced. They lie on the same plane as the original charged line, however they are perpendicular to it and parallel to the surface of the charge neutralizing plate. The saddle-like three-dimensional shape of the "original" phase deviation is shown in Figure 51a, while the flattened three-dimensional shape of the phase deviation with the two additional charged lines used to compensate for the astigmatism is shown in Figure 51b. It evident how for a region of few tens of microns around the tip the phase profile of the electrostatic sorter is almost identical to the ideal one with this new design. Furthermore, in the case the charged lines are substituted by needles kept at a specific electrostatic potential it has been found that the two astigmatism-compensating needles need to be kept at a potential that is approximately half of the "unwrapper"-needle and of opposite sign.

In the former paper by McMorran *et al*. [138] they also recognized that is was possible to obtain the same phase profile of the OAM-sorter phase corrector element, Eq.125, using a set of alternating oppositely polarized electrodes. This is possible since the phase distribution of the "corrector" is a solution of Laplace's equation in 2D ($\nabla^2\varphi_C(u,v)=0$). The resulting 2D electrostatic potential solution of Laplace's equation can be approximated by a 3D potential if said potential slowly varies along z. The alternating electrodes configuration can be used as a phase corrector as long as the longitudinal height D of the electrodes is much longer than the period d, and the thin grating condition is satisfied, i.e., $\lambda D \ll d^2$, so that the variation of the potential in the longitudinal direction is negligible over the depth. In particular, the phase of the corrector can be written as:

$$\varphi_C(u,v) = C_E D V_C exp\left(-\frac{2\pi u}{d}\right) cos\left(-\frac{2\pi v}{d}\right) \tag{211}$$

where (u, v) are the same transverse coordinates used by Berkhout and colleagues [136] to describe the transformed field in the corrector plane, V_C is the absolute value of the applied potential to each electrode. Even in this case, a grounded surface must be positioned at a certain distance (much greater than d) to satisfy the Dirichlet boundary conditions. Furthermore, they found that as long as the depth D was much larger than the period d, i.e., the potential is constant in z over a much longer length scale than it varies in u and v, the fringing fields were insignificant.

"Experimental demonstration of an electrostatic orbital angular momentum sorter for electron beams"

Amir H. Tavabi[1], Paolo Rosi[2], Enzo Rotunno[3], Alberto Roncaglia[4], Luca Belsito[4], Stefano Frabboni[2,3], Giulio Pozzi[1,5], Gian Carlo Gazzadi[3], Peng-Han Lu[1,6], Robert Nijland[7], Moumita Ghosh[7], Peter Tiemeijer[7], Ebrahim Karimi[8], Rafal E. Dunin-Borkowski [1], Vincenzo Grillo[3]

1. Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany
2. Dipartimento FIM, Universitá di Modena e Reggio Emilia, 41125 Modena, Italy
3. Centro S3, Istituto di nanoscienze-CNR, 41125 Modena, Italy
4. Istituto per la Microelettronica e i Microsistemi-CNR, 40129 Bologna, Italy
5. Department of Physics and Astronomy, University of Bologna, 40127 Bologna, Italy
6. RWTH Aachen University, 52074 Aachen, Germany
7. Thermo Fisher Scientific, PO Box 80066, 5600 KA Eindhoven, The Netherlands
8. Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Published (doi: https://doi.org/10.1103/PhysRevLett.126.094802)

Abstract

The component of orbital angular momentum (OAM) in the propagation direction is one of the fundamental quantities of an electron wave function that describes its rotational symmetry and spatial chirality. Here, we demonstrate experimentally an electrostatic sorter that can be used to analyze the OAM states of electron beams in a transmission electron microscope. The device achieves post-selection or sorting of OAM states after electron-material interactions, thereby allowing the study of new material properties such as the magnetic states of atoms. The required electron-optical configuration is achieved by using microelectromechanical systems technology and focused ion beam milling to control the electron phase electrostatically with a lateral resolution of 50 nm. An OAM resolution of $1.5\hbar$ is realized in tests on controlled electron vortex beams, with the perspective of reaching an optimal OAM resolution of $1\hbar$ in the near future.

Forewords

In this work we report the realization of a working Electrostatic Quantum Sorter for Electron Vortex Beams. I participated in writing the manuscript, 3D shaped the devices via FIB milling both in Italy and in Jülich, characterized them by acquiring the phase images in different configurations (by varying the applied potentials on the devices' electrodes) and took part in the experiments that we carried out to benchmark our Quantum Sorter in which we were able to reach an experimental OAM sorting resolution of $1.5\hbar$. Both the characterization and the benchmarking were carried out at the Forschungszentrum Jülich using the Titan HOLO TEM (TFS).

Introduction

Although electron microscopy was pioneered in the early 1930s [26], the discipline continues to be developed both technically and methodologically [217,218]. A modern transmission electron microscope can be used to characterize materials with sub-0.1-nm spatial resolution [219–221], to provide three-dimensional microstructural and compositional information [210], to record electron energy-loss spectra with a sensitivity of a few meV [222,223] and to achieve new horizons in protein characterization [224,225]. One of the most significant advances has been the theoretical and experimental development of spherical aberration correction [33,35,226], which requires the precise manufacture, alignment and control of the magnetic fields of electron-optical elements using computer-aided procedures and electronics. Recently, other concepts have been taken from light optics and used to explore new directions in electron optical instrumentation. For example, synthetic holograms have been used to realize electron vortex beams [20–22,66,67,81,92,100,101,152], non-diffractive beams [227,228] and new measurement schemes [23,69,229]. The latter developments have been used to introduce the component of orbital angular momentum (OAM) in the mean propagation direction (henceforth defined *z*) as a "new" degree of freedom in transmission electron microscopy. However, the use of synthetic holograms is detrimental to electron beam coherence and intensity, motivating the use of electrostatic and magnetostatic fields to achieve electron beam control [138,143,230].

Here, we realize an *electrostatic* "sorter" for electrons that can be used to measure the OAM components of an electron beam. The z component of the OAM operator is proportional to the gradient of the electron wavefunction with respect to the azimuthal coordinate θ, according to the expression $L_z = -i\,\hbar\partial_\theta$. Its eigenstates are electron vortex beams, which are characterized by azimuthal phase terms and can be described in the form $e^{-i\ell\theta}$. The measurement of an OAM state is therefore equivalent to finding the value of ℓ. We previously reported the realization of such a device based on two holographic phase elements fabricated from SiN [23]. In the present study, we demonstrate a complete setup that is based on the use of controllable electrostatic fields and can be retrofitted to an existing electron microscope.

An OAM sorter is intended to study an electron beam after its interaction with a sample [72,75,197] . So far much work has been done in the post-selection of the linear momentum after elastic and, more interestingly, inelastic scattering of electron beams exploring new territories for electron microscopy [231,232]. However, little has been achieved in terms of control of the OAM final state. Still the cylindrical symmetry is the most natural basis for atom scattering. EELS with OAM post selection could be used to assess magnetic dichroism [72,75,197], discriminate multipolar transitions [233] and study two dimensional materials [234], and for more exotic applications [235]. However, efficient and accurate measurement is required for this to become a valuable tool in materials science and cryoelectron microscopy [235].

Previous methods of OAM measurement [129,130,132,133,236–239] were based on diffraction by components such as amplitude apertures. In contrast, an OAM sorter provides a decomposition of all OAM channels at the same time by making use of a conformal mapping from Cartesian to log-polar coordinates. Such a coordinate transformation permits radial and azimuthal degrees of freedom to be decoupled. Since it corresponds to a unitary change of basis, all of the components of the electron beam that have a chosen OAM value are propagated into a single region of space with virtually no loss and superior efficiency to the use of a pitchfork hologram [162], despite the simplicity of the latter concept. Although the Stern-Gerlach-like approach described in Ref. [238] also offers a potentially near-unitary transformation, this method does not decouple radial and azimuthal degrees of freedom.

Here, we describe the technical steps that are required to realize an *electrostatic* OAM sorter. We present preliminary results that demonstrate its successful operation, including the recording of OAM spectra in a transmission electron microscope.

Whereas an OAM sorter that is based on the use of synthetic holograms follows an equivalent light-optical holographic approach [136], McMorran et al. showed theoretically [138] that a set of electrodes is able to reproduce the required phase elements. This proposal was improved by Pozzi *et al*. [143], who proposed the use of additional electrodes outside the area of the beam, as well as by the realization that all conformal mappings can be reproduced by the use of near-harmonic phase elements, such as those produced by electrodes [240]. Conformal mappings in optics involve transformations of coordinates by imparting appropriate phase gradients (and therefore in-plane velocities) to wavefunctions and allowing them to evolve to corresponding diffraction planes. They can be achieved using two phase elements and make use of quantum mechanical changes of basis to measure quantities such as OAM. According to theory [138,143], a first sorter element S1 (or “unwrapper”) can take the form of a single long needle located in front of an electrostatic mirror. To a first approximation, such a needle can be modelled as a straight line of constant charge density or, equivalently, as an ellipsoid-shaped electrode that corresponds to a desired equipotential contour. An electrostatic mirror (*i.e.*, a grounded flat electrode) should be positioned in front of the needle to ensure charge neutrality. The needle introduces a phase cuspid, which unwraps the electron beam so that an azimuthal phase gradient is transformed into a linear gradient in the stationary phase approximation. This

transformation is completed in the far-field diffraction plane, where a second phase element S2 (or "sorter corrector") removes the transformational phase and the beam can continue without re-wrapping again. The second element can take the form of an in-plane sequence of parallel needle-shaped electrodes of alternating opposite polarity.

It is convenient to use complex notation to describe mappings from complex coordinates $u = x + iy$ to *u'=x'+iy'*, where *x, y* and *x', y'* are orthogonal to the main propagation direction. The first conformal mapping can be expressed $u' = sf \ln(u/L)$, where *s* is a transformation scale parameter (approximately equivalent to the average tilt imparted by S1). *f* is the focal length of the lens system between the elements and *L* is a characteristic length (corresponding to the length of the needle in the electrostatic case). Element S1 applies a phase distribution of the form

$$\varphi_1 = \frac{s}{\lambda} \Re e\left\{\mathrm{u} \ln\left(\frac{u}{L}\right) - u\right\}, \tag{212}$$

whereas element S2 applies a compensating phase distribution of the form

$$\varphi_2 = \frac{Ls}{\lambda} \Re e\left\{\exp\left(\frac{u}{sf}\right)\right\}, \tag{213}$$

where $\Re e$ is the real value of the function and λ is the electron wavelength (See the Supplementary Material). Based on this initial design, we developed devices that can be fitted into an electron microscope. Without altering the microscope significantly, *i.e.*, without adding lenses, we realized OAM sorter elements S1 and S2 using apertures that were fabricated using microelectromechanical systems (MEMS) technology, one of which was placed in the back focal plane of the objective lens and the second in a selected area diffraction (SAD) plane. The aperture holders were custom-made and equipped with housings for 4 x 11 mm MEMS chips with sockets based on those used in a Thermo Fisher NanoEx-i/v specimen holder. MEMS fabrication involved two-dimensional patterning of device electrodes using optical lithography and deep reactive ion etching of silicon-on-insulator wafers using a recessed etching isolation technique. Up to 8 electrical pins could be used inside the microscope by connecting internal to external sockets through the body of each aperture holder. The use of MEMS technology with a planar geometry in a restricted space required several changes from the theoretical geometry. Needle S1 was replaced by three needles, as recently reported calculations [143] showed that astigmatism introduced by a finite needle can be compensated by the use of lateral electrodes oriented perpendicular to it. The shapes of the needles were initially defined using MEMS fabrication and then adjusted to achieve more precise (on a 50 nm scale) ellipsoidal shapes using focused ion beam milling. A virtually infinite series of electrodes in device S2 was approximated by the use of 11 electrodes. Smart use of a planar topology with links behind the connection pads permitted the even pins to be connected to each other, while the odd pins could be driven independently.

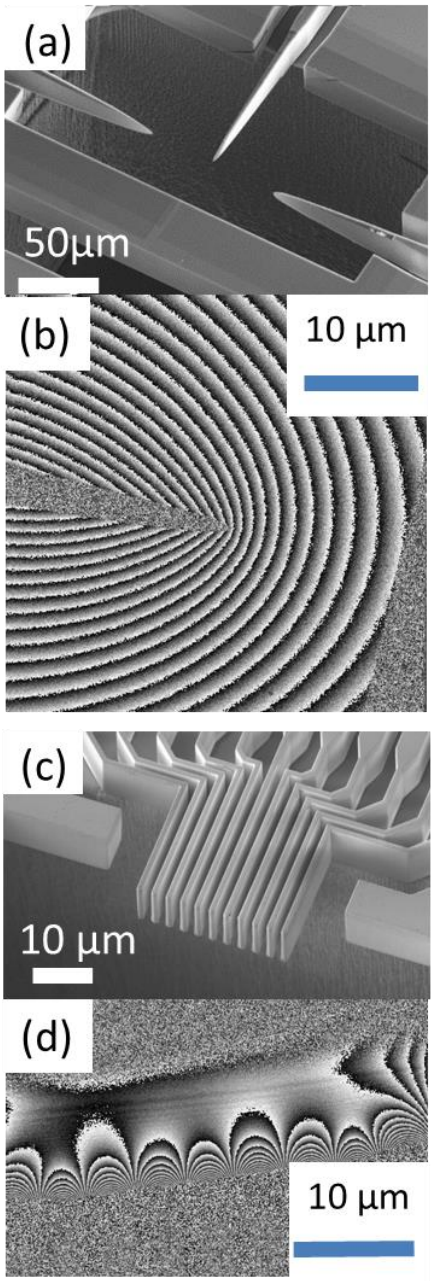

Figure 52: (a, c) SEM images of electrostatic sorting elements S1 (unwrapper) and S2 (corrector). (b, d) Electron optical phase contour maps showing the projected electrostatic potential recorded using off-axis electron holography for S1 and S2, respectively. For S1, the bias applied to the main tip was reduced to 1 V for this measurement.

Figure 52 shows off-axis electron holography measurements of electrically-biased MEMS chips inserted in the specimen plane of a standard Thermo Fisher NanoEx-i/v holder, confirming the expected phase distributions. The phase profile for S1 is almost ideal. The S2 phase contours are also almost ideal, in particular close to the central pins. Although the influence of truncation resulting from the use of only 11 needles is visible, it does not affect the region of the central 5 electrodes significantly.

Figure 53(a) shows a schematic diagram of the positions of the phase elements and their effect on the electron beam. Experiments were performed at 300 kV in a Thermo Fisher Titan G2 60-300 transmission electron microscope, which is equipped with an X-FEG electron emitter and an image spherical aberration corrector. The illumination system was set to spot size 9 and the three condenser lens system of the

microscope was used to achieve a probe convergence semi-angle of 2 mrad in the specimen plane. Although a specimen can be located in a standard position in the objective lens, for test purposes a synthetic hologram, which is referred to here as a “generation” hologram, was placed in the second condenser aperture plane, in order to generate a known electron beam (typically a superposition of vortex beams). The microscope was operated in “microprobe” mode, with the objective lens excited to a standard pre-set value. The lenses and multipoles in the image aberration corrector were switched off, in order to achieve a larger focal distance between the sorting elements. In this configuration, the second sorter element S2 was located in the diffraction plane of the first sorter element S1 (mounted in the objective (OBJ) aperture plane), which is conjugate to the specimen plane. An OAM spectrum could then be recorded on the detector with the microscope set to diffraction mode. This microscope has two SAD planes (SAD1 and SAD2). In the present study, the upper one (SAD1) was used.

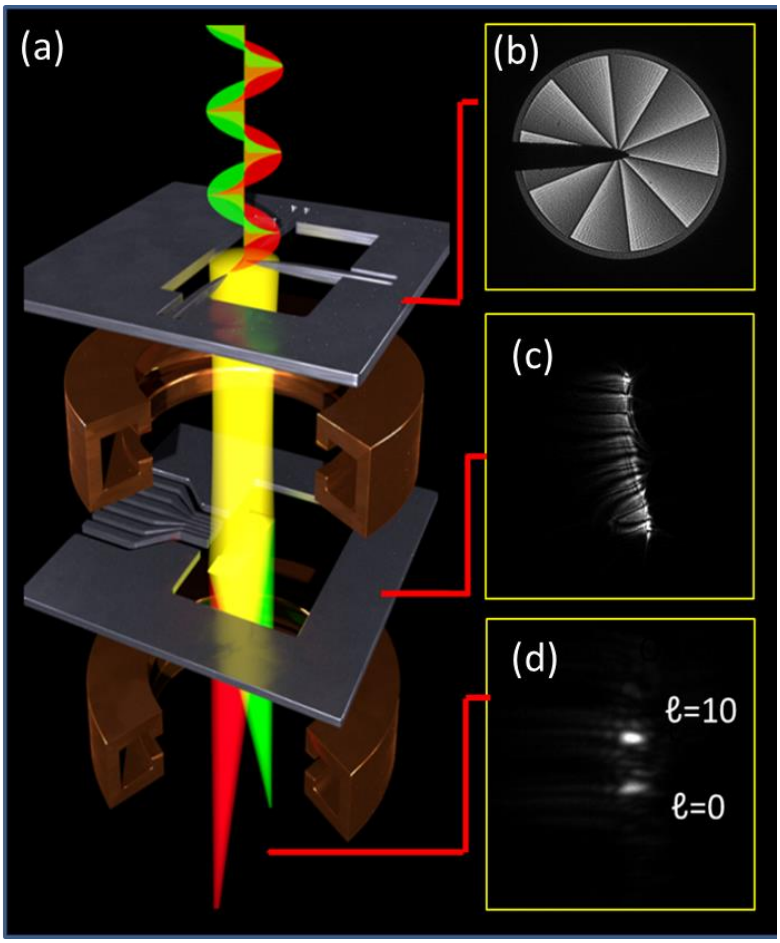

Figure 53: (a) Schematic diagram of part of an electron microscope column that contains a sorter. (b-d) Experimental images showing the evolution of the electron wavefunction for a nominal superposition of $|\ell = 10\rangle$ and $|\ell = 0\rangle$, including (b) generation, (c) conformal transformation to polar co-ordinates and (d) transformation into an OAM spectrum.

The device was tested by using “generation” holograms in the condenser aperture plane to create electron beams with known OAM states [23]. The “generation” holograms were nano-patterned SiN membranes, in which thickness modulations were used to create intended phase distributions. With the image corrector

unit turned off, an effective focal distance of $f \approx 530$ mm could be achieved. A voltage applied to the main needle of V_c = -9.1 V corresponded to a scale factor $s \approx 3$ µrad. The lateral needles of element S1 were kept at V_L = -0.5 V_c. For the S2 element, the needles were biased with alternating voltages $|V_{S2}| = 16$ V (with 8 V on the external ones). A change of applied bias could be compensated by a lateral shift of the beam. The alignment of the setup was found to be demanding, as the diffraction plane of element S1 had to be matched in position, rotation and magnification to element S2. Under standard working conditions, a rotation angle of ~23° was measured between the OBJ aperture and the SAD1 aperture. This angle was compensated by designing a rotated element S2, with the residual rotation adjusted by applying a minor excitation to the adaptor and transfer lenses of the aberration corrector. The magnification introduced by element S1 was adjusted by tuning the bias of the main S1 needle (with the voltages of the lateral needles tuned accordingly). The position of the diffraction pattern on element S2 was adjusted by using the image shift coils.

Figure 53(b-d) show the evolution of an electron beam from a "generation" hologram mounted in the condenser aperture plane (reproduced almost exactly at the entrance of the S1 element) to the diffraction plane of element S1. The images show unwrapping of the vortex into Cartesian coordinates and finally an OAM spectrum. In Figure 53(c), the logarithmic radial coordinate is on the horizontal axis, while the azimuthal coordinate is on the vertical axis. If element S2 works perfectly, then the result is expected to comprise vertically-separated peaks that describe the OAM spectrum of the beam. In practice, slight bending is observed, perhaps as a result of residual charging of the needle or imperfect centring of the needle on the optic axis.

Experimental spectra corresponding to different electron beam combinations are shown in Figure 54 for synthetic holograms with the indicated nominal OAM decompositions. A near-quantitative comparison with an independent evaluation (based on image analysis) of the OAM composition for a superposition of beams with $\ell = \pm 5$ is reported in the Supplementary Material, alongside secondary electron microscopy (SEM) images of generation holograms. The experimental cross talk between different OAM channels is shown in Figure 54(b) for input beams with values of ℓ between 0 and 6. The values of cross talk depend on the specific details of each experiment.

The OAM resolution $\Delta\ell$, which is defined here as the full width at half maximum of the spectrum, is found to be between 1.5 and 2, compared to an optimal value of unity [138]. This discrepancy is thought to result from imperfect alignment, while the background under each peak is likely to result from a slight mismatch between the size of the diffraction pattern of element S1 and the size of element S2. The experimental OAM spectrum corresponding to a "standard" (azimuthally uniform) beam shown in Figure 54(c) is reproduced nearly perfectly in a simulation shown in Figure 54(d) by assuming a 1.1% error in the excitation of S1 and an 80 nm lateral misalignment between the two sorting elements(See the Supplementary Material). The image shows also the comparison experiment and simulation for one case of a near complete correction of effect. A more thorough comparison of simulation and experiment will be explained in a further coming paper.

A “standard” beam obtained using a conventional condenser aperture should have an azimuthally-uniform wavefront. The beam that enters the sorter is almost exactly in a plane conjugate to that of the condenser aperture and should decompose to only an $\ell = 0$ component. The observed fringes result from misalignment. The absence of practically any material in the electron beam path (excluding the Sorter 1 electrode) should rule out a spurious effect from the generation hologram. The simulations shown in Figure 54(d) and an analytical model presented in the Supplementary Material suggest that this mismatch is also responsible for the slight loss of resolution in the peaks.

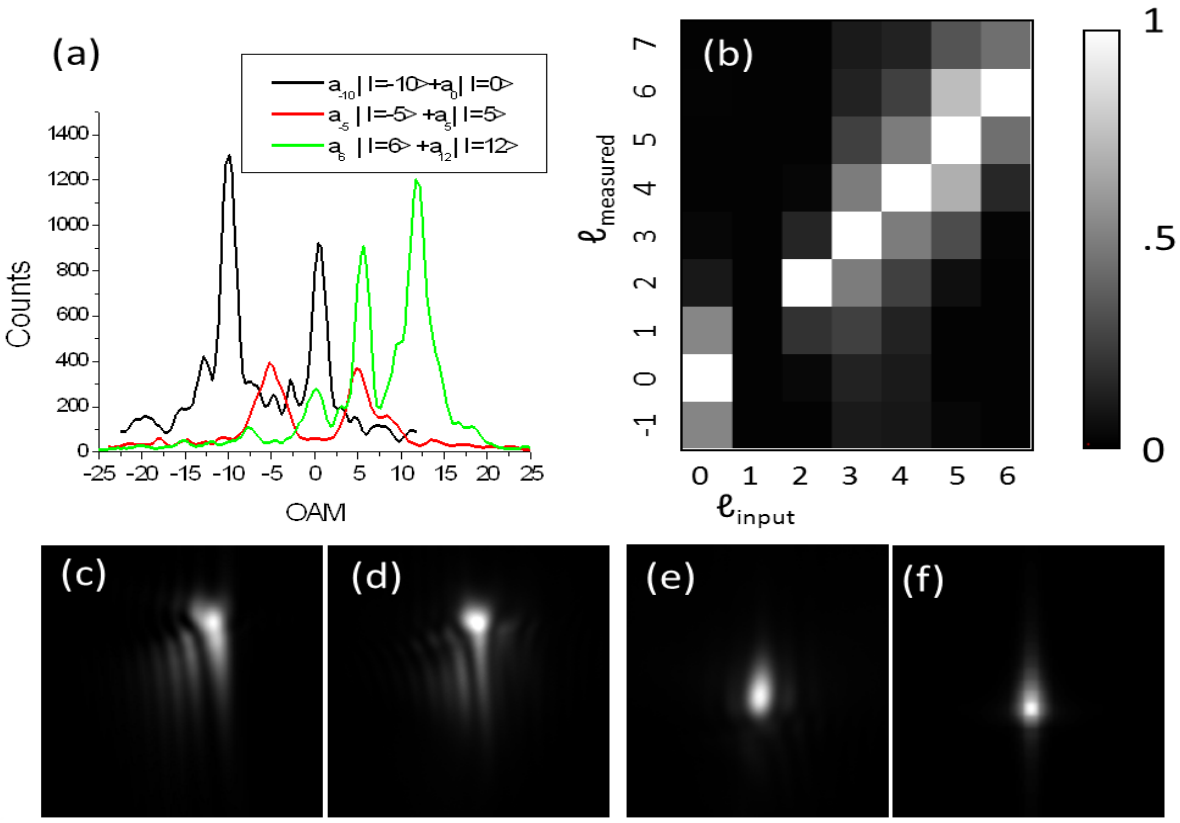

Figure 54: (a) Experimentally measured OAM spectra for three test electron beams with the indicated nominal OAM compositions. (b) Experimental evaluation of cross talk between the nominal OAM value and the measured distribution. The figure is normalized so that the maximum value in each column is 1. No data are available for ℓ_{input} =1 since no hologram with ℓ_{input} =1 was created (it would have created a confused spectrum with no possibility of internal calibration).(c,e) Experimental and (d,f) simulated OAM spectra for a “standard” (ℓ = 0) beam. A match is obtained for incomplete(c,d) and near complete (e,f) compensation between diffraction from S1 and the phase of S2. (See text for details).

On the assumption that the phase mismatch is the primary limiting factor, two solutions can be proposed to improve the resolution:

1) Since the phase mismatch is proportional to the phase gradient, the OAM resolution can be improved by reducing *s*, *i.e.*, the voltage applied to element S1, while maintaining the stationary phase condition $\frac{sR}{\lambda} \gg \ell_{MAX}$ [143], where *R* is the size of the beam at the entrance of the sorter and ℓ_{MAX} is the maxmum value of OAM to be measured. If this approach is used, then the product *s f* must be

matched to the periodicity of S2, which is fixed by the geometry of the S2 electrodes. For this reason, we increased the focal length by switching off the aberration corrector.

2) In order to limit the phase gradient (and considering the geometric imperfection of sorter 1), it helps to limit the angle of acceptance of the sorter, for example by using an aperture. In our tests without a sample, we simply reduced the convergence angle for this purpose.

Given the need for precise alignment, we believe that automatic diagnosis and alignment will be required regularly to reduce shift and magnification errors, in a manner similar to that used in adaptive optics in telescopes [241] and for the adjustment of spherical aberration correction over time (see [241] and references therein).

In summary, we have described a reliable OAM sorter for electron beams that is based on electrostatic phase elements and makes use of multiple-electrode-based control of electron wave functions. Experimental OAM spectra of test beams are consistent with predictions. The OAM resolution is measured to be between $\Delta\ell \approx 1.5$ and 2. In the future, improved alignment using computer-based diagnostics is expected to provide an optimal resolution of $\Delta\ell = 1$ with a reduced background signal. The attainment of an ideal OAM resolution with the maximum lossless efficiency of the electrostatic approach has been already foreseen as the key in close future to atomic column resolution magnetic measurements trough dichroism [241].

More in general our demonstration of the use of MEMS-based technology to fabricate new components for charged particle optics offers a breakthrough for the measurement of previously inaccessible physical quantities, which are not limited to orbital angular momentum [240].

Supplementary information

A1: Complex notation expanded

According to the main text, the phase distributions introduced by the sorters are

$$\varphi_1 = \frac{s}{\lambda}\,\Re e\left\{\mathrm{u}\,\ln\left(\frac{u}{L}\right) - u\right\}, \tag{214}$$

$$\varphi_2 = \frac{Ls}{\lambda}\,\Re e\left\{\exp\left(\frac{u}{sf}\right)\right\}. \tag{215}$$

Expansion of the complex notation for sorter 1 in the form

$$\varphi_1 = \frac{s}{\lambda}\,\Re e\left\{(\mathrm{x}+\mathrm{iy})\left[ln\left(\frac{\sqrt{x^2+y^2}}{L}\right) + i\mathrm{atan}\,(y,x)\right] - (x+iy)\right\} \tag{216}$$

leads to the expression

$$\varphi_1 = \frac{s}{\lambda}\left\{x\,ln\left(\frac{\sqrt{x^2+y^2}}{L}\right) - y\,\mathrm{atan}(y,x) - x\right\} \tag{217}$$

For sorter 2,

$$\varphi_2 = \frac{Ls}{\lambda}\left\{\exp\left(\frac{x}{sf}\right)cos\left(\frac{y}{sf}\right)\right\} \tag{218}$$

These are the equations that are conventionally used to describe an OAM sorter. Their gradients can be defined in the form of Wirtinger derivatives $\frac{\partial\varphi_1}{\partial\bar{u}} := (\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$ as

$$\begin{aligned}\frac{\partial\varphi_1}{\partial\bar{u}} &= \frac{s}{\lambda}\, ln(u/L)\\ \frac{\partial\varphi_2}{\partial\bar{u}} &= \frac{L}{f\lambda}\exp\,(u/sf)\end{aligned} \tag{219}$$

A2: Holograms and generated beams

We reported scanning electron microscopy (SEM) images of some of the "generator" holograms. The first example corresponds to an inline vortex, which is characterized by the thickness function

$$T = t_1 \bmod(n\,\theta, 2\pi) + t_0 \tag{220}$$

where "mod" refers to a remainder following division of the two arguments. The OAM spectrum of such a mask has been calculated in Ref. [23].

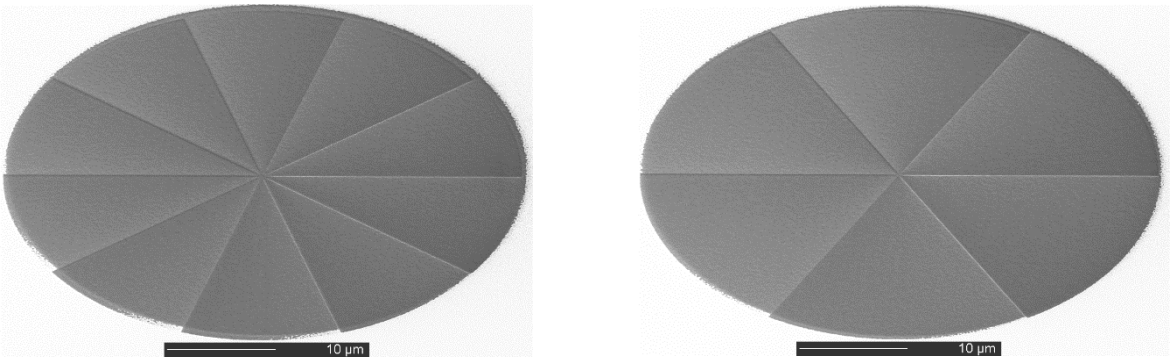

Figure 55: Tilted SEM images of holograms with angular frequencies corresponding to ℓ = 10 and ℓ = 6.

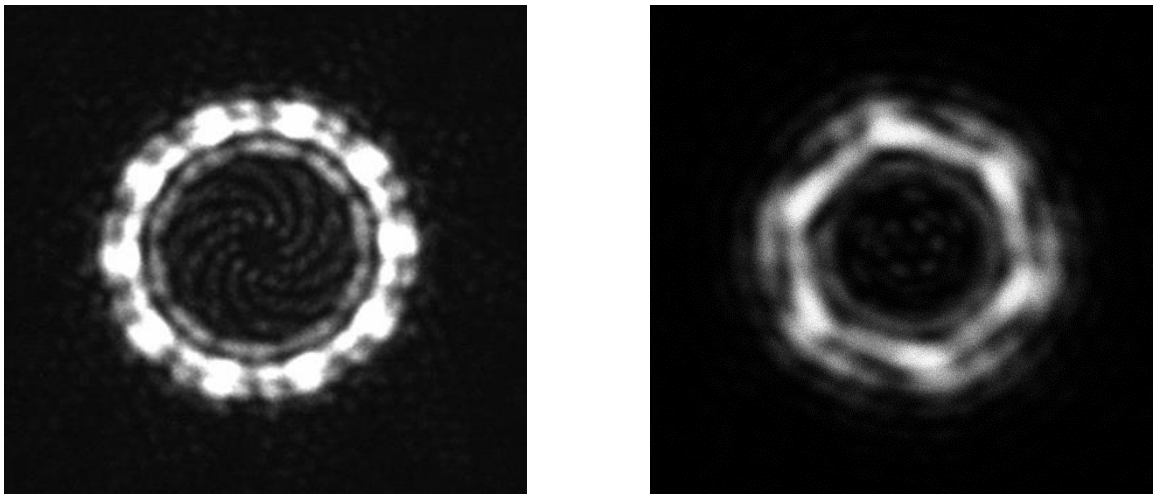

Figure 56: Beams corresponding to the holograms in *Figure 55*. 10-fold and 6-fold modulations in intensity result from the presence of a strong ℓ=0 component, which is visible in experimental OAM spectra.

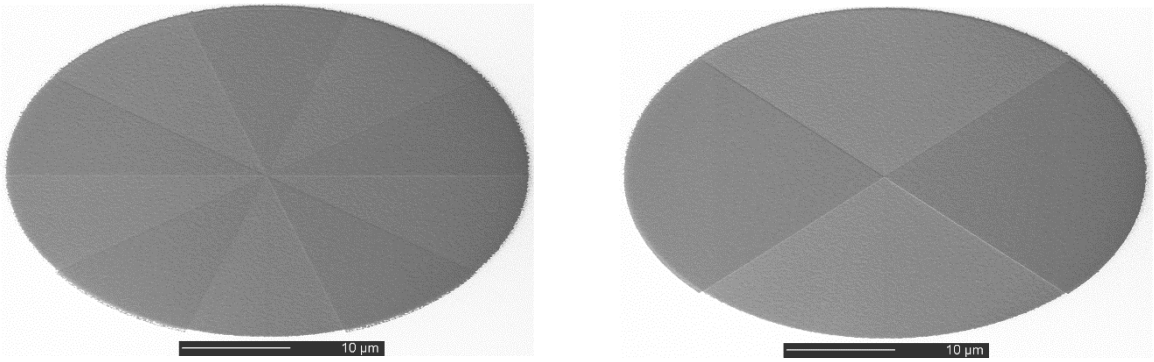

Figure 57: Tilted SEM images of holograms for the generation of petal beams for ℓ = ±5 (left) and ℓ = ±2 (right).

The second mask is based on a thickness modulation of the form

$$\mathrm{T} = \begin{cases} t_0 & if\ 0 < \mathrm{mod}(n\theta, 2\pi) < \pi \\ t_0 + t_1 & if\ \pi < \mathrm{mod}(n\theta, 2\pi) < 2\pi \end{cases} \tag{221}$$

A3: Validation of OAM decomposition

In this section, OAM decomposition is validated by using direct images of the electron beam in real space with the probe close to its waist.

Normally, the determination of the OAM decomposition of a beam from a real space image is impossible because the phase of the beam is lost during the measurement. However, in the case of a petal beam, which is given by a superposition of beams of vortices with opposite OAM $\ell = \pm n$, one can simplify the calculation since the beam wavefunction only has a real component. In fact, by writing a vortex as $\psi = R_{|\ell|}(\rho)\exp(i\,\ell\theta)$ where $R_{|\ell|}(\rho)$ is a radial function the corresponding petal beam is $\psi = R_{|\ell|}(\rho)\sin(\ell\theta)$.

The inline hologram that we used to generate the petal beams is described by Eq. 221. For this specific hologram, the angular and radial degrees of freedom are decoupled and the azimuthal part of the wavefunction can be factorized in the form

$$\Theta(\theta) = \sum_{\ell} c_\ell \exp(i\,\ell\theta) \tag{222}$$

The coefficients of OAM expansion for a perfect phase hologram are

$$c_\ell = \begin{cases} \dfrac{-2ni}{\ell\pi}\exp(i\delta/2)\sin\left(\dfrac{\delta}{2}\right) & if \quad \ell = mn \\ \exp(i\delta/2)\cos\left(\dfrac{\delta}{2}\right) & if \quad \ell = 0 \\ 0 & otherwise\ , \end{cases} \tag{223}$$

where δ is the phase corresponding to a depth t_1 of milling and m is an odd integer.

To a first approximation, the series can be limited to only the first terms, for which $\ell = \pm n$ and $\ell = 0$. The wavefunction is then given by the expression

$$\psi = 2\,|c_n|\sin(n\theta) + \mathrm{i}|c_0| \tag{224}$$

The expression differs by that of an exact petal beam because of the presence of the term c_0 that arises if the hologram depth is not perfectly calibrated to have $\delta = \pi$, since in practice it is very difficult to calibrate with such high precision the milling.

The absolute value has been here used to highlight that for a perfect phase hologram the 0th order and the first nonzero diffractions have exactly a $\pi/2$ phase difference.

The resulting image can be expressed in the form

$$I(\theta) = |\psi|^2 = 4|c_n|^2\sin^2(n\theta) + |c_0|^2 = 2|c_n|^2 - 2|c_n|^2\cos(2n\theta) + |c_0|^2 \tag{225}$$

Where again we omitted the radial dependence. In fact, in the approximation that the azimuthal and radial degrees of freedom of the image are factored, the image can be decomposed in terms of azimuthal Fourier components as

$$I(\theta) = \sum_{\ell} I_\ell \exp(i\,\ell\theta) \tag{226}$$

These coefficients can be measured experimentally and compared with the prediction based on the sorter based OAM decomposition.

Azimuthal Fourier decomposition of the image therefore only has components I_ℓ with ℓ=±2n and 0. By using Eq.225 and Eq.226, we find that

$$\begin{gathered} I_0 = 2|c_n|^2 + |c_0|^2 \\ I_{2n} = |c_n|^2 \end{gathered} \tag{227}$$

A more realistic description of the hologram should include absorption and therefore an amplitude modulation superimposed onto the phase with the same frequency. Although the formulae are given in Ref. [23], it is sufficient to write $\delta \rightarrow \delta + ia$, where $a \ll \delta$. In the same way, Eq. A8 can be rewritten with real coefficients $c_{1r}, c_{1i}, c_{0r}, c_{0i}$ in the form

$$\psi = 2(c_{1r} + ic_{1i})\sin(n\theta) + \mathrm{i}(c_{0r} + ic_{0i}) \tag{228}$$

where $c_{1i} \ll c_{1r}$ and $c_{1i} \ll c_{1r}$. After a few simple steps, we find that the main contribution of the additional terms is

$$I_n = -4c_{0i}c_{1r} + 4c_{0r}c_{1i} \tag{229}$$

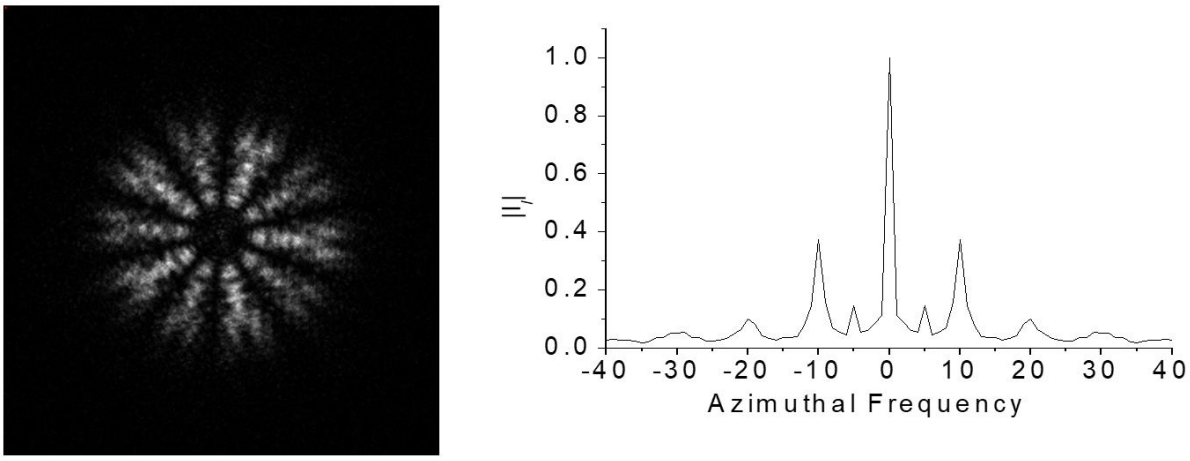

Figure 58: (a) Image of a petal beam of Figure 57(left) corresponding to $\ell = \pm 5$. (b) OAM coefficients of this image obtained by its digital analysis.

We therefore expect an additional Fourier component that is related to the superimposed and unwated amplitude modulation effect of the hologram, whereas the contribution from such components is very small.

An experimental defocused image of a petal beam generated by the hologram in Figure 57(left) is shown in Figure 58 alongside a digitally-calculated azimuthal Fourier decomposition of the same image. The procedure for extracting the I_ℓ coefficients follows previous work [23] and is based on a digital version of the OAM sorter, where the intensity is mapped to polar coordinates and digitally Fourier transformed. The results indicate that the dominant azimuthal Fourier coefficients I_ℓ correspond to $\ell = \pm 10$ and 0. This is consistent with the assumption that the meaningful values of c_ℓ correspond to $\ell = \pm 5$. Moreover, there is a slight absorption effect that introduces an image term I_ℓ at $\ell = \pm 5$.

Using both expressions in Eq.227 we can find $|c_0|^2 = I_0 - 2I_{2n}$. Using this equation, we find that a contribution c_0 is present, with an intensity ratio $\frac{|c_0|^2}{|c_1|^2} \approx 50\%$ that is approximately consistent with the spectrum shown in Figure 54.

A4: Analytical model for phase mismatch

The experimental OAM spectrum from a uniform beam is not a single sinc peak, but contains extra oscillations and broadening, We show here that these deformations arise primarily from a size mismatch, which can be estimated from Eq.213 and expressed in a real coordinate formalism using Eq.218 or otherwise by

$$\varphi = \frac{Ls}{\lambda} \cos(Kx') \exp(-Ky') \tag{230}$$

where $K = \frac{1}{sf}$.

The beam can be approximated as a line (as shown in Figure 59 and Ref. [23]) at a given value of $y' = sf \ln\left(\frac{R}{L}\right)$, which corresponds to the outer rim of the beam in the S1 plane at radius $R = \sqrt{x^2 + y^2}$.

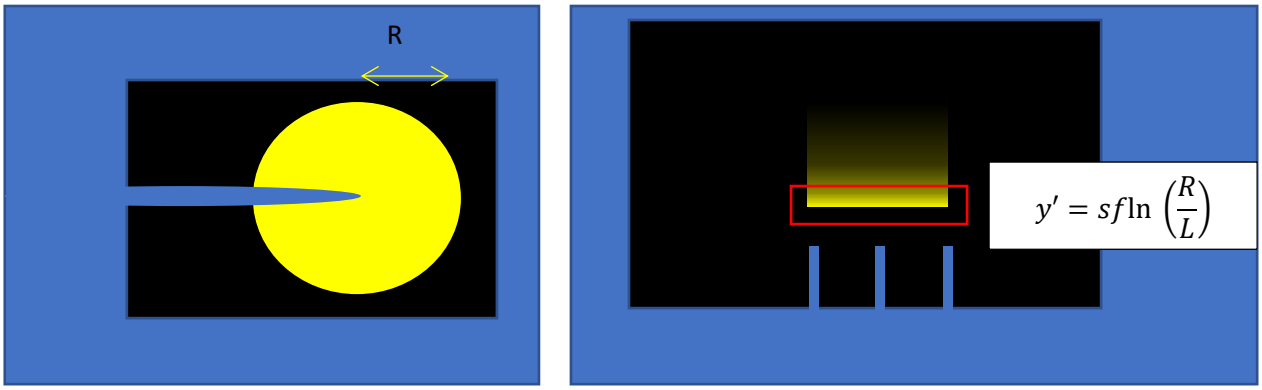

Figure 59: Schematic illustration of a uniform beam approaching sorter 1 and sorter 2.

The unmatched phase that results from the mismatch $m = \Delta K/K$ between element S2 and the size of the diffracted beam is

$$\Delta\varphi = \frac{s\mathrm{R}}{\lambda}\left[\cos\left(\frac{x'}{sf}\right) - \cos\left(\frac{x'}{sf}(1+\mathrm{m})\right)\right] \approx \frac{s\mathrm{R}}{\lambda}\mathrm{m}\left(\frac{x'}{sf}\right)^2 - \frac{s\mathrm{R}}{\lambda}\frac{1}{6}\mathrm{m}\left(\frac{x'}{sf}\right)^4 \quad (231)$$

where Taylor series in *m* and *x'* have been truncated at 4th order. The quadratic term can be difficult to notice in experiments, since the microscope operator adjusts parameters such as defocus, which can compensate for it. The leading uncompensated term produces a maximum phase shift $\Delta\varphi_{max} \approx 16\ sR\ m/\lambda$. For $sR/\lambda = 1$ and a value for *m* of 4%, the phase profile has the shape shown in Figure 60.

A "flat" phase region is obtained over approximately half of the angular range. Changing the value of *m* does not change the shape of the curve, but only the phase scaling factor. One can conclude that: 1) The central "flat" region is responsible for the main OAM peak, whereas the phase tails are responsible for the OAM spectrum background; 2) In most cases, the flat region extends over only about half of the *x'* range, which accounts for the fact that $\Delta\ell \approx 2$ is observed in most cases; 3) It is possible to improve the OAM resolution by reducing *s, i.e.*, the voltage applied to element S1, or *R*, the size of the beam at the entrance of the sorter.

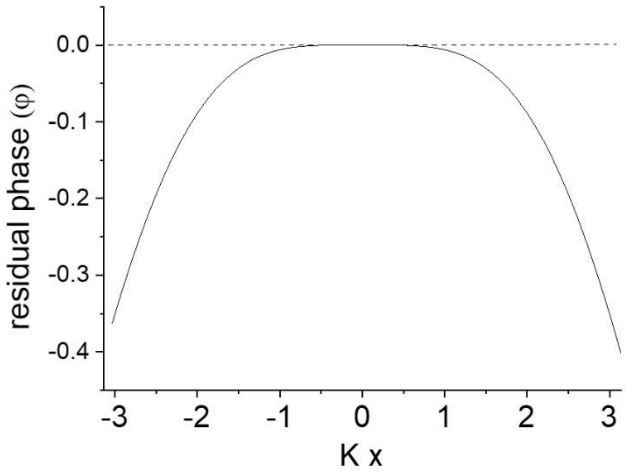

Figure 60: Residual phase when sorter 2 does not perfectly compensate the phase of diffraction of sorter 1.

A5. Lens + sorter simulation

We implemented a numerical method to simulate the OAM spectroscopy experiment based on a full wave calculation and free space propagation, in order to quantitatively describe the effects of several aberrations and misalignments on the final resolution. The beam is free-space propagated between the elements using the Fresnel-Kirchhoff integral

$$U_z(u,v) = \frac{e^{ikz}}{i\lambda z}\iint U_0(x,y)\, e^{-ik\frac{xu+yv}{z}} dxdy \tag{232}$$

The lens is defined by a quadratic phase element

$$T = exp\left(\frac{i(x^2+y^2)}{2f\lambda}\right) \tag{233}$$

The sorters are defined by the phase distributions reported in the text. Calculations were performed numerically on an 8K x 8K pixel matrix using an FFT algorithm for convolutions. The code was written in Matlab and C.

Even for the largest allowed sampling, the phase gradient was limited by the number of pixels. In order to correctly compute the lens effect, the phase should not vary too fast. The phase difference between adjacent pixels should typically be less then π, such that

$$\frac{d\,\phi}{d\,n} < \pi \tag{234}$$

Based on this criterion, the minimum usable focal distance is

$$f_{min} = \frac{2L^2}{\lambda n} \approx 198\, mm \tag{235}$$

We are therefore limited to relatively large focal distances, making it impossible to numerically simulate the objective lens (*i.e.*, a thick and strong lens).

Fortunately, the curvature of the diverging beam from the sample and the focusing effect are nearly compensated at the entrance of the sorter in the objective back focal aperture. We can therefore simulate an equivalent configuration, in which the sorter 1 element is illuminated by a weakly convergent beam.

A second lens is finally placed after sorter 2. The simulated optical configuration is shown in Figure 61.

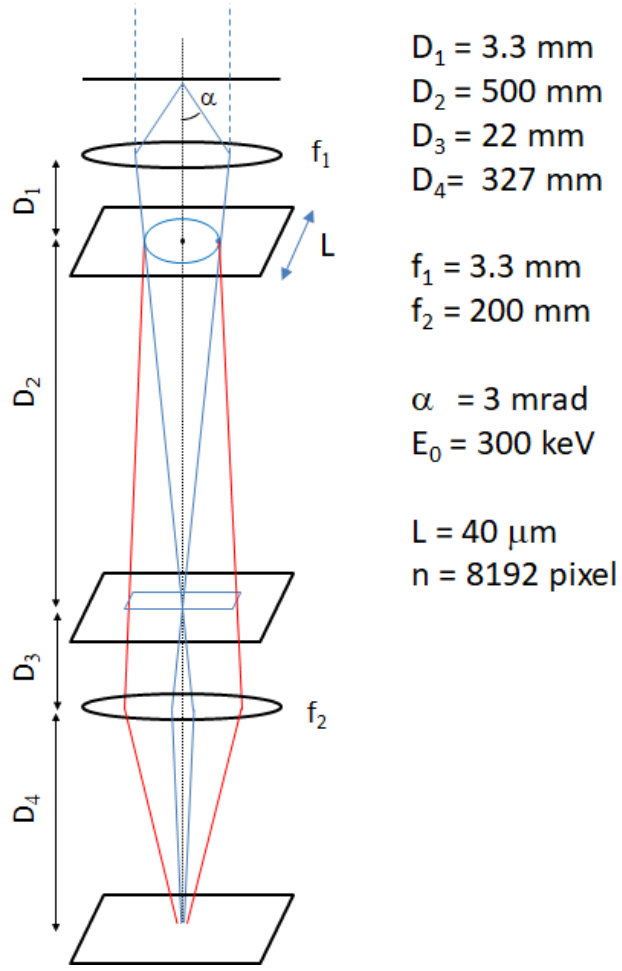

Figure 61: Schematic illustration of the sorting apparatus, including two main lenses.

We used this approach to analyse the effect of different parameters on OAM spectrum resolution. The more detrimental parameters, which should therefore be addressed with care, include misalignment between the two sorter elements, their relative rotation, the size mismatch and the defocus of the lenses. With the exception of size mismatch and relative rotation, these factors can be easily adjusted in real time during operation.

Chapter 6

"Alignment of electron optical beam shaping elements using a convolutional neural network"

E. Rotunno[1], A.H. Tavabi[2], P. Rosi[3], S. Frabboni[1,3], P. Tiemeijer[4], R.E. Dunin-Borkowski[2], V. Grillo[1]

1) *Istituto di Nanoscienze - CNR, 41125 Modena, Italy*
2) *Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany*
3) *University of Modena and Reggio Emilia, via G. Campi 213/A, 41125 Modena, Italy*
4) *Thermo Fisher Scientific, PO Box 80066, 5600 KA Eindhoven, The Netherlands*

Abstract

A convolutional neural network is used to align an orbital angular momentum sorter in a transmission electron microscope. The method is demonstrated using simulations and experiments. As a result of its accuracy and speed, it offers the possibility of real-time tuning of other electron optical devices and electron beam shaping configurations.

Forewords

In this work we report how we can use the novel and hot topic of machine learning to electron microscopy, in particular applied to the OAM sorting of electron vortex beams. For this work I helped in writing down the manuscript and collected some of the experimental results that have been analysed and compared to simulations to validate this approach. It must be noted that, while we used machine learning for pattern recognition to understand which detrimental effects where present in our system and what was their magnitude, this method can be generalized to eventually correct aberration, for example.

Introduction

The evolution of resolution and optics in electron microscopy has involved major steps, including the introduction of spherical aberration correction [33,35], chromatic aberration correction [242,243] and monochromators [222,244–246]. Each increase in complexity has resulted in an increase in the difficulty of instrument control. Although the basic concept of the operation of each lens and optical element is known, the overall behaviour of the microscope is not predictable in detail and the quality of microscope performance is limited by the skill of the user.

One of the most significant recent developments in electron microscopy is electron beam shaping through the use of material-based holograms [113,115,116,207] and, more recently, electron optical components based on microelectromechanical systems technology [121,139,143,186]. Electron beam shaping can be used to generate vortex beams [21,22,216,247], non-diffracting beams [90], compact aberration

correctors [213–215] and analysers of quantum states by means of unitary wave transformations [23,139]. These ideas are often inspired by light optics, where their implementation is easier. Both standard electron beam control and new ideas of electron beam shaping require more automated control of the electron column, both to increase the speed and reproducibility of electron optical alignment and to reduce the demand on the operator.

In visible light optics, every element can be positioned manually and aligned separately, with, for example, adaptive optics providing improvements in telescope optics [241]. In electron microscopy, hardware aberration correctors formed from high order multipoles permit the limitations of cylindrically symmetric lenses to be circumvented. They are presently controlled using semi-analytical models [248–251] based on parametrized aberrations and on the effect of each multipole excitation. However, such an approach cannot easily be adapted to more unconventional optics, such as electron beam shaping. Moreover, a general approach is required to control a full microscope.

Here, we use a convolutional neural network (CNN) [252,253], which is now implemented in many other scientific disciplines [254–258] and is able to learn from a large set of training images to extrapolate a detail or the value of a parameter tagged to each image [259,260]. The success of the technique is based on the fact that it permits any parametrical space to be treated, no matter how complex, provided that enough data are fed to the learning algorithm. It is also often more “robust” to noise than an analytical model.

Here, we are motivated by the specific case of an orbital angular momentum (OAM) sorter [23,138,139], which makes use of electron beam shaping to measure an electron beam’s component of OAM in the propagation direction by decoupling the azimuthal and radial degrees of freedom. Apart from diffraction, it provides the first complete example of a lossless unitary base change that “diagonalizes” a quantum operator using wave manipulation. Recent research [240] suggests that it could be the first of many useful wave transformations to revolutionize the concept of measurement in electron microscopy. The implementation of an OAM sorter requires precise alignment and control of two optical phase elements. No simple analytical model can be used to completely predict the effect on a final OAM spectrum of the available control parameters. In this paper, we perform a quantitative comparison with experiments to show the use of a CNN to control the alignment parameters of an OAM sorter.

Methods

- **Sorter misalignment**

An OAM sorter comprises two phase elements, which can each be based on synthetic holography or electrostatic potentials and are used to spatially separate the different OAM components of an electron beam. In the stationary phase approximation, the first element imparts a coordinate transformation from Cartesian to log-polar coordinates to an electron wave. The transformation phase is removed by exact phase

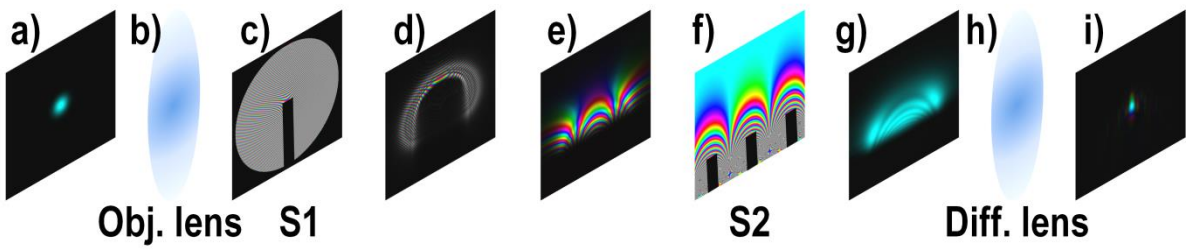

Figure 62: Schematic diagram showing the evolution of an electron beam through an orbital angular momentum sorter. See text for details.

compensation of second phase elements positioned in the Fraunhofer plane of the first element. The working principle is shown schematically in Figure 62.

In the most natural implementation of the OAM sorter, the microscope is operated in scanning TEM (STEM) mode, with a small electron probe formed in the sample plane (Figure 62a). The electron beam is then diffracted by the objective lens (Figure 62b) into a disk in its back focal plane, where the first element of the sorter is located (Figure 62c). The phase shift associated with the first element (S1) can be described mathematically by the expression:

$$\varphi_{S1}(x,y) = \frac{ks}{f}\left(y\tan^{-1}\frac{y}{x} - x\log\left(\frac{\sqrt{x^2+y^2}}{L}\right) + x\right) \tag{236}$$

where k is the electron wavevector, f is the focal distance between the two sorter elements and s and L are scaling parameters. Element S1 can be fabricated using a long electrically charged conductive needle [139]. The scaling parameters s and L depend on the electrical bias applied to the needle and on its length, respectively [122]. Element S1 introduces a conformal mapping from Cartesian to log-polar coordinates (Figure 62c and d), which is concluded in the selected area aperture plane (Figure 62 f). Element S2 is located in this plane, which is conjugate to the sample plane. In its experimental realization, it comprises a periodic array of conductive needles, which are charged positively and negatively in an alternate fashion. The phase shift introduced by element S2 can be written in the form

$$\varphi_{S2}(u,v) = -\frac{ksL}{f}e^{-\frac{u}{s}}\cos\left(\frac{v}{s}\right), \tag{237}$$

where *u* and *v* are coordinates in the diffraction plane. This phase shift is designed to match and compensate (Figure 62g) the phase shift of the electron beam, in order to prevent further S1-based evolution of the electron beam shape upon propagation. A second lens (Figure 62h) collapses the electron beam to a rod (Figure 62i), whose position with respect to the optical axis is proportional to the OAM carried by the electron beam. If several OAM components are present, then each component is focused to a different position, thereby generating a spectrum.

As a result of the stationary phase condition, the phase shifts of elements S1 and S2 vary rapidly. In the S2 plane, a phase change of $\Delta\phi = 2\pi$ typically occurs over less than 200 nm (in the conditions used in the

present work). Therefore, precise alignment of the elements is important to obtain complete phase compensation. This compensation involves controlling rotation, size, focus and translation.

Figure 63 shows the influence of different parameters on the resolution of an OAM spectrum. The three columns show electron optical simulations of the primary misalignments between the diffraction of S1 (left column) and S2 (centre column) and their effect on an OAM spectrum (right column). The simulations were

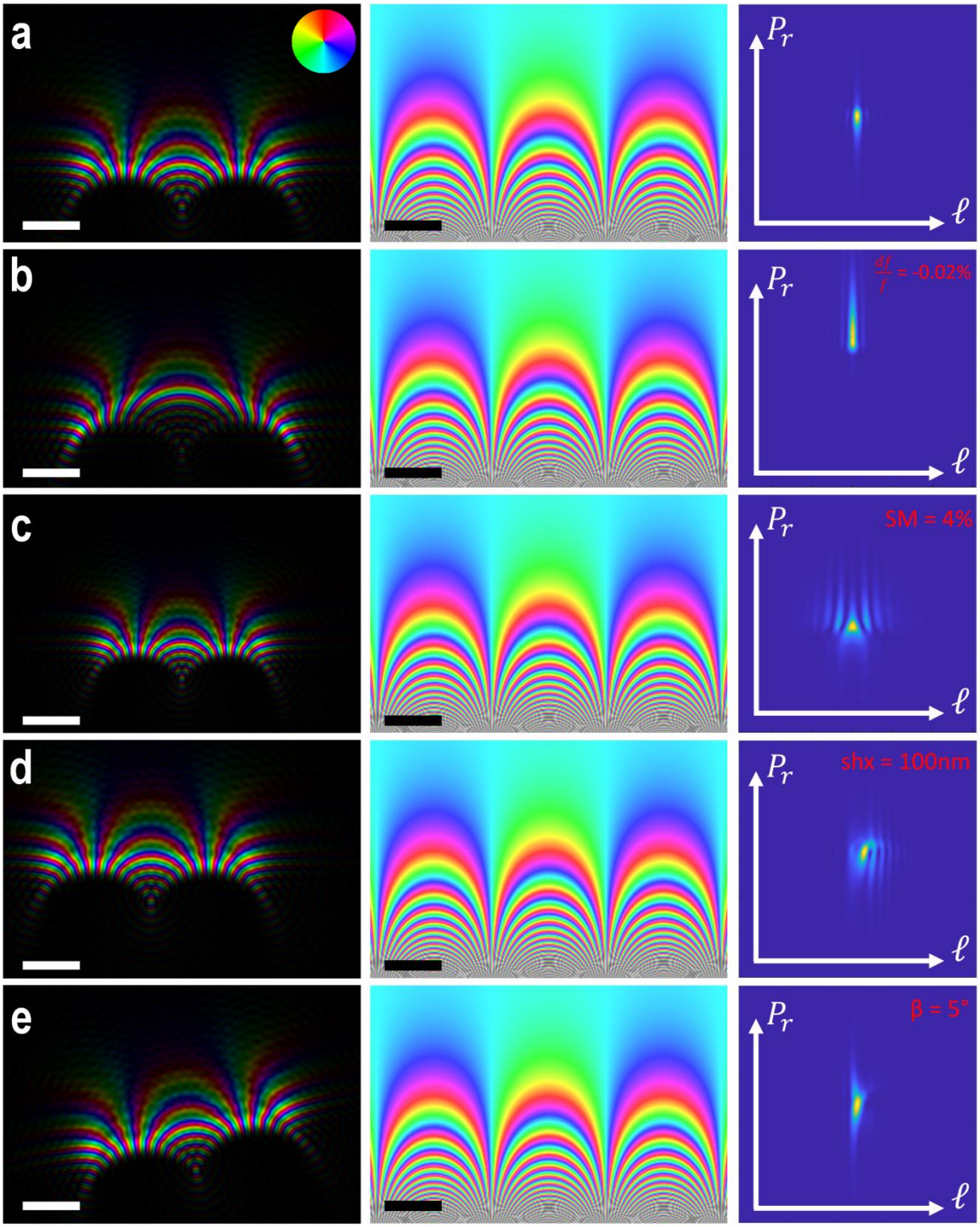

*Figure 63: Primary misalignment conditions between the diffraction of S1 (left column) and S2 (centre column) and their effects on an OAM spectrum (right column). a) Perfect condition, with the beam and S2 aligned. b) Effect of objective lens defocus (*df*). c) Effect of size mismatch (*SM*). d) Effect of rigid shift (*shx *and* shy*). e) Effect of rotation (*β*). Scale bar in real space images: 10 µm. The OAM spectrum spans over a range of 40ħ. The real space misalignment (left column) is magnified for better visualization. The real parameters that were used for the simulation are shown as labels.*

performed for a 300 kV electron beam defined by a 2 mrad hard aperture traveling in vacuum. The sorter parameters are $s = 3\ \mu\mathrm{rad}$ and $L = 40\ \mu\mathrm{m}$. The algorithm that was used for the calculations is described below. The first row corresponds to the ideal situation, in which the phase of the beam is matched perfectly and compensated by the phase of element S2 (centre column). The OAM spectrum then only features a narrow spot at $\ell = 0$ (right column). In the second row (Figure 63b), the defocus of the objective lens (df) is considered. The conformal mapping is then not completed in the plane in which element S2 is located. As a result, a residual curvature is present in the beam and the phase pattern is distorted. The main effect of this aberration is broadening of the beam in the radial direction (P_r, which depends on the sign of the defocus. The OAM resolution is preserved instead, meaning that a moderate amount of defocus *f* can be tolerated if only the OAM component is of interest [72,75,197,233,261].

A more prominent source of resolution loss is the size mismatch (SM) between the beam and element S2 (Figure 63 c). This aberration appears when the electron beam scale in the S2 plane is larger (smaller) than the scale dictated by the periodicity of element S2. Since the periodicity of S2 is bound to the physical distance between the electrodes and cannot be modified, correction for this aberration entails changing the beam size. The size of the electron beam in the S2 plane, in turn, depends on the product of the focal distance between the two planes (another physical parameter that cannot be changed) and the excitation parameter s of element S1 (Eq. 236). An optimal potential then needs to be applied to the main electrode of element S1. A deviation from this value introduces an SM aberration to the OAM spectrum. The main effect of the SM aberration is the introduction of background fringes and a loss of OAM resolution. A difference of a few % is sufficient to produce the effect reported in the right column of Figure 63c.

The third effect is associated with a rigid shift of the electron beam in one or both directions (*shx* and *shy*) with respect to element S2. The primary effect of this misalignment is an asymmetrical background fringe pattern. A misalignment of a few nm, which is small compared to the periodicity of element S2 (typically $20\ \mu\mathrm{m}$), is sufficient to reduce the OAM resolution. Fortunately, TEMs offer high precision control over the beam position and this aberration can be fixed easily, even manually.

The last row reports the effect of relative rotation (*i.e.*, orientation mismatch β) of the electron beam with respect to element S2.

In addition to these effects, which are inherent to S1/S2 alignment, the defocus of the diffraction lens located after element S2 was also considered. Its effect is to modify the shape of the OAM spectrum. All other aberrations of the microscope, in particular astigmatism in different planes and coma, were neglected. At this stage, we assume that precise alignment is carried out before starting an experiment and it is not affected by a change of the sorter parameters. Aberrations of the main lenses can be included in the training of the CNN by adding more adjustable parameters. However, special care is devoted to the finite lateral coherence of the beam, as described below.

Even for a perfectly aligned OAM sorter, other issues can affect final experimental spectra. In particular, if the electron beam is not centred with respect to the optical axis in the sample plane or S1 plane, the OAM spectrum will be broadened. This is a normally unwanted effect that arises from the physical definition of OAM. The OAM operator, like its classical counterpart, is related to the specific choice of a pole, which corresponds here to the optical axis.

- **CNN and training dataset**

Figure 64 shows the structure of the CNN that was optimized here. It is composed of 5 convolution layer filters, each of which is followed by an average-pooling layer filter. Two fully connected layers lead to the output, corresponding to a total of 2,480,326 trainable parameters. The chosen activation function was the rectified linear unit (ReLU). The learning algorithm used was Adam [262] and the learning rate was 0.001, while the loss function was the root mean square difference between the predicted misalignment coefficients and the true coefficients. The neural network was implemented using the Keras library [263] and the TensorFlow backend [264].

The CNN was trained on a dataset of 20,000 simulated images (+ 2000 images for validation) for random values of the six misalignment parameters.

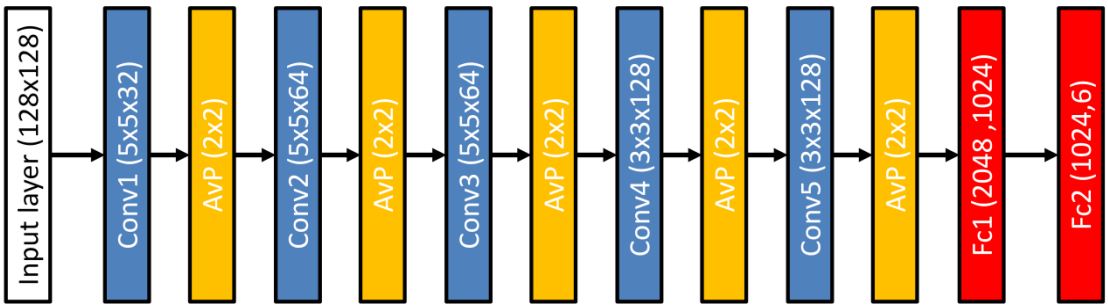

Figure 64: Structure of the optimized convolution neural network. Blue, yellow and red boxes correspond to convolution layers, average pooling layers and fully connected layers, respectively. Filter sizes are shown between parentheses.

Training can also be performed experimentally by changing the sorter parameters. Such a procedure is expected to improve the fitting because it will not be limited to an ideal numerical model. However, it requires complex automatic control of the microscope and long-term stability of microscope alignment. These problems are avoided here by training the CNN on simulated images.

In order to produce the training dataset, a numerical method was used to simulate the OAM spectroscopy experiment, based on a full wave calculation and free space propagation, in order to describe the effects of different aberrations and misalignments on the final resolution. The beam was free-space propagated between the elements using the Fresnel-Kirchhoff integral

$$U_z(u,v) = \frac{e^{ikz}}{i\lambda z} \iint U_0(x,y)\, e^{-ik\frac{xu+yv}{z}} dxdy \tag{238}$$

while the lenses were defined by quadratic phase elements of the form

$$T = exp\left(\frac{i(x^2 + y^2)}{2f\lambda}\right) . \tag{239}$$

The calculations were performed numerically on an 8k x 8k mesh using a Fourier transform algorithm for convolutions. Versions of the code were written in Matlab (for testing purposes) and C (for fast parallel computing). Entries from the training database are shown in Figure 65 (top line).

Before being fed to the CNN, spectra were pre-processed to account for decoherence effects. Spatial and temporal coherence effects in electron microscopy are usually described by convolution and/or multiplication by suitable damping functions. A systematic treatment of coherence effects on the resolution of an OAM sorter is beyond the scope of the present paper and will be presented elsewhere. We accounted for decoherence using Monte Carlo simulations by averaging over OAM spectra obtained for different beam positions in the sample plane. The size of the deflection was evaluated by considering broadening of the probe between 0.5 and 1 Å. When magnification and propagation were accounted for, a best match was obtained for Gaussian broadening of 0.5ħ in the OAM direction. Conversely, nearly no broadening appeared to be necessary in the radial direction. Without accounting for decoherence, the CNN did not converge to the correct value.

The CNN was trained on the simulation database for 20 cycles through the full training dataset (epoch), with a computation time of 380 s per epoch, reaching a limit RMS error on the predicted parameter of 0.0038. Training was interrupted after 20 epochs, as the CNN started to show signs of overtraining, *i.e.*, the RMS error on the training dataset became lower than that estimated on the validation dataset, suggesting that the CNN was memorizing features from the training images and losing its generalization ability. In

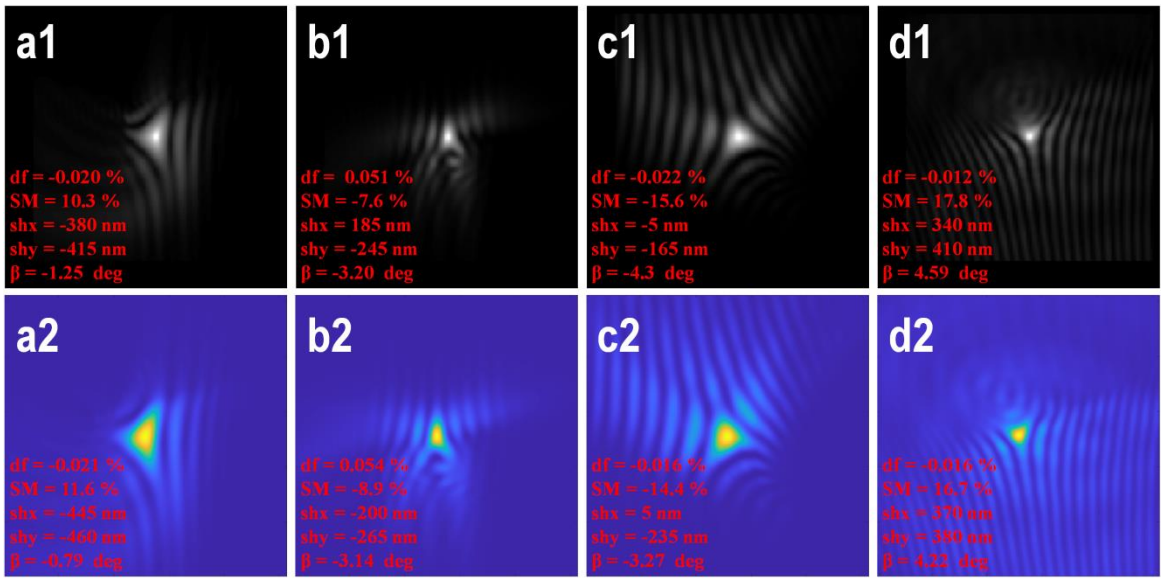

Figure 65: Demonstration of the predictive ability of the CNN for a comparison between simulated reference images (upper row) and simulations performed using the parameters predicted by the CNN (lower row).

order to validate the fitting accuracy of the CNN, the predicted misalignment coefficients were fed back into the simulation algorithm. The resulting images (Figure 65, lower row) agree with the real images.

Experimental results

Experiments were performed at 300 kV in a Thermo Fisher Titan G2 60-300 TEM equipped with an X-FEG emitter and an image spherical aberration corrector. The illumination system was set to spot size 9 and the three condenser lens system was used to achieve a probe convergence semi-angle of 2 mrad in the specimen plane. No sample was used, in order to allow free propagation of the electron beam in vacuum. The microscope was operated in "microprobe" mode with the objective lens at a standard pre-set value. The image aberration corrector was switched off, in order to achieve a larger focal distance between the sorting elements. In this configuration, element S2 was located in the diffraction plane of element S1 (mounted in the objective (OBJ) aperture plane), which is conjugate to the sample plane. The OAM spectrum could then be imaged on the detector with the microscope set to diffraction mode. The objective lens current was kept to a standard eucentric preset value, while focusing was achieved by changing the C3 excitation.

The device was first tuned to find a good working condition. The polarization of the main needle of element S1 was set to $V_c = 6.40$ V, corresponding to a scale factor $s \approx 3$ µrad. All experimental images were rescaled and rotated to match the scale and rotation of the training dataset before using the CNN. Calibration was achieved by sorting electron beams with known OAM signatures, as reported in [139]. Recorded OAM spectra were labelled based on the most sensitive parameters, *i.e.*, the main potential applied to element S1 and the main defocus *f*. Mechanical instabilities of the aperture holders in which the sorter elements were mounted made precise calibration impossible.

The upper row of Figure 66 shows experimental spectra recorded for different values of the potential applied to element S1 over a symmetrical range (6 - 6.8 V) about a reference value of 6.4 V, in order to study the influence of SM misalignment. The images were fitted using the CNN. The fitting parameters provided by the CNN were used to simulate spectra, in order to evaluate the qualities of the fits. The simulated spectra are

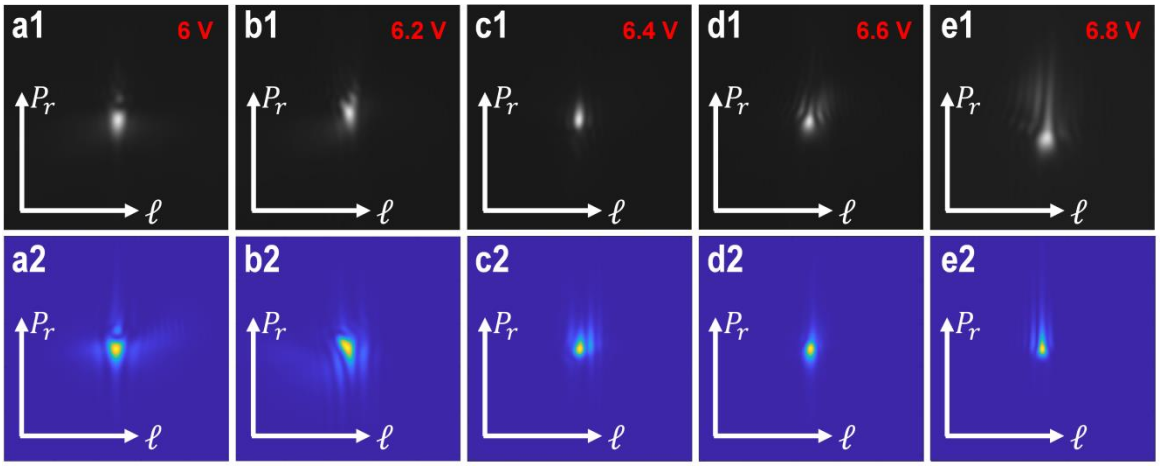

Figure 66: Upper row: Experimental spectra recorded for different values of the potential of element S1 over the symmetrical range (6 - 6.8 V). Lower row: Simulated spectra for best-fitting parameters obtained from the CNN.

shown in the lower row of Figure 66. This comparison demonstrates the ability of the CNN to identify and analyse the main features of the spectra, in particular the arrowhead-like central peak, which points towards or away from element S2 (*i.e.*, up or down in the reference system used in Figure 66), depending on the sign of SM.

The second parameter that was investigated was the defocus of the first lens (*df*). Spectra were again recorded over a symmetrical range about a reference value that provided the best resolution during manual alignment. The experimental results are reported in the upper row of Figure 67, alongside fitting results in the lower row. The CNN is again able to reproduce the experimental images satisfactorily. The primary feature recognized by the CNN is streaking of the spectrum towards or away from element S2 (*i.e.*, up or down in the reference system used in Figure 67), depending on the sign of the defocus.

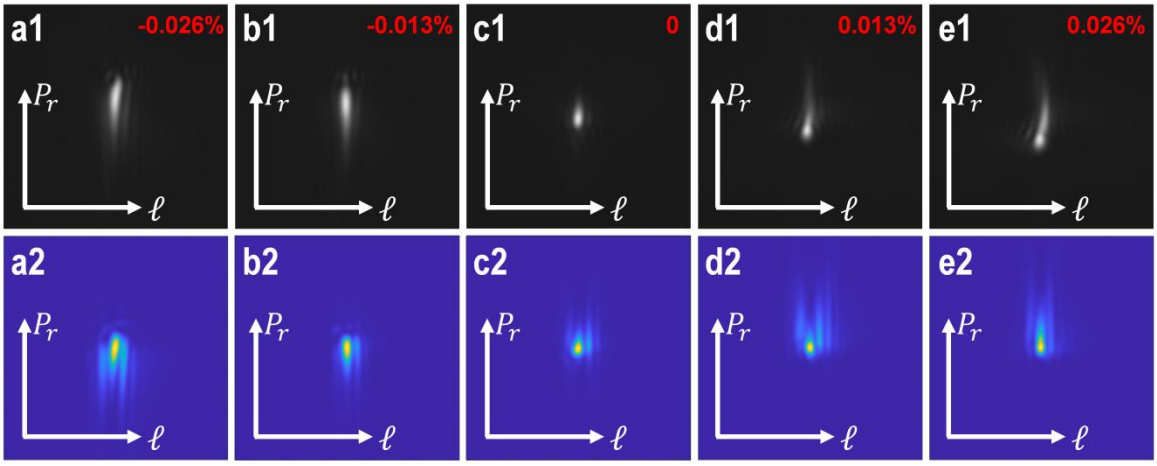

*Figure 67: Upper row: Experimental spectra recorded for different values of the defocus of the objective lens (*df/f*). Lower row: Simulated spectra for best-fitting parameters obtained from the CNN.*

Minor differences between the experimental and simulated spectra are thought to result from the algorithm used for simulation of the training dataset, which is based on an ideal sorter model. In reality, the phase of the sorter may be affected by imperfections in the shapes of the needles or contaminants on them. Conventional lens aberrations have also been neglected in the training model. A possible way to overcome these limitations would be to finalize the CNN training directly on experimental images after pre-training is performed on a simulated database.

Figure 68 shows an assessment of the accuracy of the quantification, in the form of a comparison between the estimated and experimental values. Figure 68a shows the expected linear trend between the estimated SM values and the real voltages applied to element S1. Figure 68b shows a comparison between the estimated defocus values and the real defocus values (Figure 67), which were obtained by varying the current (I) according to the expression

$$\mathrm{d}f = -2f\frac{\mathrm{d}I}{I}\ . \tag{240}$$

Although the linear trend again suggests successful fitting of the sorter misalignments, in practice misalignment effects tend to combine together, resulting in a complicated parametric dependence. The CNN was therefore allowed to extrapolate values of all of the parameters. In our best manual alignment (Figure 66c and Figure 67c), the S1 potential was underestimated and the defocus was overestimated, illustrating

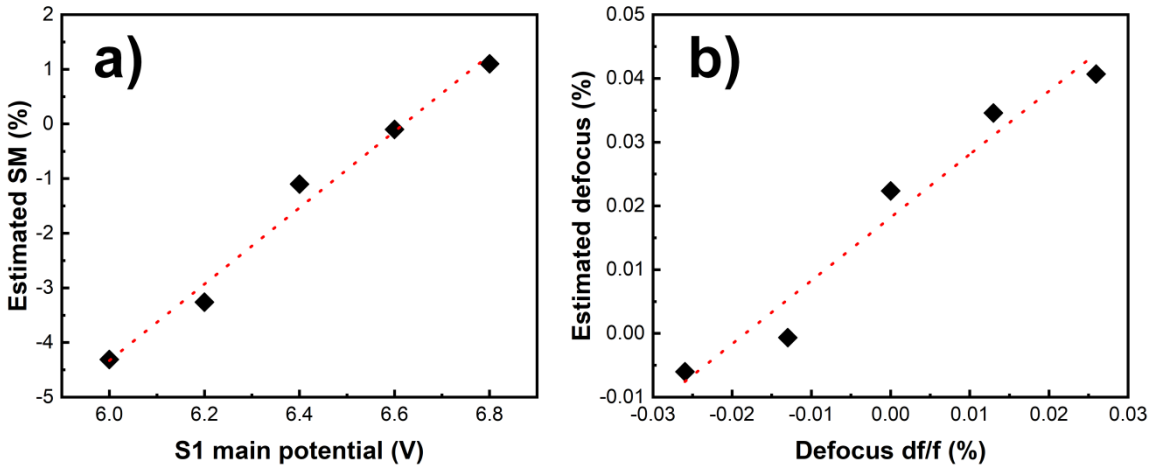

Figure 68: Comparison between experimental sorter parameters and CNN predictions shown as a function of a) S1 main potential and b) deocus (df).

how aberrations can partially compensate for each other, just as spherical aberration and defocus can partially compensate under Scherzer conditions for the electron probe. This compensation highlights the fact that perfect alignment would be difficult to achieve manually.

The fitting of an experimental image using the CNN was found to only take (on average over 1000 iterations) 56 ms on a conventional laptop, including image pre-processing. The computational time is therefore negligible compared to the acquisition time. For this reason, we anticipate that a CNN can provide real-time control and feedback about alignment accuracy during experiments.

Conclusions

We have demonstrated that a neural network can be used to determine alignment parameters for the complex electron optical configuration of an OAM sorter, for which the effects of misalignment cannot easily be managed analytically or adjusted manually. The CNN is capable of determining parameters such as defocus and sorter electrode excitation from a single spectrum image. Such an approach can be applied in real time to align other complex optical systems, such as spherical aberration correctors, based on minimal experimental data. We envisage that in the future experimental devices will be able to self-diagnose and communicate with operators in real time.

Chapter 7

“A sorter for electrons based on magnetic elements”

Giulio Pozzi[1,2], Paolo Rosi[3], Amir H. Tavabi[1], Ebrahim Karimi[4], Rafal E. Dunin-Borkowski [1], Vincenzo Grillo[5]

1. Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany
2. Department of Physics and Astronomy, University of Bologna, 40127 Bologna, Italy
3. Dipartimento FIM, Universitá di Modena e Reggio Emilia, 41125 Modena, Italy
4. Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada
5. Centro S3, Istituto di nanoscienze-CNR, 41125 Modena, Italy

Abstract

The orbital angular momentum (OAM) sorter is an electron optical device for the measurement of an electron's OAM. It is based on two phase elements, which are referred to as an “unwrapper" and a “corrector" and are located in Fourier conjugate planes. The simplest implementation of the sorter is based on the use of electrostatic phase elements, such as a charged needle for the unwrapper and electrodes with alternating charges or potentials for the corrector. Here, we use a formal analogy between phase shifts introduced by charges and vertical currents to propose alternative designs for the sorter elements, which are based on phase shifts introduced by magnetic fields. We use this concept to provide a general guide for phase element design, which promises to provide improved reliability of phase control in electron optics.

Forewords

In this work we propose a novel design for an OAM sorter based on magnetic elements. We propose two different designs for a complete OAM sorting system, one based on elements carrying a current and another based on magnetic materials. For this work I helped in writing the manuscript and worked on the mathematics reported. The possibility to use magnetic elements resolves some of the shortcomings of the electrostatic device such as charging of the elements under the electron beam and chromatic sensibility (i.e., small changes in the energy of the electron beam require to re-tune the electrostatic sorter optics).

Introduction

A key strength of transmission electron microscopy is its ability to make use of information from both diffraction space and real space through the smart use of electron optics [28]. However, not long after the introduction of techniques such as electron magnetic circular dichroism [71], it became clear that electron microscopy requires a technique that can be used to measure an electron's orbital angular momentum (OAM) [21], which is neither a real space nor a diffraction space property. OAM is a quantum vectorial operator, whose component along the average propagation direction for a paraxial beam can be written in

position representation as the azimuthal derivative of the wave. Its eigenstates are the now well-known vortex beams, whose phase winds in the azimuthal direction [20,22,66,265].

The introduction of electron vortex beams has triggered developments in physics that include new measurement tools [67,69,81,92,100]. Whereas the generation of OAM eigenstates is relatively simple, their measurement and decomposition has proved to be more cumbersome. In particular, methods that have been used to measure OAM composition [130,132,141,239] typically suffer from an inability to cleanly separate the radial and angular degrees of freedom of the electron beam [266].

A new device, which is referred to as an “OAM sorter", promises to overcome this problem by introducing a conformal transformation from Cartesian to polar coordinates to separate different OAM components [23,136,138,143]. The realization and application of this device is an active topic in electron optics. It has been realized using both holographic [23] and electrostatic elements [139], based on theoretical concepts [138,143]. However, its implementation is still not straightforward as a result of practical problems. When using electrostatic elements, a needle-shaped electrode is typically used to perform the transformation, while a periodic array of alternatively-charged electrodes in the corresponding diffraction plane is used to correct the transformation phase. When using holograms, similar phase distributions can be imprinted directly in the form of thickness modulations of the diffraction gratings. In the long term, a needle can become contaminated as a result of exposure to the electron beam, resulting in a change in its electrostatic field, whereas charging and wear can affect a grating material. In addition, the shape of a needle needs to be controlled accurately. Whereas a line of constant charge should be placed in the path of the electron beam, the device is in practice based on the use of conductive electrodes that can only be used to control potentials. A further problem in the electrostatic device results from a chromatic effect on the phase.

It is well known that magnetic lenses are less problematic than electrostatic lenses. Here, we introduce the concept of using magnetic fields as phase elements in the sorter by building on a formal analogy between the phase imparted to an electron beam by a charge and a vertical current. This analogy, which has been suggested in recent experimental work [121,230], is presented systematically in the first section of this paper, in order to demonstrate that the effects of electrostatic phase elements can be reproduced by using either currents or permanent magnets. Key advantages of using a current are that it is more stable than a charge distribution, it can be controlled directly and its (magnetic) effect on the phase is independent of incident electron energy. After a short description of the operation of an ideal sorter device, we introduce new configurations of sorter elements that are based on currents and magnets. Just as for the electrostatic case, we discuss additional corrections that are required to remove residual astigmatism and we assess the experimental feasibility of the concept.

The ideal sorter

We begin by recalling, for the sake of completeness, the basic equations that describe an ideal sorter, first developed for light optics. The key component of the optical sorter transforms azimuthal position in the input beam into linear transverse position in the output beam, which is located in the Fraunhofer diffraction plane of the input beam. An input image comprising concentric circles is then transformed into an output image of parallel lines. However, this transformation introduces a phase distortion that needs to be corrected by a second element. The complete system then comprises two optical elements, with the first element transforming the image and the second element correcting for phase distortions introduced by the first element. The second element is positioned in the Fourier plane of the first element, which can be achieved by using a Fourier-transforming lens of focal length f [136]. The phase profile of the transforming optical element, which is referred to as an “unwrapper" or “sorter 1", is given by the expression

$$\phi_1(x,y) = \frac{d}{\lambda f}\left[y\, arctan(x,y) - x\, log\left(\frac{\sqrt{x^2+y^2}}{b}\right) + x \right], \tag{241}$$

which performs the conformal mapping $(x,y) \rightarrow (u,v)$ to its Fourier plane (u,v), with

$$u = -\frac{d}{2\pi} log\left(\frac{\sqrt{x^2+y^2}}{b}\right) \tag{242}$$

and

$$v = \frac{d}{2\pi} arctan(x,y) \tag{243}$$

In Eq. 241, λ is the wavelength of the incident electron beam. The parameter d is the length of the transformed beam, while b translates the transformed image in the u direction and can be chosen independently of d [136]. The function $arctan(x,y)$ refers to the arc tangent of $y = x$, taking into account which quadrant the point (x,y) is in. The transforming optical element contains a half-line of discontinuity along the negative x axis, which defines the axis around which the OAM is measured and whose end corresponds to both the origin of the coordinate system and the location of the tip (see below).

The second optical element, which is referred to as the “corrector" or “sorter 2", introduces a phase correction of the form

$$\phi_2(u,v) = -\frac{db}{\lambda f} exp\left(-2\pi\frac{u}{d}\right) cos\left(2\pi\frac{v}{d}\right). \tag{244}$$

By adding a second Fourier transforming lens of focal length F after the phase correcting element, OAM states can be separated in its focal plane. When using refractive elements instead of diffractive spatial light modulators, in order to have an efficient and compact mode transformer [267] it has been found to be

convenient to add to ϕ_1 and ϕ_2 a thin convergent spherical lens of focal length f equal to the distance between the two elements, with the transmission function

$$\psi_f(x,y) = exp\left[\frac{-i\pi(x^2+y^2)}{\lambda f}\right]. \tag{245}$$

Sorter elements based on currents

Whereas a phase shift can be introduced by a local variation in refractive index and relative optical distance in visible light optics, it is related to electromagnetic potentials in electron optics. According to the standard high energy phase object approximation [36], the electron optical phase shift $\varphi(x,y)$ can be written in the form

$$\varphi(x,y) = C_E \int_{-\infty}^{\infty} V(x,y,z)dz - \frac{2\,\pi\,e}{h}\int_{-\infty}^{\infty} A_z(x,y,z)dz \tag{246}$$

where $C_E = \frac{2\pi}{\lambda}\frac{e}{E}\frac{E_0+E}{2E_0+E}$, e is the absolute value of the electron charge, h is Planck's constant, λ is the de Broglie electron wavelength, E is the electron's energy, E_0 is the electron's rest mass energy and $A_z(x,y,z)$ is the z component of the magnetic vector potential. The z axis of the coordinate system is aligned to the optical axis of the transmission electron microscope and has the same direction as the incident electron beam.

The equations that link V and A_z to their sources can be reduced to an identical from by noticing that (in the Coulomb Gauge)

$$\nabla^2 V = -\frac{\rho}{\epsilon_0} \tag{247}$$

and

$$\nabla^2 A_z = \mu_0 j_z \tag{248}$$

where μ_0 is the vacuum permeability and the charge density ρ plays the same role as j_z, which describes the component of the current density parallel to the main propagation direction of the electron beam. According to these expressions, in-plane components of the current density do not contribute to the phase shift. They can eventually produce small, point-dependent wave translations, which can be neglected in most cases [230]. Taken together, the three equations show that, in the phase object approximation, the phase shift is proportional to the z integral of the potentials. The equation for the phase shift can then be written in the form [268]

$$\nabla_\perp^2 \varphi = -\frac{C_E\sigma}{\epsilon_0} - \frac{2\,\pi\,e}{h}\mu_0\zeta \tag{249}$$

where $\sigma = \int_{-\infty}^{+\infty}\rho dz$ and $\zeta = \int_{-\infty}^{+\infty} j_z dz$.

In this approximation, an inclined current in the plane (x, z) is equivalent to a vertical current j_z. A device that is used to control the phase using a charge distribution can therefore be reproduced by a series of vertical currents. This analogy is particularly relevant because it is often not the charge distribution but the equipotential distribution that can be controlled experimentally. Mathematically, the problem of finding appropriate boundary conditions for the potential in three dimensions can be transformed to the easier problem of finding projected sources of field in two dimensions. This difference is important when making conformal transformations of the electron wavefunction using harmonic phase elements, as suggested in the recent work of Ruffato *et al.* [240].

Here, we use this analogy to propose new alternative OAM sorter phase elements, which are based on magnetostatic phase shifts. Their design is obtained simply by substituting charges by vertical currents, or by equivalent Ampere currents for magnetic materials. More broadly, the work of Ruffato et al. [240] provides a landscape of possible conformal transformations and a generalisation of the sorter concept. The ultimate generalisation would allow sorting of any set of wavefunctions, as recently proposed for visible light optics [269]. Unfortunately, the requirement for a harmonic phase does not allow the direct application of this formalism to electron optics and will require some modifications. Nevertheless, magnetic phase elements will be an essential ingredient in this direction.

Unwrapper or sorter 1

We begin by considering a closed circuit carrying the current I between the tip of the device at $O = (0, 0, 0)$ and the points $A = (0, a, 0)$ and $B = (0, a, c)$, as shown in Figure 69.

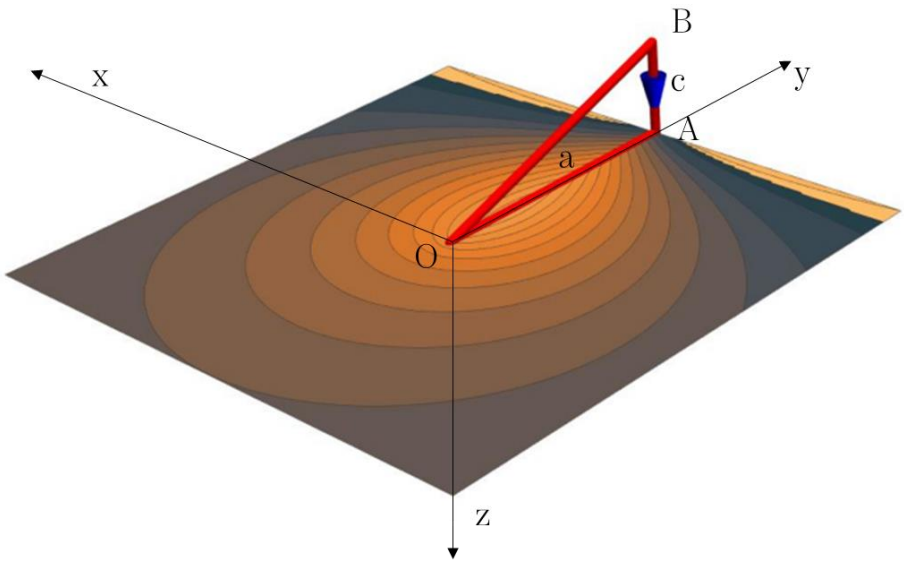

Figure 69: Current circuit and corresponding phase shift for a magnetic sorter 1 element. The current closes just behind the element. See text for details.

In order to calculate the phase shift due to a straight section of the current-carrying wire at position $r_P = (x_P, y_P, z_P)$ directed along $\mathrm{d}\boldsymbol{l} = (dx_P, dy_P, dz_P)$ and, more generally, of a closed loop, we make use of the fifth formulation of Ampere's law in terms of the vector potential, according to the expression [270]:

$$\mathrm{A_z(x,y,z)} = -\frac{\mu_0}{4\pi}\,\mathrm{I}\oint \frac{1}{\mathrm{r}}\mathrm{d}\,l \tag{250}$$

where $r = \sqrt{(\mathrm{x}-\mathrm{x_P})^2 + (\mathrm{y}-\mathrm{y_P})^2 + (\mathrm{z}-\mathrm{z_P})^2}$ and $\mathrm{d}l = (\mathrm{dx_P}, \mathrm{dy_P}, \mathrm{dz_P})$.

As the phase shift described by Eq. 246 depends only on the z component of the vector potential, horizontal components of the wire can be neglected. It is convenient to parametrize the vertical segment AB in the form $r_P = (0, a, c\,s)$, where $0 < s < 1$ and $\mathrm{d}l = (0, 0, c\,ds)$. Similarly, the oblique segment BO can be written $r_P = (0, a\,s, c\,s)$, where $dl = (0, -a, -c)ds$. The sum of the contributions of the two elements to the z component of the vector potential is therefore

$$\mathrm{dA_z(x,y,z)} = \frac{\mu_0}{4\,\pi}\,\mathrm{I\,c\,ds}\left[\frac{1}{\sqrt{(\mathrm{x})^2 + (\mathrm{y}-\mathrm{a})^2 + (\mathrm{z}-\mathrm{cs})^2}} - \frac{1}{\sqrt{(\mathrm{x})^2 + (\mathrm{y}-\mathrm{sa})^2 + (\mathrm{z}-\mathrm{cs})^2}}\right] \tag{251}$$

Making use of Eq. 246, we find for the contribution to the phase shift

$$\mathrm{d\varphi(x,y)} = \frac{e\mu_0}{2\mathrm{h}}\; I\; c\;\mathrm{ds}\{\log(x^2 + (y-a)^2) - \log(x^2 + (y-as)^2)\} \tag{252}$$

The total phase shift is then obtained by integrating the above expression in $\mathrm{d}s$ between 0 and 1, resulting in the expression

$$\varphi(\mathrm{x,y}) = \frac{ce\mu_0 I}{2\;a\;h}\;[(y\;\; log((a-y)^2 + x^2)\;+$$
$$2x\arctan\left(\frac{y-a}{x}\right) + 2a - y\log(x^2+y^2) - 2x\arctan\left(\frac{y}{x}\right)\Big] \tag{253}$$

In the limit of large a, we can make the approximations

$$\mathrm{y\log((a-y)^2 + x^2) - y\log(x^2+y^2) \simeq -y\log\left(\frac{x^2+y^2}{a^2}\right)} \tag{254}$$

and

$$\mathrm{2x\arctan\left(\frac{y-a}{x}\right) - 2x\arctan\left(\frac{y}{x}\right) \simeq -2x\;Sign[a]\;\arctan[-Sign[a]y, x]} \tag{255}$$

where $Sign[x]$ is the sign function, so that

$$\mathrm{\varphi(x,y) = \frac{c\,e\,\mu_0\,I}{2ah}\left(-2xSign[a]\;\arctan[-Sign[a]y, x] - y\log\left(\frac{x^2+y^2}{a^2}\right)\right)} \tag{256}$$

If constant and linear terms are neglected, then the functional phase shift of the ideal sorter described by Eq. 241 is recovered, once it is noted that the discontinuity is aligned along the y axis and lies on the positive or negative side depending on the sign of a. The result is identical to that obtained in the electrostatic case [143]

if the charge distribution is substituted by the current according to the analogy described above, provided that

$$\frac{d}{\lambda f} = 2C_E C_V = \frac{c\, e\, \mu_0\, I}{a\, h} \tag{257}$$

where $C_V = \frac{K}{4\pi\varepsilon_0}$, K is the constant charge density of a charged line and ε_0 is the dielectric constant of vacuum.

Corrector or sorter 2

In this case, we consider two current distributions that have the shapes of a square wave and a sinusoid, as sketched in Figure 70(a) and Figure 71, respectively.

Square wave

A square wave current is equivalent to a periodic array of vertical wires at positions $(x, y) = (0, n\, d/2)$ carrying alternatively opposite currents, as the horizontal sections of the wire do not contribute to the phase. The phase shift of a single vertical wire of length $2L$ carrying current I at the origin [230] can be obtained from the first term of Eq. 12 with $a = 0$, in the form

$$\varphi(x, y) = \frac{e\, \mu_0}{h} IL\, log[x^2 + y^2] \tag{258}$$

Although it is a simple matter of addition to construct a finite square wave, it is more interesting to consider an infinite periodic arrangement of vertical wires at positions $(0, n\, d/2)$, which can be described by the expression

$$\sum (-1)^n \frac{e\, \mu_0}{h}\, I\, L \log\left[\left(x - n\frac{d}{2}\right)^2 + y^2\right] \tag{259}$$

As shown elsewhere, e.g., in [271], this sum can be expressed analytically as

$$\varphi(\mathrm{x}, \mathrm{y}) = \frac{e\mu_0}{2\mathrm{h}} I\, L \log\left(\frac{\cosh\left(\frac{2\pi y}{d}\right) - \cos\left(\frac{2\pi x}{d}\right)}{\cosh\left(\frac{2\pi y}{d}\right) + \cos\left(\frac{2\pi x}{d}\right)}\right) \tag{260}$$

For $y \gg d$, the leading term in this Fourier series reads

$$\varphi(\mathrm{x}, \mathrm{y}) = \frac{2e\mu_0}{h}\, \mathrm{I\, L} \exp\left(-\frac{2\pi|y|}{d}\right) cos\left(\frac{2\pi x}{d}\right) \tag{261}$$

Apart from a multiplying constant, this expression is coincident with the required functional form for sorter 2 described by Eq. 244. A cosine phase map of such an array is shown in Figure 70(b).

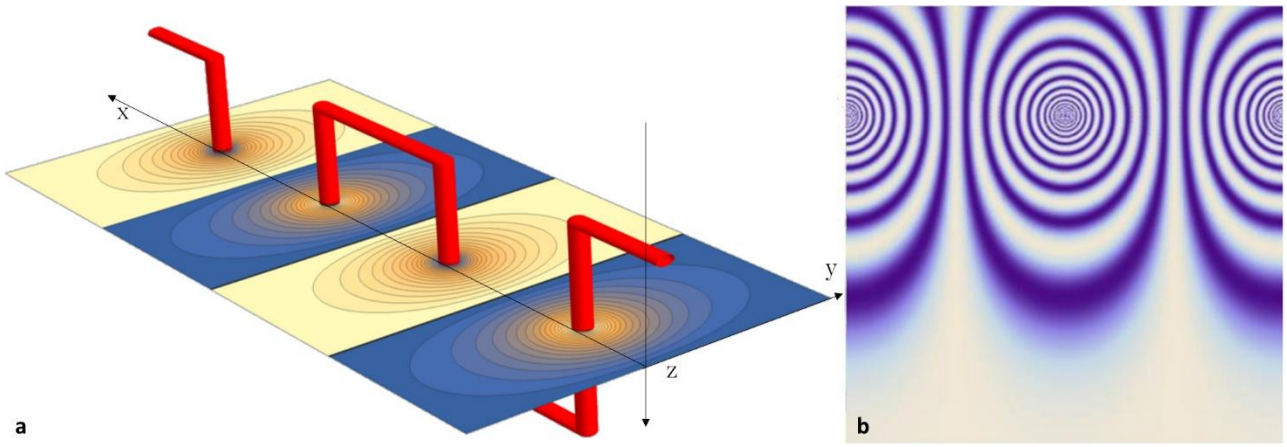

Figure 70: Square wave current distribution suitable for the sorter 2 element. a) Three-dimensional structure of the current, with a plane cutting it at half of the total amplitude and showing a contour map of the phase shift introduced by the current distribution. Blue and yellow correspond to opposite signs of the phase shift. b) Cosine phase map calculated for a periodic distribution of vertical currents of opposite sign. See text for details.

Sinusoid

We describe a sinusoidal current by the parametric equation

$$\mathrm{r_P} = \left(x, 0, L\, sin\left(\frac{2\pi x}{d}\right)\right) \tag{262}$$

with

$$\mathrm{d}l = \left(1{,}0, L\frac{2\pi}{c}\, cos\left(\frac{2\pi x}{d}\right)\right)\mathrm{d}x \tag{263}$$

Even if the current is continuous, its z component varies as a cosine and it is equivalent to a cosine distribution of charges, resulting in the ability to generate the sorter phase explicitly without the need for approximations. The phase shift is then given by the expression

$$\varphi(x,y) = \frac{2e\mu_0}{h}\, I\, L\, exp\left(-\frac{2\pi|y|}{d}\right) cos\left(\frac{2\pi x}{d}\right) \tag{264}$$

The current and corresponding phase shift are shown in Figure 71.

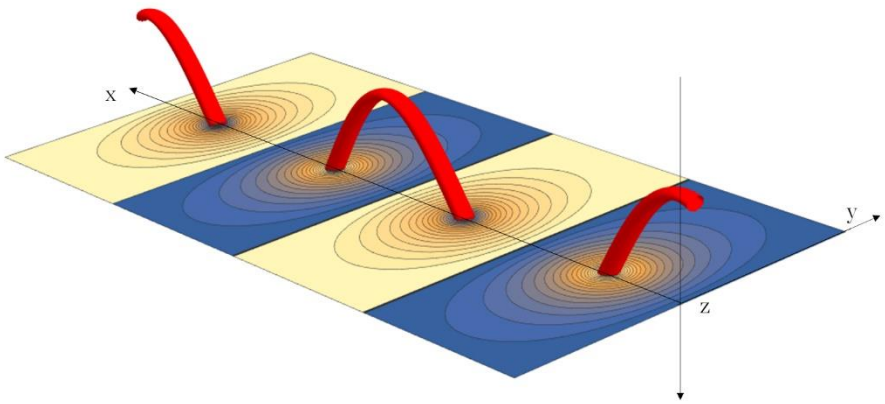

Figure 71: Sinusoidal current distribution suitable for the sorter 2 element. The three-dimensional structure of the current is illustrated, with a plane cutting it at half of the total amplitude and showing a contour map of the phase shift introduced by the current distribution. Blue and yellow correspond to opposite signs of the phase shift. See text for details.

Sorter elements based on magnetic materials

Unwrapper or sorter 1

The magnetic effect of a current n the phase shift can be obtained equivalently from a magnetized element. The analogy is obtained by shaping the edge of the magnetized element based on the current distribution. We begin by considering a magnetic lamina of small thickness t with vertices at the points $O = (0, 0, 0)$, $A = (0, a, 0)$ and $B = (0, a, c)$ and with a magnetization vector $\boldsymbol{M}$ in the x direction, i.e., $\boldsymbol{M} = (M;\ 0;\ 0)$, as shown in Figure 72. The magnetic vector potential of an elementary magnetic dipole with moment $\boldsymbol{m} = (m, 0, 0) = (M\ t\ dy\ dz, 0, 0)$, located at the origin of the coordinate system, is given by the expression [270]

$$\mathrm{A_z} = \left\{0, -\frac{\mu_0\ \mathrm{m\ z}}{4\pi(\mathrm{x^2 + y^2 + z^2})^{3/2}}, \frac{\mu_0\ \mathrm{m\ y}}{4\pi(\mathrm{x^2 + y^2 + z^2})^{3/2}}\right\} \tag{265}$$

where μ_0 is the vacuum permeability.

In the phase object approximation described by Eq. 246, the corresponding electron optical phase shift takes the form

$$\varphi(x, y) = -\frac{e\ \mu_0\ m\ y}{h(x^2 + y^2)} \tag{266}$$

As the phase shift does not depend on z, the contribution to the phase shift at the point (x, y) due to a column of dipoles at $(0, y_0, 0)$ of height $y_0 = c/a$ (Figure 72) is

$$\mathrm{d}\varphi(x, y) = -\frac{c\ e\ \mu_0\ M\ t\ y_0(y - y_0)}{a\ h\ (x^2 + (y - y_0)^2)}\mathrm{d}y_0 \tag{267}$$

By integrating the above equation from 0 to a, the phase shift associated with the triangular magnetic lamina turns out to be

$$\varphi(\mathrm{x, y}) = \frac{c\ e\ \mu_0 Mt}{2\ a\ h}[y\ log((a - y)^2 + x^2) + 2x\ arctan\left(\frac{y - a}{x}\right) + 2a - y\ log(x^2 + y^2) - 2x\ arctan\left(\frac{y}{x}\right)] \tag{268}$$

By making use of the same approximations as in Eq. 254 and Eq. 255, we obtain the expression

$$\varphi(x, y) = \frac{ce\mu_0 Mt}{2ah}\left(-2x\mathrm{Sign}[a]\arctan[-\mathrm{Sign}[a]y, x] - y\log\left(\frac{x^2 + y^2}{a^2}\right)\right) \tag{269}$$

Therefore, by considering the correspondence between magnetization and Amperian currents, the magnetic device is shown to be equivalent to a current of $I = -M\ t$ circulating on the edges of the lamella.

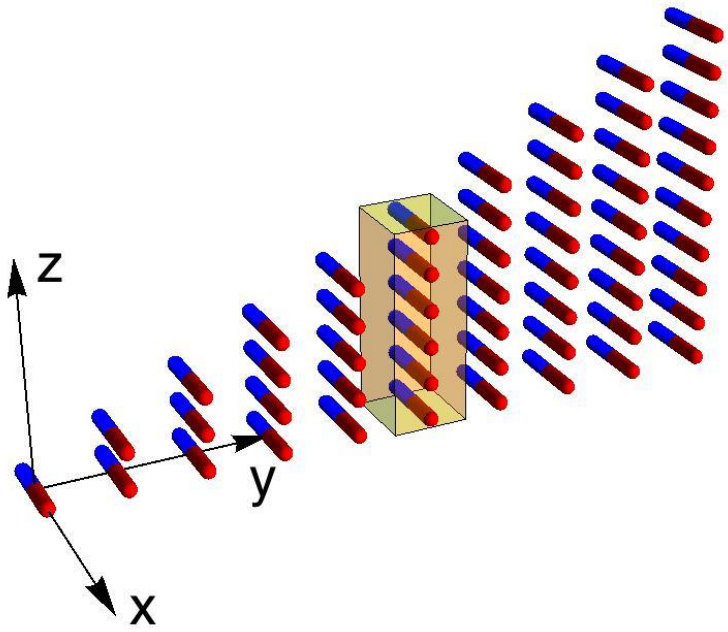

Figure 72: Discrete representation of a continuous triangular distribution of elementary magnetic dipoles. The shaded prism represents a volume of integration along z.

Corrector or sorter 2

As a result of the projection feature of the phase object approximation, any periodic structure is able to give, in vacuum, the phase shift required by the corrector, or sorter 2, element of the ideal sorter. We consider, as an illustrative example, a periodic array of alternating magnetic stripe domains of width w, aligned along y, in a thin specimen lying in the half-plane $x < 0$. The phase shift introduced by such an array (Figure 73) has been calculated in analytical form previously [272] and is given by the expression

$$\varphi(\mathrm{x},\mathrm{y}) = \frac{N}{2\,\pi}\,\Re\left((2\,e^{\frac{i\,\pi\,y}{w}}\,H(-x)\,\Phi\left(e^{\frac{2\,i\,\pi\,y}{w}},2,\frac{1}{2}\right) - \right.$$

$$\left.\mathrm{Sign}(-\mathrm{x})\,\mathrm{e}^{-\frac{\pi(|-x|-iy)}{w}}\,\Phi\left(e^{\frac{2\,\pi(|-x|-\,i\,y)}{w}},2,\frac{1}{2}\right)\right) \qquad (270)$$

where $N = \frac{2e\mu_0 Mwt}{h}$ is the number of flux quanta trapped in the domain, $H(x)$ is the Heaviside step function and $\phi(z,s,a)$ is the Lerch transcendent function [273]. The phase shift of a basic period of this structure is shown in Figure 73(a) for $N = 20$, illustrating the triangular shape of the phase in the specimen corresponding to the case of zero-width magnetic domain walls. The slow decrease in phase in the vacuum region is visualized more clearly in the form of a contour map in Figure 73(b), from which the internal phase contribution has been removed.

The leading Fourier term in the former expression

$$\varphi(x,y) = \frac{4\,N}{\pi}\,exp\left(-\pi\frac{x}{w}\right)cos\left(\pi\frac{y}{w}\right) \qquad (271)$$

has the analytical form required for sorter 2, as described in Eq. 244.

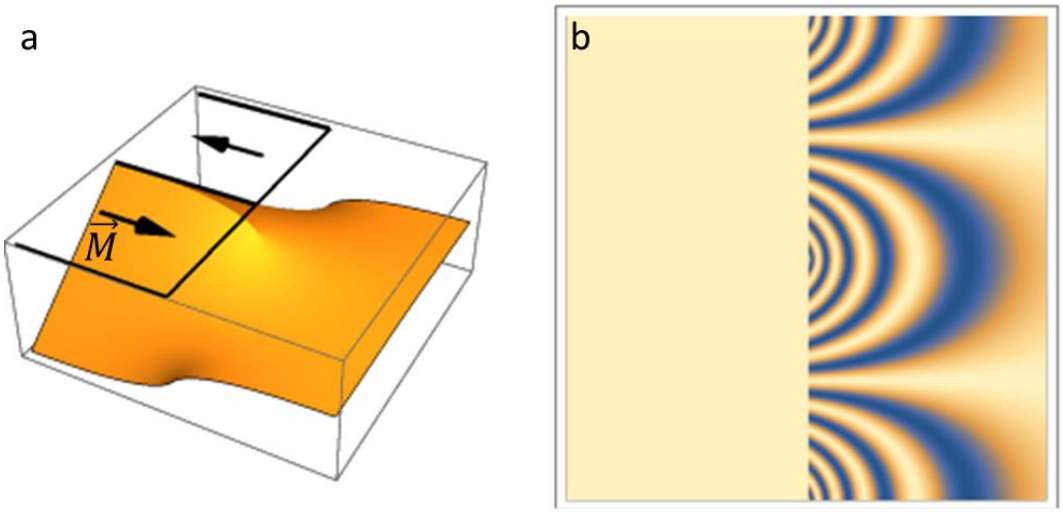

Figure 73: (a) Three-dimensional representation of the phase shift for a basic period of an infinite array of alternately-oriented stripe domains aligned along the y direction in the half-plane $x < 0$. (b) Corresponding phase contour map in the vacuum region for $x > 0$.

Design of a realistic sorter

In our former paper on an electrostatic sorter [143], we recognized that the use of a line charge of finite length for sorter 1, combined with the need to have a mirror line for neutrality purposes, was responsible for an astigmatic contribution to the phase shift, which could wash out the sorting effect completely. We were able to compensate for this astigmatic contribution by adding two additional line charges (and their images) perpendicular to the line charge, thereby restoring the performance of the device. As magnetic elements do not, in general, require compensating mirror images, a simpler structure comprising only three elements (sorter 1 and two perpendicularly aligned elements) was investigated here. The results were disappointing, indicating that the mirror charges were also required for compensation.

In order to better appreciate the influence of astigmatism and its correction, we first consider a single magnetic element of length $a = 80\mu\mathrm{m}$ (with the tip at the origin of the coordinate system), acting as a sorter 1 in a device in which a length of the transformed beam of $d = 10\mu\mathrm{m}$ over a focal length $f = 0.5\mathrm{m}$ is required. These values are similar to the conditions used for our last experiments performed for an electrostatic sorter [139]. We are interested in a field of view of $30\mu\mathrm{m} \times 30\mu\mathrm{m}$ centred on the tip.

As an illuminating beam, we consider a petal beam [23], which is defined by the equations

$$\psi_{ill} = \mathcal{N}\{\exp[\mathrm{i}l\arctan(\mathrm{x},\mathrm{y})] + \exp[-\mathrm{i}l\arctan(\mathrm{x},\mathrm{y})]\} \quad \text{for} \quad \mathrm{x}^2 + \mathrm{y}^2 < R^2 \tag{272}$$

$$\psi_{ill} = 0 \quad \text{for} \quad \mathrm{x}^2 + \mathrm{y}^2 > R^2 \tag{273}$$

where $\mathcal{N}$ is a normalization constant.

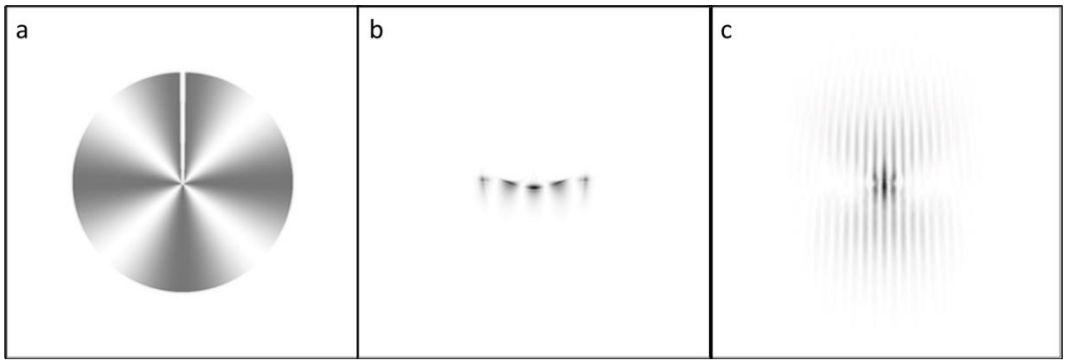

Figure 74: (a) Petal beam with an OAM $|\ell| = 2$ intensity distribution. (b) Beam intensity distribution before the sorter 2 element and after it has interacted with the sorter 1 element. (c) Blurred OAM spectrum obtained without astigmatism compensation/correction.

Figure 74(a) shows the intensity distribution of the beam, with $R = 10\mu\mathrm{m}$ and $\ell = 2$, in the sorter 1 plane, over a field of view of $30\mu\mathrm{m} \times 30\mu\mathrm{m}$. An amplitude function introduced to take into account the finite dimension of the device is also shown. The intensity before the sorter 2 plane is shown in Figure 74 (b). Four intensity minima are visible, corresponding to $\ell = \pm 2$. However, there is also a distortion caused by the finite dimensions of sorter 1. A Fraunhofer image calculated for a focal length $F = 2\mathrm{m}$ in Figure 74 (c) shows that the astigmatism blurs the sorting effect completely.

In order to optimize the device without resorting to a replica of the electrostatic device, we introduced two additional magnetic elements at a distance of $40\mu\mathrm{m}$, with their tips aligned with sorter 1. We then investigated the effect of rotating them with respect to sorter 1, as shown in Figure 75. For two triangular elements that have lengths of $80\ \mu\mathrm{m}$, compensation is achieved at an angle of $10°$, as shown in Figure 76(a). The focal length of the Fraunhofer lens has been increased to $F = 5\ \mathrm{m}$, in order to emphasize the variations and features of the intensity distribution in the diffraction image. Deterioration of the image at 0 and $20°$ is shown in Figure 76 (b) and Figure 76 (c), respectively. We also tested the effect of placing the return vertical current at a greater distance and of having a trapezoidal instead of a triangular current circuit, but this modification was not as beneficial as the angular compensation.

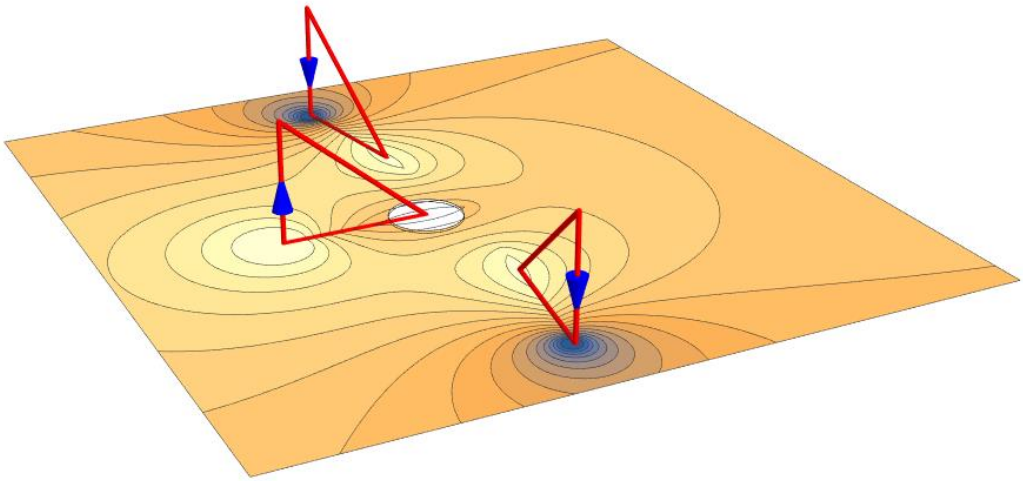

Figure 75: Phase map calculated for sorter 1 with astigmatism correction elements. The currents of the sources are indicated schematically. The typical beam position is indicated by a circle.

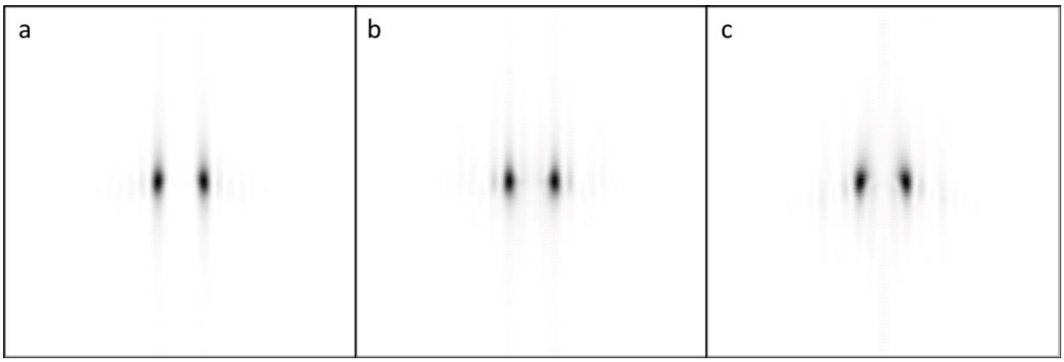

Figure 76: OAM spectrum obtained using a magnetic sorter with a length of $80\ \mu m$ *and with a lateral astigmatism corrector oriented at an angle of: (a) 10° (corresponding to Figure 75), (b) 0° and (c) 20°.*

We compared the results of Eq. 256 numerically with our previous experimental results for electrostatic sorter 1, where the average tilt of the beam was $d/f\ =\ 20\ \mu\mathrm{rad}$ [139]. A similar result could be obtained for a current of $67\ mA$ for the case $c = \frac{a}{2} = 40\mu\mathrm{m}$. Such a current requires detailed engineering, but is compatible with MEMS technology. For sorter 2 fabricated in the form of either a square wave or a sinusoidal current distribution, the minimum current required for correction, for $z\ =\ 0$ in Eq.261, is again $67\ mA$ for $b = a = 80\mu\mathrm{m}$ and $L = 20\mu\mathrm{m}$.

It should be noted that for sorter 2 this is the minimum value required for correction. However, precise tuning is not required, since the position of the beam can be used to compensate for the exact phase, with the values of the parameters increased according to the exponential factor in Eq.261. For the realisation of sorter elements using magnetic materials, strong magnetic anisotropy is needed so that a slab of triangular or rectangular shape remains magnetized in the out-of-plane direction (i.e., in the direction perpendicular to the beam propagation axis). For FePt, assuming a saturation magnetisation of $M\ =\ 10^6 A/\mathrm{m}$, a current of $I\ =\ 67\ mA$ corresponds to $t = 67\ n\mathrm{m}$ for sorter 1. For the stripe domain configuration of sorter 2 (Eq.271) and the same value of magnetisation M, minimum correction is achieved for a specimen thickness of $0.2\ \mu\mathrm{m}$ and a domain width $w = {}^{d}/_{2}\ =\ 5\mu\mathrm{m}$. These values correspond to the geometrical configuration considered above and should be changed for different geometrical parameters.

Conclusions

We have described an analogy between magnetic and electrostatic phase elements, which has general importance for fabricating customized phase elements, but is particularly relevant for the present case of an orbital angular momentum sorter. We have used analytical models to derive the shape of the current distribution that is required to achieve a completely magnetic version of the sorter. For the first sorting element, a new design with no direct analogue to the electrostatic case has been introduced to control the astigmatism that is introduced by the device. Numerical calculations indicate that phase shifts analogous to the electrostatic case can be obtained with currents that are within experimentally achievable ranges. The proposed devices have the important advantage that the field sources can be controlled directly, in contrast

to the electrostatic case, for which the effective charge could only be controlled through the appropriate shaping of electrodes and by neglecting mutual induction of the elements. The new approach promises to provide a more accurate way to produce near-ideal OAM sorting over a large range of beam sizes.

Chapter 8

"Focused Ion Beam fabrication of Janus bimetallic cylinders acting as drift tube Zernike phase plates for electron microscopy"

Paolo Rosi[1], Gian Carlo Gazzadi[2], Stefano Frabboni[1, 2], Vincenzo Grillo[2], Amir H. Tavabi[3], Rafal E. Dunin-Borkowski[3], and Giulio Pozzi[3, 4]

1. Dipartimento FIM, Universitá di Modena e Reggio Emilia, 41125 Modena, Italy
2. Centro S3, Istituto di nanoscienze-CNR, 41125 Modena, Italy
3. Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany
4. Department of Physics and Astronomy, University of Bologna, 40127 Bologna, Italy

Abstract

Modern nanotechnology techniques offer new opportunities for fabricating structures and devices at the micron and sub-micron level. Here, we use focused ion beam techniques to realize micron-sized Janus bimetallic cylinders acting as drift tube devices, which are able to impart a controlled phase shift to an electron wave. The phase shift results from the presence of contact potentials in the cylinders, in a similar manner to the electrostatic Aharonov-Bohm effect in bimetallic wires. We use electron Fraunhofer interference to demonstrate that such bimetallic structures introduce phase shifts that can be tuned to desired values by varying the dimensions of the pillars, in particular their heights. Such devices are promising for electron beam shaping and for the realization of electrostatic Zernike phase plates (i.e., devices that are able to impart a constant phase shift between an unscattered and a scattered electron wave) in electron microscopy, in particular cryo-electron microscopy.

Forewords

In this work I present our early findings on a novel method to modify the phase of an incoming electron beam that we envision might be used in the future to build phase plates for beam shaping or phase contrast microscopy. In this work I prepared the samples via FIBID and FIB milling using the dual-beam machine of the CNR-Nano in Modena and analysed them at the TEM of the CIGS facility in Modena. I also analysed all the experimental data and wrote most of the manuscript.

Introduction

Developments in state-of-the-art electron microscopes (including aberration correctors, field emission guns and single particle detectors), in the fabrication of sub-micron devices (using focused ion beam instruments, electron beam and optical lithography, etc.) and in biological specimen preparation (e.g., the use of frozen-hydrated specimens in cryo-electron microscopy) have stimulated renewed interest in using phase plates as

devices that can be used to apply phase shifts to scattered electrons with respect to unscattered beams, in order to improve contrast in images of weak phase objects (including un-stained biological molecules) more efficiently than using standard methods, such as defocusing the objective lens.

Phase plates are of paramount importance in cryo-electron microscopy for increasing the contrast of unstained biological molecules. Beyond electron microscopy, phase control of charged particles is relevant in the general field of structured matter waves, in particular for protons and light ions in the sub-MeV regime [123,274].

A survey of the history of phase plates has been presented by Nagayama [275], while the state of the art in the field has been reported by Glaeser [276], who summarised the strengths of each device, highlighted remaining problems and presented future perspectives. In his analysis, he listed electrostatic drift tubes, i.e., electrically biased drift tubes surrounded by grounded guard electrodes [277], as a promising approach. One of the problems with fabricating such devices is the need to connect and control the electrical bias to each nanoscale electrode, with limited space available for cabling. Here, we show how similar devices can be realized by making use of contact potential differences between metals to generate potential differences. This concept loses the advantage of tuneability, but is simpler in realization and offers prospects for further reductions in device dimensions. A desired constant phase shift can be chosen by selecting the dimensions of the device, in particular its height. A similar use of contact potentials has been reported by two groups [278,279] for a two-metal Einzel lens configuration (Boersch phase plate) [276], which has a different configuration from the present “drift tube" concept.

Basic theoretical considerations

The working principle of the proposed device can be understood within the framework of the Aharonov-Bohm effect [280–282] in its electrostatic version [283,284] as demonstrated previously for a bimetallic wire [41,285–287]. It is straightforward to demonstrate the equivalence of the electron optical phase shifts produced by magnetic and electrostatic dipoles that are rotated with respect to each other by 90° [36]. For a toroidal instead of a linear geometry, a circular distribution of elementary magnetic dipoles, which is shown in discrete form in Figure 77(a), is equivalent to the circular distribution of electrostatic dipoles shown in Figure 77(b). Each arrangement produces a constant phase shift between an electron beam passing inside and outside the toroid. On a macroscopic level, this equivalence is retained between the toroidal magnet shown in Figure 77(c) and a cylindrical charge distribution with a radial electric dipole moment, such as the drift tube shown in Figure 77(d).

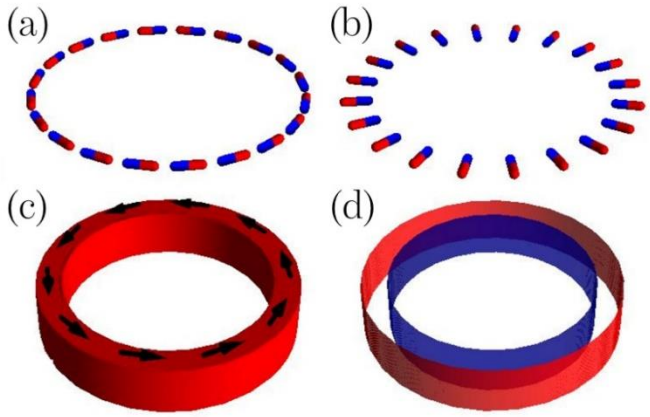

Figure 77: Analogies between arrangements of magnetic (left) and electric (right) dipoles. (a) Toroidal magnetic flux tube and (b) toroidal ring of electrostatic dipoles, represented by discrete sets of dipoles. (c) Toroidal magnet and (d) drift tube formed from coaxial cylinders of opposite charge.

In the present case, instead of using a dipole-based model we consider cylindrically symmetrical charges, whose simplest electrostatic field distribution is represented by a uniformly charged ring, for which the problem of finding the electrostatic field and potential in all space has been solved [288–290]. In cylindrical coordinates (ρ, θ, z), the potential of a ring of radius R carrying total charge Q takes the form

$$V(\rho, z) = \frac{1}{4\pi\epsilon_0} \frac{Q}{\sqrt{(\rho + R)^2 + z^2}} \frac{2}{\pi} K\left[\frac{4\rho R}{(\rho + R)^2 + z^2}\right] \tag{274}$$

where

$$K(m) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m\,sin^2\,\theta}} \tag{275}$$

is the complete elliptic integral of the first kind defined for $0 < m < 1$ [291]. Just as for isolated charges [122], the electron optical phase shift of a charged ring can only be calculated if another charged ring of opposite charge is present, as the convergence of the integral is then ensured by the charge neutrality. In the phase object approximation, the phase shift is proportional to the projected electrostatic potential. The z coordinate is therefore not significant, but only the different radii matter. Unfortunately, the function K makes it difficult to obtain an analytical result for the phase shift by direct evaluation of the integral.

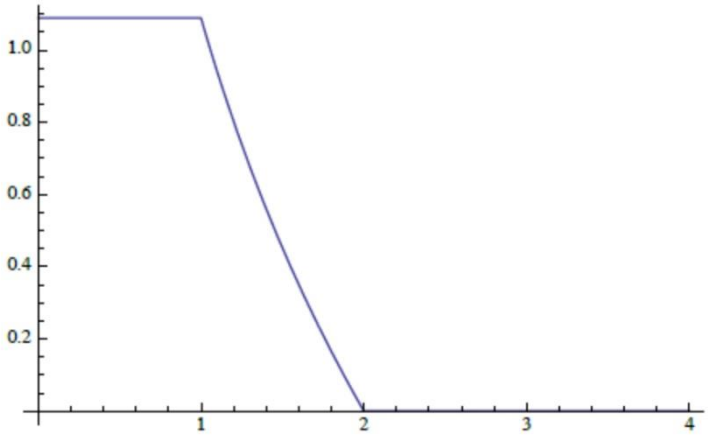

Figure 78: Trend of the electron optical phase shift of two opposite line charges at radii 1 and 2, shown in arbitrary units.

Instead, the numerical evaluation of the coordinate dependent part can be carried out using Mathematica [273], as shown in Figure 78.

This numerical result suggests a way to find an analytical counterpart, based on the idea that the two-dimensional projected potential (and hence the electron optical phase shift) satisfies Poisson's equation [292–294]

$$\nabla^2\varphi(x,y) = -C_E\frac{1}{\epsilon_0}\sigma_p(x,y) \tag{276}$$

where C_E is an interaction constant that takes a value of $7.3\frac{rad}{V\mu\mathrm{m}}$ for 200 keV electrons and $10.1\frac{rad}{V\mu\mathrm{m}}$ for 80 keV electrons.

Furthermore, $\rho(x,y,z)$ is the charge density and $\sigma_\rho(x,y)$ is the projected charge density, given by the expression

$$\sigma_p(x,y) = \int_{-\infty}^{\infty}\rho(x,y,z)dz \tag{277}$$

By applying the Gauss theorem, we can infer that the phase inside a charged ring is constant, whereas outside the ring it is equal to that obtained if the charge were concentrated at the centre of the ring, in the form

$$\varphi(r) = -C_E\frac{1}{4\pi\epsilon_0}Q\,log\left(\frac{x^2+y^2}{R^2}\right) \tag{278}$$

According to this equation, the phase inside the ring is constant and equal to 0.

For two oppositely charged rings with radii R_1 and $R_2 > R_1$, with the positive charge on the inner ring R_1, the phase shift is given by the expression

$$\varphi(r) = 0 \quad \text{for} \quad r > R_2 \tag{279}$$

$$\varphi(r) = -C_E\frac{1}{4\pi\epsilon_0}Q\,log\left(\frac{r^2}{R_2^2}\right) \quad \text{for} \quad R_1 < r < R_2 \tag{280}$$

$$\varphi(r) = -C_E\frac{1}{4\pi\epsilon_0}Q\,log\left(\frac{R_1^2}{R_2^2}\right) \quad \text{for} \quad 0 < r < R_1 \tag{281}$$

These equations are consistent with the numerical results reported in Figure 78. As a result of the projection of the charge described by Eq.277, this result holds not only for a charged ring but also for the cylindrical charge distribution shown in Figure 77(d). If the region between the charged rings is shielded by an opaque aperture, then this setup acts as a Zernike phase plate, with a phase difference that can be varied by changing the charge on the coaxial cylinders.

Specimen preparation

In order to test this proposal, we fabricated Janus bimetallic cylinders and measured their phase shifts with respect to nearby holes of nearly equal diameter. We recorded Fraunhofer images of each pair of holes, from

which the phase shift of the drift tube with respect to the nearby hole could be measured in the form of a lateral displacement of the interference fringes with respect to the diffraction envelope [30]. This effect was enhanced by optimizing the radii of the holes and their separations [30,36]. An estimate of the phase difference can be obtained by assuming that the potential inside a cylinder of height h is constant and equal to the contact potential difference V_C between the two dissimilar metals, in our case Pt and Al, which is $V_C = 1\mathrm{V}$ [295,296]. The corresponding phase difference is given by the expression [28,36,297]

$$\phi = \mathrm{C_E V_C h} \tag{282}$$

The expected phase shift is 2π for a tube length $h{\sim}0.86\mu\mathrm{m}$ for 200 keV electrons.

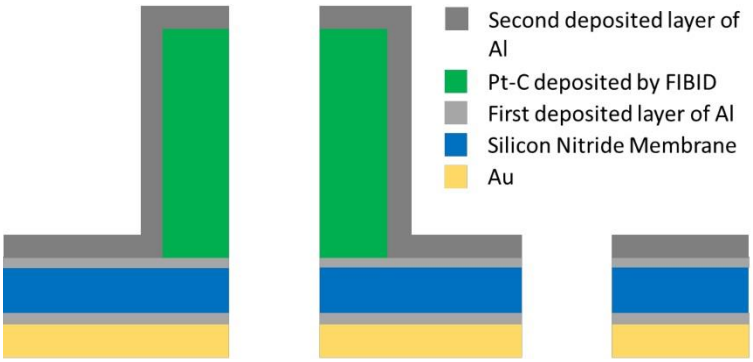

Figure 79: Schematic diagram of the cross-sectional structures of the cylinders and adjacent holes.

As the contact potential is affected greatly by the state of the surfaces and the purity of the materials, we chose to make a series of Janus bimetallic cylinders with nominal heights of 0, 0.5, 0.75, 2.5 and 4 μm, the first of which served as reference.

We used focused ion beam (FIB) techniques in an FEI strata DB235M equipped with a Ga+ liquid metal ion source (LMIS) operated at 30 keV to realize toroidal versions of bimetallic wires. The starting point was a commercial PELCO® ultra-low-stress silicon nitride support film with nine $100 \times 100\ \mu\mathrm{m}$ windows and a membrane thickness of 200 nm. The frame supporting the membranes was a standard 3-mm-diameter round silicon frame of thickness $200\mu\mathrm{m}$. We evaporated ~ 50 nm of Al onto each side of the membrane. Pairs of circular holes, with a pillar on one hole in each pair, were then fabricated in four steps to realize the cross-sectional structures shown in Figure 79. First, Pt-C cylinders were deposited by FIB-induced deposition (FIBID) using a Pt-organic precursor gas (trimethylmethylcyclopentadienylplatinum (IV), $(CH_3)3CH_3C_5H_4Pt$, vaporized by heating to 44 °C). The ion beam current used during deposition was ~ 50 pA. For each cylinder, the ion beam was scanned during gas injection over a ring-shaped area with an inner radius of $1.15\ \mu\mathrm{m}$ and an outer radius of $1.5\ \mu\mathrm{m}$. Five pillars of varying height were deposited. The total scan time was ~ 1 minute for the shortest pillar and ~ 9 minutes for the tallest pillar (Figure 80.e). A distance of at least $35\ \mu\mathrm{m}$ was maintained between pillars and no more than three pillars were deposited on each window. In the second step, a bimetallic structure was created by depositing ~ 100 nm of Al over the entire sample using thermal evaporation. In the final step, a double aperture was formed by creating pairs of holes with radii of $1.2\ \mu\mathrm{m}$

using FIB milling, one through each tube and one adjacent to it. The distances between the centers of the adjacent holes was approximately twice the diameter of each hole. Here, the ion beam current was increased to ~ 100 pA. It was found to be necessary to deposit a further ~ 150 nm of Au onto the "lower" side of the membrane using sputter coating, in order to make the support completely opaque to electrons at 200 kV.

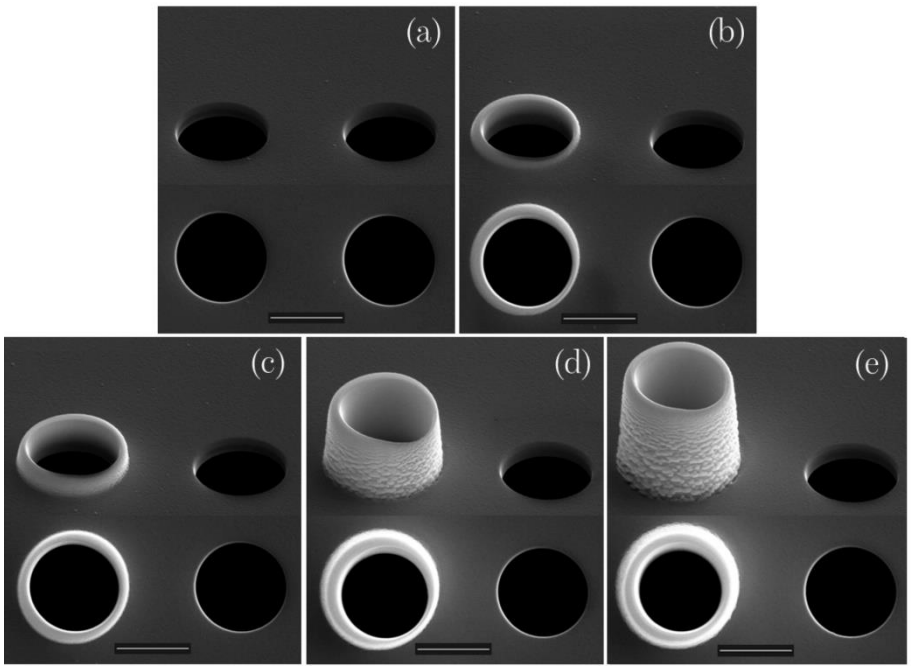

Figure 80: SEM tilted and top views of the double circular aperture structures. One aperture in each pair takes the form of a FIBID-fabricated Al-covered Pt-C tube. Several representative values of tube height are shown. The scale bar in each image is $2\mu m$*. The measured pillar heights are (a) 0 (i.e., no pillar structure), (c)* $(0.4 \pm 0.1)\ \mu m$*, (c)* $(0.7 \pm 0.1)\ \mu m$*, (d)* $(2.2 \pm 0.1)\ \mu m$ *and (e)* $(3.3 \pm 0.1)\ \mu m$*.*

Figure 80 shows scanning electron microscopy (SEM) images recorded in tilted and top view configurations. A reference structure of two holes without a fabricated pillar structure is shown in Figure 80(a). The tilted views shown in Figure 80(b-d) allow the pillar heights to be measured. The nominal values of the heights for the pillars shown are 0.5, 0.75, 2.5 and 4μm, while the measured heights from the SEM tilted views are $(0.4 \pm 0.1)\ \mu$m, $(0.7 \pm 0.1)\ \mu$m, $(2.2 \pm 0.1)\ \mu$m and $(3.3 \pm 0.1)\ \mu$m, respectively.

Experimental results

Experiments were carried out at 80 and 200 keV on a Talos F200S G2 TEM in low magnification, low angle diffraction mode (camera length 1.4 km). Results obtained from the structures shown in Figure 80 are reported in Figure 81. Values of the phase difference were measured by performing first to the data based on the formula

$$I(x) = b\frac{J_1^2(xD)}{(xD)^2}[1 + \mu cos(xd + \Delta\varphi)] \tag{283}$$

which was derived by considering the Fraunhofer images of two circular apertures [30]. The multiplicative factor is the Fourier transform of a circular aperture (i.e., an Airy disk), while the interference term in square brackets includes a damping factor μ, which accounts for partial coherence effects resulting from the finite

dimensions of the electron source (i.e., spatial coherence) [30], J_1 is the Bessel function of the first kind, b is a fitting parameter that allows a non-normalized intensity profile to be fitted, $\Delta\varphi$ is the phase difference between the apertures and D and d are coefficients that are proportional to the aperture diameter and separation, respectively. As a result of the high coherence of the Schottky field emission gun, the influence of spatial coherence was found to be less than 10% of the overall intensity. The values of phase difference between the two apertures, which were obtained modulo 2π from the fitting procedure, are reported in Figure 81 on the right of each line scan plot. The Fraunhofer images and corresponding line scans show that pattern changes with pillar height due to changes in the phase difference between the holes. For the double aperture with the smallest cylinder (Figure 81b), the phase difference is almost π at both accelerating voltages. A similar structure in the intensity profile can be observed for the double aperture with a pillar height of $2.2\mu\mathrm{m}$ (Figure 81d) at 80 kV, but not at 200 kV, which instead shows a marked asymmetry, suggesting that the phase difference between the two holes is neither 0 not 2π.

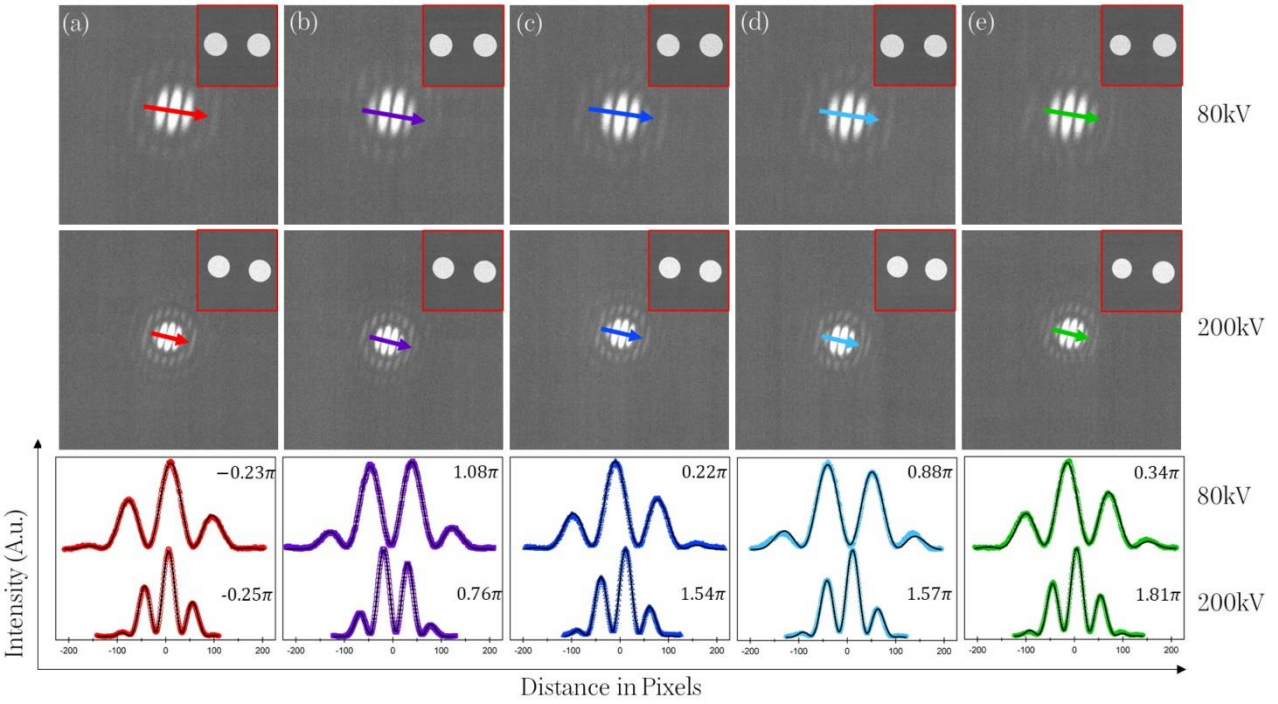

Figure 81: TEM images and diffraction patterns of the double apertures shown in Figure 80: (a) reference, (b) 0.4 μm, (c) 0.7 μm, (d) 2.2 μm, (e) 3.3 μm, The first row shows data collected at 80 keV, while the second row shows data collected at 200 kV. Line scans of the diffraction patterns are shown in the lower two rows, with markers representing experimental data points and solid black lines fits. The direction of the line scan is marked by an arrow in each diffraction pattern. The phase differences (modulo 2π) obtained from fits to Eq. 283 are indicated in the upper right corner of each line scan.

The shape of each line scan plot shows that at both accelerating voltages the reference (with no pillar on either of the holes) is slightly asymmetrical. Fits of the data to Eq.283 indicate that at both accelerating voltages the phase difference is almost $\frac{\pi}{4}$. This unexpected effect, in the absence of a pillar, probably results from the metal-insulating structure of the holes, as well as some asymmetry between them resulting from the fabrication process. Similar behaviour was recently observed in another paper with circularly symmetrical

structures built on a Si_3N_4 supporting layer [187]. Metal deposition reduced charging effects but did not completely remove it.

We plotted the total phase difference as a function of cylinder height. The total phase difference was calculated by adding an integer multiple of 2π to each phase difference value obtained from the fitting procedure, so that for each height $\Delta\varphi_{tot} = \Delta\varphi_{fit} + 2n\pi$. The value of n for each pillar height was chosen to be the rounded-off integer result of the division $\frac{h_{pillar}}{h_{2\pi}}$.We estimated an initial plausible value of $h_{2\pi}$ for each accelerating voltage from the trend of the fitted phase difference reported in Figure 81. In Figure 81(c), the phase difference at 80 keV should have already surpassed 2π since that calculated from the fit is 0.22π, while it is close to being reached at 200 keV. The two first guesses were therefore $h_{2\pi\,80\,\mathrm{keV}} = 0.6\mu\mathrm{m}$ and $h_{2\pi\,200\,\mathrm{keV}} = 0.8\mu\mathrm{m}$. We then fitted the linear function $\Delta\varphi_{tot} = \beta h_{pillar}$ to the data and tried other values of $h_{2\pi}$to minimize χ^2 (evaluated using Pearson's χ^2test). The physical meaning of β can be understood when we compare the linear fitting function to Eq.282, so that $\beta = C_E V_{C-exp}$. The results are shown in Figure 82. However, it is important to note that the result is not unequivocal, since we have a small number of data points.

The angular coefficient for the linear fit with the smallest value of χ^2 is $\beta_{200\,kV} = (2.41 \pm 0.06)\frac{\pi}{\mu\mathrm{m}}$ for the 200 kV series, while $\beta_{80\,keV} = (3.13 \pm 0.03)\frac{\pi}{\mu\mathrm{m}}$ at 80 kV. The ratio between the two angular coefficients is $r_\beta = \frac{\beta_{80\,keV}}{\beta_{200\,kV}} \approx 1.3$, which is consistent with the ratio between the interaction constants $r_{C_E} = 1.3836$.

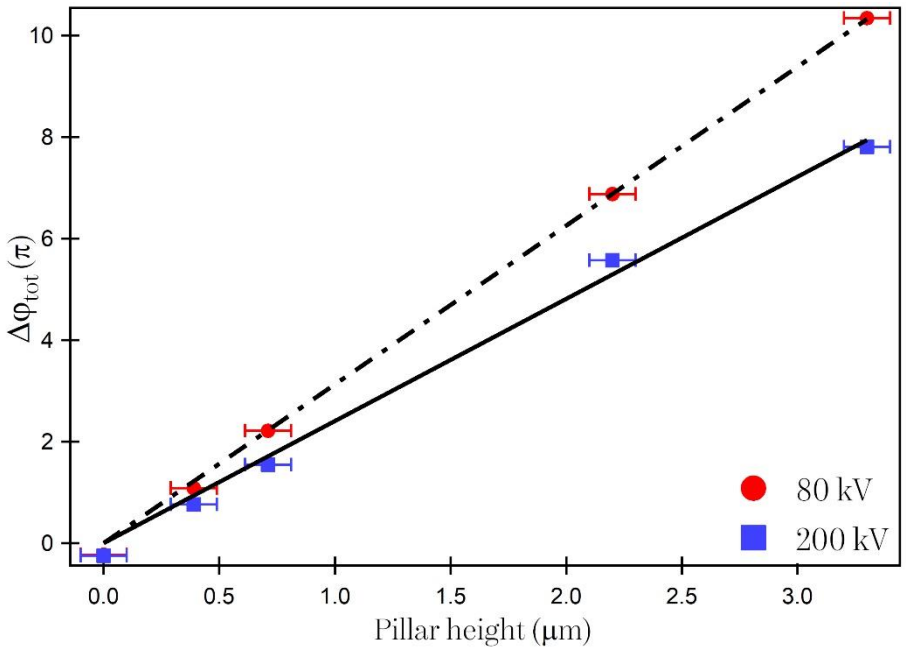

Figure 82: Plot of the total phase difference as a function of the pillar height at 200 kV (blue squares) and 80kV(red circles). We added a suitable multiple of 2π to the experimental phase differences obtained from fitting the data. We fitted our data with a linear fitting function which allows us to estimate first the height corresponding to a 2π phase shift and then it gives us an approximate value of the contact potential.

From the two angular coefficients, it is also possible to estimate the pillar height corresponding to a a full 2π phase shift, where $h_{2\pi\_fit} = \frac{2\pi}{\beta}$, so that for 200 keV electrons the resulting pillar height is $\approx 0.83\mu\mathrm{m}$, while for 80 keV electrons the pillar height is $\approx 0.64\mu\mathrm{m}$. By comparing the $h_{2\pi}$ resulting from fitting the 200 keV data to the expected the expected value ($h_{2\pi} \approx 0.86\mu\mathrm{m}$ as estimated from Eq. 282), it can be seen that the two values are in agreement within a small margin of error. We can then determine an approximate value for the potential difference experienced by the electrons when they travel through the pillar to be $V_{C-exp} \approx 1.04\ \mathrm{V}$.

Conclusions

The fitted values of the phase difference shown in Figure 81 are consistent with the generation of constant phase shifts by Janus bimetallic cylindrical structures that behave as drift tubes. More systematic experiments are required in the future to assess the influence on the phase shift of experimental parameters associated with the fabrication process, as well as possible effects of electron beam induced charging of both the structures and their supports. When improved control of the phase shift is achieved, it may be possible to use such structures as Zernike phase plates in transmission electron microscopy, as they are more compact than tuneable drift tube devices and therefore promise to allow a greater number of spatial frequencies to contribute to phase contrast images.

Beyond their use as Zernike phase plates, we foresee the application of such bimetallic devices for different forms of electron beam shaping. For example, they could be used to replace electrostatic elements in applications such as spiral phase plates [120,121,298–300], Hilbert phase plates and electrostatic orbital angular momentum sorters [139,143]. They could also be used to replace electro static phase plates for conformal mapping operations [240]. A two-dimensional array of such cylinders could be used to produce an arbitrary phase landscape without the problems that come with using material-based holograms. A similar concept was recently proposed for pixelated programmable phase plates [186,187]. When using bimetallic cylinders, the loss of in situ programmability would be compensated by the ability to fabricate a greater number of pixels per unit area. Indeed, it is possible to reduce of a factor 2 the radial dimension of the bimetallic hollow cylinder and when paired with the lack of necessity of bringing the contact-lines to each pixel, then this results in a greater pixel density with respect to a corresponding tunable device. A clear advantage should be the possibility to better reproduce the desired phase profile. On the other hand, apart from a loss of intensity from the increase in the pixel density, the main limiting factors in this case would be the width of the cylinder walls that cannot be reduced with the present technique and the decoherence effect (now at 10%) that would affect more dramatically the beam when using a smaller hole.

Chapter 9 – Conclusions

This thesis work reports the main research results that I obtained in my PhD. I have reported my contribution in the field of Electron Vortex Beams (EVBs), electron beams that carry an Orbital Angular Momentum (OAM), focusing in particular on how to efficiently measure the OAM of a generic beam. To aid the reader in achieving a better understanding in this research field, in this work I first introduced the theory behind electron microscopes (in particular a Transmission Electron Microscope - TEM), the dual-beam machine that I used to fabricate a multitude of Synthetic Computer-Generated Holograms (S-CGHs) and to modify the 3-D shape of the electrodes for our experiments on the demonstration of a working electrostatic Orbital Angular Momentum (OAM) sorter (device that we developed and that allows to measure the OAM spectrum of an EVB). I then introduced the concept of OAM in relation to Vortex Beams, focusing especially on Electron Vortex Beams (EVBs), showing how in a TEM it is possible to both generate free-space EVBs and to measure their OAM. Chapter 4 is meant to set the reader up to the methods used in the last three decades, first in photon optics and later in electron optics, to create structured beams using diffraction gratings (either amplitude ones or phase ones, but even mixed amplitude-phase ones). The original research results are presented from Chapter 5 and onwards. Since the majority of the results obtained during this PhD have already been published or are soon to be published, starting from Chapter 5 each chapter reports the text of the article plus a brief description on my role that led to the results reported. Nonetheless, when necessary, I introduced the required theoretical knowledge to aid a reader which may not be well versed in the field to understand the article.

The results reported in this thesis can be summarized as follow:

I. As shown in Chapter 5, we designed and realized a working electrostatic OAM sorter using MEMS (MicroElectroMechanical Systems) technology. The first element of the OAM sorter features three needles made of doped silicon (so that they have an almost metallic-like behaviour) in front of a grounded plate, which are connected through a custom-made MEMS holder (used to insert the MEMS chips at positions different from the sample plane in the TEM column) to an external potential generator. The second element of the OAM sorter has also been realized on a MEMS chip. The realisation of a complete electrostatic OAM sorter is important for three major reasons: first of all, since the two elements of this device use electrostatic potentials to modify the phase of the electrons wavefunction and the electrostatic potentials can be controlled externally by the user, it is possible to use the complete device at different accelerating potentials, unlike for what happens for holographic phase masks. Second, as mentioned in Chapter 3 section 3.3.2, most of the beam's intensity reaches the detector and only a small part of it is lost due to the presence of the electrodes in the beam path. Lastly, the OAM sorting resolution is improved with respect to the previously realized holographic version [23] (passing from $2.5\hbar$ to approximately $1.5\hbar$, nearing the theoretical

limit of $1\hbar$ and the peaks in the OAM spectrum are well distinguishable without deconvolution). Future refinements to the design paired computer based diagnostics (an example of which is reported in Chapter 6) may allow us to reach the theoretical resolution of $1\hbar$, which, when paired to the intrinsic almost lossless efficiency of this approach, will open the possibility to atomic column resolution in measurements on magnetic materials [72], but also to study other chiral systems such as plasmonic structures [197] or proteins [261].

II. In Chapter 6 we reported how a Convolution Neural Network can be used to determine with sufficient accuracy the size of detrimental effects thanks to its widely known capabilities in pattern recognition. Such effects cannot easily be recognized analytically or during the experiments. We first trained the Convolution Neural Network (CNN) with a set of simulated OAM spectra patterns where various detrimental effects such as defocus o mismatches were present (20000 simulated images for training + 2000 for validation), then we used it to analyse the OAM spectrum patterns. The CNN would give us a series of values as an output corresponding to the various mismatches, which we then used to simulate OAM spectra patterns. By comparing these simulated patterns to the experimental ones, we found that the two were in good agreement, meaning that one can employ a CNN for almost real time feedback on the possible required corrections during the experiments. We envisage that by directly employing the CNN during experiments we may be able to correct in real time for detrimental effects such as defocus, misalignments, wrong electrode excitation. This will allow to further improve the device sorting resolution and stability. Moreover, it should be possible to use this approach also for more general problems such as in developing adaptive optics for real time aberration.

III. Two new designs of an OAM sorter based on magnetic elements and its theoretical framework has been studied, the findings are shown in chapter 7. One design is based on elements run through by a current, while the other is based on magnetic materials. Both designs have been found to be realizable with already available technology. However, the one based on electrical currents is more advantageous with respect to the one with magnetic moment, since it is difficult to obtain such well-defined magnetic domain (in this case only an approximation of the theoretical device would be achievable). Nonetheless, there are a series of advantages that make the magnetic approach a valid alternative over the electrostatic one: first, it does not suffer from chromatic effects, i. e. changes in the electron energy do not modify its functioning. Second, this device should be better in reproducing the behaviour of the ideal OAM sorter as in the electrostatic approach the electrode suffers from the detrimental effect of mutual induction (it must be pointed out that in the electrostatic OAM sorter since the tips are quite distanced between each other, the effect of mutual induction is in most of the cases of small entity).

IV. Lastly, laterally to the main topic of my thesis, I realized a proof of concept for a new approach on the realisation of phase plates. Via FIB deposition and a subsequent metal evaporation we fabricated a series of bimetallic cylinders of various height. Thanks to the contact potential between the two dissimilar metals, that arises from the difference in the work-function of the two, we can compare our structures to two concentric toroidal charge distributions of varying height. We foresee that each cylinder can be used as a pixel of a complex phase landscape. Using the specimen with different height we were able to quantify the phase effect. Despite the 2π phase ambiguity, we were able to characterize the phase per unit thickness of the different cylinders by measuring the phase effect at two different accelerating potentials (80kV and 200kV), the results that we obtained were consistent with each other. This is an important result since it opens the way for more systematic experiments and for the realisation of a pixelated phase mask which should not suffer of the same drawbacks that other commonly used phase masks suffer [301], such as degradation under prolonged beam exposure. However, it is still true that this device unlike the one reported in Chapter 5 or in Ref. [186] or Ref. [187] is not tuneable. Nonetheless, this approach should allow to pack a large number of hollow cylinders per unit area with respect to Ref. [186] or Ref. [187].

As demonstrated throughout this thesis, the ability to freely manipulate the phase of an electron beam can allow to overcome some of the classical limits of electron microscopy, which is why there is such high interest in this field. Improvements in the design of the electrostatic OAM sorter and the real time application of the CNN to control the microscope may allow us to reach in the near future the theoretical limits of our devices. The possibility to acquire OAM-resolved EELS spectra will allow us to study with one more degree of freedom magnetic and chiral plasmonic structures. Not only this, but the OAM sorter can also be used to study other chiral structures such as proteins. On a more generic note, we predict that properly designed electrostatic or magnetic phase plates based on MEMS technology when used to modify charged particle optics will allow us to probe fundamental quantum mechanical properties (other than just the OAM) that were previously inaccessible [240].

Acknowledgments

All the research work presented in this thesis and its results have received funding from the European Unions Horizon 2020 Research and Innovation Programme under Grant Agreement No 766970 Q-SORT (H2020-FETOPEN-1-2016- 2017). Other funding agencies that have supported a part of the scientific works here reported are: the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 823717 (project ESTEEM3); the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 405553726 TRR 270; the Ontario Early Researcher Award (ERA); Canada Research Chairs (CRC) Program.

Bibliography

[1] A. Einstein, *Zur Elektrodynamik Bewegter Körper*, Ann. Phys. **322**, 891 (1905).

[2] A. Einstein, *Über Die von Der Molekularkinetischen Theorie Der Wärme Geforderte Bewegung von in Ruhenden Flüssigkeiten Suspendierten Teilchen*, Ann. Phys. **322**, 549 (1905).

[3] A. Einstein, *Über Einen Die Erzeugung Und Verwandlung Des Lichtes Betreffenden Heuristischen Gesichtspunkt*, Ann. Phys. **322**, 132 (1905).

[4] A. Einstein, *Ist Die Trägheit Eines Körpers von Seinem Energieinhalt Abhängig?*, Ann. Phys. **323**, 639 (1905).

[5] A. Einstein, *Eine Neue Bestimmung Der Moleküldimensionen*, Ann. Phys. **324**, 289 (1906).

[6] M. Planck, *Zur Theorie Des Gesetzes Der Energieverteilung Im Normalspectrum*, Verhandlungen Der Dtsch. Phys. Gesellschaft **2**, 237 (1900).

[7] A. Einstein, *Die Plancksche Theorie Der Strahlung Und Die Theorie Der Spezifischen Wärme*, Ann. Phys. **327**, 180 (1907).

[8] A. Einstein, *Zur Quantentheorie Der Strahlung*, Phys. Z. **18**, 124 (1917).

[9] L. DE BROGLIE, *Waves and Quanta*, Nature **112**, 540 (1923).

[10] G. P. THOMSON and A. REID, *Diffraction of Cathode Rays by a Thin Film*, Nature **119**, 890 (1927).

[11] C. J. Davisson and L. H. Germer, *Reflection of Electrons by a Crystal of Nickel*, Proc. Natl. Acad. Sci. **14**, 317 (1928).

[12] H. Busch, *The Mode of Operation of the Concentrating Coil in a Braun Tube*, Arch. Fur Elektrotechnik **18**, 583 (1927).

[13] M. Knoll and E. Ruska, *Das Elektronenmikroskop*, Zeitschrift Für Phys. **78**, 318 (1932).

[14] E. Schrödinger, *An Undulatory Theory of the Mechanics of Atoms and Molecules*, Phys. Rev. **28**, 1049 (1926).

[15] D. GABOR, *A New Microscopic Principle*, Nature **161**, 777 (1948).

[16] M. E. Haine and T. Mulvey, *The Formation of the Diffraction Image with Electrons in the Gabor Diffraction Microscope*, J. Opt. Soc. Am. **42**, 763 (1952).

[17] G. Mollenstedt and H. Duker, *Fresnelscher Interferenzversuch Mit Einem Biprisma Fur Elektronenwellen*, Naturwissenschaften **42**, 41 (1955).

[18] G. Moellenstedt and H. Dueker, *Beobachtungen Und Messungen an Biprisma-Interferenzen Mit Elektronenwellen*, Zeitschrift Fðr Phys. **145**, 377 (1956).

[19] G. F. Missiroli, G. Pozzi, and U. Valdre, *Electron Interferometry and Interference Electron Microscopy*, J. Phys. E. **14**, 649 (1981).

[20] M. Uchida and A. Tonomura, *Generation of Electron Beams Carrying Orbital Angular Momentum*, Nature **464**, 737 (2010).

[21] J. Verbeeck, H. Tian, and P. Schattschneider, *Production and Application of Electron Vortex Beams*, Nature **467**, 301 (2010).

[22] B. J. McMorran, A. Agrawal, I. M. Anderson, A. A. Herzing, H. J. Lezec, J. J. McClelland, and J. Unguris, *Electron Vortex Beams with High Quanta of Orbital Angular Momentum*, Science (80-.). **331**, 192 (2011).

[23] V. Grillo, A. H. Tavabi, F. Venturi, H. Larocque, R. Balboni, G. C. Gazzadi, S. Frabboni, P.-H. Lu, E. Mafakheri, F. Bouchard, R. E. Dunin-Borkowski, R. W. Boyd, M. P. J. Lavery, M. J. Padgett, and E. Karimi, *Measuring the Orbital Angular Momentum Spectrum of an Electron Beam*, Nat. Commun. **8**, 15536 (2017).

[24] M. Mousley, S. Eswara, O. De Castro, O. Bouton, N. Klingner, C. T. Koch, G. Hlawacek, and T. Wirtz, *Stationary Beam Full-Field Transmission Helium Ion Microscopy Using Sub-50 KeV He + : Projected Images and Intensity Patterns*, Beilstein J. Nanotechnol. **10**, 1648 (2019).

[25] E. Ruska, *The Development of the Electron Microscope and of Electron Microscopy*, Biosci. Rep. **7**, 607 (1987).

[26] E. Ruska, *Das Entstehen Des Elektronenmikroskops Und Der Elektronenmikroskopie (Nobel-Vortrag)*, Angew. Chemie **99**, 611 (1987).

[27] M. De Graef, *Introduction to Conventional Transmission Electron Microscopy* (Cambridge university press, 2003).

[28] H. Kohl and L. Reimer, *Transmission Electron Microscopy*, Vol. 36 (Springer New York, New York, NY, 2008).

[29] J. M. Zuo and J. C. H. Spence, *Advanced Transmission Electron Microscopy* (Springer New York, New York, NY, 2017).

[30] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Elsevier, 2013).

[31] H. Boersch, *Experimentelle Bestimmung Der Energieverteilung in Thermisch Ausgeloesten Elektronenstrahlen*, Zeitschrift Fur Phys. **139**, 115 (1954).

[32] *Jeol Glossary of TEM Terms*.

[33] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, and K. Urban, *Electron Microscopy Image Enhanced*, Nature **392**, 768 (1998).

[34] M. Haider, H. Rose, S. Uhlemann, B. Kabius, and K. Urban, *Towards 0.1 Nm Resolution with the First Spherically Corrected Transmission Electron Microscope*, J. Electron Microsc. (Tokyo). **47**, 395 (1998).

[35] O. L. Krivanek, N. Dellby, and A. R. Lupini, *Towards Sub-Å Electron Beams*, Ultramicroscopy **78**, 1 (1999).

[36] P. Pozzi, Giulio; Hawkes, *Advances in Imaging and Electron Physics - Particles and Waves in Electron Optics and Microscopy* (Academic Press, 2016).

[37] A. Claverie, *Transmission Electron Microscopy in Micro-Nanoelectronics* (John Wiley & Sons, 2013).

[38] J. Verbeeck, G. Bertoni, and H. Lichte, *A Holographic Biprism as a Perfect Energy Filter?*, Ultramicroscopy (2011).

[39] M. Takeda, H. Ina, and S. Kobayashi, *Fourier-Transform Method of Fringe-Pattern Analysis for Computer-Based Topography and Interferometry*, J. Opt. Soc. Am. **72**, 156 (1982).

[40] 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 **45**, 77 (1992).

[41] G. Matteucci, G. F. F. Missiroli, and G. Pozzi, *Electron Holography of Long-Range Electrostatic Fields*, in *Advances in Imaging and Electron Physics*, Vol. 122 (Elsevier, 2002), pp. 173–249.

[42] R. E. Dunin-Borkowski, T. Kasama, M. Beleggia, and G. Pozzi, *Lorentz Microscopy and Electron Holography of Magnetic Materials*, in *Handbook of Nanoscopy* (Wiley-VCH Verlag GmbH & Co.

KGaA, Weinheim, Germany, 2012), pp. 221–251.

[43] A. Kovács and R. E. Dunin-Borkowski, *Magnetic Imaging of Nanostructures Using Off-Axis Electron Holography*, in (2018), pp. 59–153.

[44] W. Zhou and Z. L. Wang, *Scanning Microscopy for Nanotechnology: Techniques and Applications* (Springer science & business media, 2007).

[45] P. J. Goodhew and J. Humphreys, *Electron Microscopy and Analysis* (CRC Press, 2000).

[46] T. E. Everhart and R. F. M. Thornley, *Wide-Band Detector for Micro-Microampere Low-Energy Electron Currents*, J. Sci. Instrum. **37**, 246 (1960).

[47] L. A. Giannuzzi, *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice* (Springer Science & Business Media, 2004).

[48] J. Orloff, L. Swanson, and M. Utlaut, *High Resolution Focused Ion Beams: FIB and Its Applications: Fib and Its Applications: The Physics of Liquid Metal Ion Sources and Ion Optics and Their Application to Focused Ion Beam Technology* (Springer Science & Business Media, 2003).

[49] H. H. Andersen and H. L. Bay, *Sputtering Yield Measurements*, in *Sputtering by Particle Bombardment I* (Springer, 1981), pp. 145–218.

[50] P. Sigmund, *Sputtering by Ion Bombardment Theoretical Concepts*, in (1981), pp. 9–71.

[51] J. F. Ziegler, M. D. Ziegler, and J. P. Biersack, *SRIM – The Stopping and Range of Ions in Matter (2010)*, Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms **268**, 1818 (2010).

[52] Z. L. Liau and J. W. Mayer, *Ion Bombardment Effects on Material Composition*, in *Treatise on Materials Science and Technology*, Vol. 18 (Elsevier, 1980), pp. 17–50.

[53] M. Nastasi, N. Michael, J. Mayer, J. K. Hirvonen, and M. James, *Ion-Solid Interactions: Fundamentals and Applications* (Cambridge University Press, 1996).

[54] J. Masajada and B. Dubik, *Optical Vortex Generation by Three Plane Wave Interference*, Opt. Commun. **198**, 21 (2001).

[55] J. F. Nye and M. V. Berry, *Dislocations in Wave Trains*, Proc. R. Soc. London. A. Math. Phys. Sci. **336**, 165 (1974).

[56] M. S. Soskin and M. V. Vasnetsov, *Singular Optics*, in (2001), pp. 219–276.

[57] M. R. Dennis, K. O'Holleran, and M. J. Padgett, *Chapter 5 Singular Optics: Optical Vortices and Polarization Singularities*, in (2009), pp. 293–363.

[58] H. Wolter, *Zur Frage Des Lichtweges Bei Totalreflexion*, Zeitschrift Für Naturforsch. A **5**, 276 (1950).

[59] W. Braunbek and G. Laukien, *Features of Refraction by a Semi-Plane*, Optik (Stuttg). **9**, 174 (1952).

[60] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, (1999).

[61] C. Cohen-Tannoudji, B. Diu, F. Laloe, and B. Dui, *Quantum Mechanics (2 Vol. Set)*.

[62] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, Vol. 3 (Elsevier, 2013).

[63] V. Fock, *Bemerkung Zur Quantelung Des Harmonischen Oszillators Im Magnetfeld*, Zeitschrift Für Phys. **47**, 446 (1928).

[64] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, *Orbital Angular Momentum of*

Light and the Transformation of Laguerre-Gaussian Laser Modes, Phys. Rev. A **45**, 8185 (1992).

[65] K. Y. Bliokh, *Geometrical Optics of Beams with Vortices: Berry Phase and Orbital Angular Momentum Hall Effect*, Phys. Rev. Lett. **97**, 043901 (2006).

[66] K. Y. Bliokh, Y. P. Bliokh, S. Savel'ev, and F. Nori, *Semiclassical Dynamics of Electron Wave Packet States with Phase Vortices*, Phys. Rev. Lett. **99**, 190404 (2007).

[67] J. Harris, V. Grillo, E. Mafakheri, G. C. Gazzadi, S. Frabboni, R. W. Boyd, and E. Karimi, *Structured Quantum Waves*, Nat. Phys. **11**, 629 (2015).

[68] J. Rusz and S. Bhowmick, *Boundaries for Efficient Use of Electron Vortex Beams to Measure Magnetic Properties*, Phys. Rev. Lett. **111**, 105504 (2013).

[69] V. Grillo, T. R. Harvey, F. Venturi, J. S. Pierce, R. Balboni, F. Bouchard, G. Carlo Gazzadi, S. Frabboni, A. H. Tavabi, Z.-A. Li, R. E. Dunin-Borkowski, R. W. Boyd, B. J. McMorran, and E. Karimi, *Observation of Nanoscale Magnetic Fields Using Twisted Electron Beams*, Nat. Commun. **8**, 689 (2017).

[70] C. Hébert and P. Schattschneider, *A Proposal for Dichroic Experiments in the Electron Microscope*, Ultramicroscopy **96**, 463 (2003).

[71] P. Schattschneider, S. Rubino, C. Hébert, J. Rusz, J. Kuneš, P. Novák, E. Carlino, M. Fabrizioli, G. Panaccione, and G. Rossi, *Detection of Magnetic Circular Dichroism Using a Transmission Electron Microscope*, Nature **441**, 486 (2006).

[72] E. Rotunno, M. Zanfrognini, S. Frabboni, J. Rusz, R. E. Dunin Borkowski, E. Karimi, and V. Grillo, *Orbital Angular Momentum Resolved Electron Magnetic Chiral Dichroism*, Phys. Rev. B **100**, 224409 (2019).

[73] E. Rotunno, A. H. Tavabi, E. Yucelen, S. Frabboni, R. E. Dunin Borkowski, E. Karimi, B. J. McMorran, and V. Grillo, *Electron-Beam Shaping in the Transmission Electron Microscope: Control of Electron-Beam Propagation Along Atomic Columns*, Phys. Rev. Appl. **11**, 044072 (2019).

[74] D. Song, Z. Wang, and J. Zhu, *Magnetic Measurement by Electron Magnetic Circular Dichroism in the Transmission Electron Microscope*, Ultramicroscopy **201**, 1 (2019).

[75] M. Zanfrognini, E. Rotunno, J. Rusz, R. E. Dunin Borkowski, E. Karimi, S. Frabboni, and V. Grillo, *Dynamical Diffraction Effects in STEM Orbital Angular Momentum Resolved Electron Energy-Loss Magnetic Chiral Dichroism*, Phys. Rev. B **102**, 184420 (2020).

[76] M. V. Berry, *Paraxial Beams of Spinning Light*, in edited by M. S. Soskin (1998), pp. 6–11.

[77] K. Y. Bliokh and F. Nori, *Transverse and Longitudinal Angular Momenta of Light*, Phys. Rep. **592**, 1 (2015).

[78] P. Schattschneider, V. Grillo, and D. Aubry, *Spin Polarisation with Electron Bessel Beams*, Ultramicroscopy **176**, 188 (2017).

[79] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial, *Measuring the Orbital Angular Momentum of a Single Photon*, Phys. Rev. Lett. **88**, 257901 (2002).

[80] K. Y. Bliokh, M. R. Dennis, and F. Nori, *Relativistic Electron Vortex Beams: Angular Momentum and Spin-Orbit Interaction*, Phys. Rev. Lett. **107**, 174802 (2011).

[81] K. Y. Y. Bliokh, I. P. P. Ivanov, G. Guzzinati, L. Clark, R. Van Boxem, A. Béché, R. Juchtmans, M. A. A. Alonso, P. Schattschneider, F. Nori, and J. Verbeeck, *Theory and Applications of Free-Electron Vortex States*, Phys. Rep. **690**, 1 (2017).

[82] S. M. Barnett, *Relativistic Electron Vortices*, Phys. Rev. Lett. **118**, 114802 (2017).

[83] K. Y. Bliokh, M. R. Dennis, and F. Nori, *Position, Spin, and Orbital Angular Momentum of a Relativistic Electron*, Phys. Rev. A **96**, 023622 (2017).

[84] Y. D. Han, T. Choi, and S. Y. Cho, *Singularity of a Relativistic Vortex Beam and Proper Relativistic Observables*, Sci. Rep. **10**, 7417 (2020).

[85] J. Durnin, *Exact Solutions for Nondiffracting Beams I The Scalar Theory*, J. Opt. Soc. Am. A **4**, 651 (1987).

[86] J. Durnin, J. J. Miceli, and J. H. Eberly, *Diffraction-Free Beams*, Phys. Rev. Lett. **58**, 1499 (1987).

[87] J. Durnin, J. J. Miceli, and J. H. Eberly, *Durnin, Miceli, and Eberly Reply*, Phys. Rev. Lett. **66**, 838 (1991).

[88] D. McGloin and K. Dholakia, *Bessel Beams: Diffraction in a New Light*, Contemp. Phys. **46**, 15 (2005).

[89] Y. Lin, W. Seka, J. H. Eberly, H. Huang, and D. L. Brown, *Experimental Investigation of Bessel Beam Characteristics*, Appl. Opt. **31**, 2708 (1992).

[90] V. Grillo, E. Karimi, G. C. Gazzadi, S. Frabboni, M. R. Dennis, and R. W. Boyd, *Generation of Nondiffracting Electron Bessel Beams*, Phys. Rev. X (2014).

[91] P. Schattschneider and J. Verbeeck, *Theory of Free Electron Vortices*, Ultramicroscopy **111**, 1461 (2011).

[92] S. M. Lloyd, M. Babiker, G. Thirunavukkarasu, and J. Yuan, *Electron Vortices: Beams with Orbital Angular Momentum*, Rev. Mod. Phys. **89**, 035004 (2017).

[93] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Vol. 55 (US Government printing office, 1964).

[94] L. G. Gouy, *Sur Une Propriété Nouvelle Des Ondes Lumineuses*, Comptes Rendus l'Académie Des Sci. Paris **110**, 1251 (1890).

[95] R. W. Boyd, *Intuitive Explanation of the Phase Anomaly of Focused Light Beams*, J. Opt. Soc. Am. **70**, 877 (1980).

[96] A. E. Siegman, *Lasers University Science Books*, Mill Val. CA **37**, 169 (1986).

[97] S. Feng and H. G. Winful, *Physical Origin of the Gouy Phase Shift*, Opt. Lett. **26**, 485 (2001).

[98] K. Y. Bliokh, P. Schattschneider, J. Verbeeck, and F. Nori, *Electron Vortex Beams in a Magnetic Field: A New Twist on Landau Levels and Aharonov-Bohm States*, Phys. Rev. X **2**, 41011 (2012).

[99] P. Schattschneider, T. Schachinger, M. Stöger-Pollach, S. Löffler, A. Steiger-Thirsfeld, K. Y. Bliokh, and F. Nori, *Imaging the Dynamics of Free-Electron Landau States*, Nat. Commun. **5**, 4586 (2014).

[100] B. J. McMorran, A. Agrawal, P. A. Ercius, V. Grillo, A. A. Herzing, T. R. Harvey, M. Linck, and J. S. Pierce, *Origins and Demonstrations of Electrons with Orbital Angular Momentum*, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **375**, 20150434 (2017).

[101] H. Larocque, I. Kaminer, V. Grillo, G. Leuchs, M. J. Padgett, R. W. Boyd, M. Segev, and E. Karimi, *'Twisted' Electrons*, Contemp. Phys. **59**, 126 (2018).

[102] R. Shiloh, P.-H. Lu, R. Remez, A. H. Tavabi, G. Pozzi, R. E. Dunin-Borkowski, and A. Arie, *Nanostructuring of Electron Beams*, Phys. Scr. **94**, 034004 (2019).

[103] M. W. Beijersbergen, R. P. C. Coerwinkel, M. Kristensen, and J. P. Woerdman, *Helical-Wavefront Laser Beams Produced with a Spiral Phaseplate*, Opt. Commun. **112**, 321 (1994).

[104] G. A. Turnbull, D. A. Robertson, G. M. Smith, L. Allen, and M. J. Padgett, *The Generation of Free-*

Space Laguerre-Gaussian Modes at Millimetre-Wave Frequencies by Use of a Spiral Phaseplate, Opt. Commun. **127**, 183 (1996).

[105] A. G. Peele, P. J. McMahon, D. Paterson, C. Q. Tran, A. P. Mancuso, K. A. Nugent, J. P. Hayes, E. Harvey, B. Lai, and I. McNulty, *Observation of an X-Ray Vortex*, Opt. Lett. **27**, 1752 (2002).

[106] M. V Berry, *Optical Vortices Evolving from Helicoidal Integer and Fractional Phase Steps*, J. Opt. A Pure Appl. Opt. **6**, 259 (2004).

[107] J. Leach, E. Yao, and M. J. Padgett, *Observation of the Vortex Structure of a Non-Integer Vortex Beam*, New J. Phys. **6**, 71 (2004).

[108] J. B. Götte, K. O'Holleran, D. Preece, F. Flossmann, S. Franke-Arnold, S. M. Barnett, and M. J. Padgett, *Light Beams with Fractional Orbital Angular Momentum and Their Vortex Structure*, Opt. Express **16**, 993 (2008).

[109] R. Shiloh, Y. Lereah, Y. Lilach, and A. Arie, *Sculpturing the Electron Wave Function Using Nanoscale Phase Masks*, Ultramicroscopy **144**, 26 (2014).

[110] A. Béché, R. Winkler, H. Plank, F. Hofer, and J. Verbeeck, *Focused Electron Beam Induced Deposition as a Tool to Create Electron Vortices*, Micron **80**, 34 (2016).

[111] V. Y. Bazhenov, M. V Vasnetsov, and M. S. Soskin, *Laser Beams with Screw Dislocations in Their Wavefronts*, Jetp Lett **52**, 429 (1990).

[112] N. R. Heckenberg, R. McDuff, C. P. Smith, H. Rubinsztein-Dunlop, and M. J. Wegener, *Laser Beams with Phase Singularities*, Opt. Quantum Electron. **24**, S951 (1992).

[113] E. Mafakheri, A. H. Tavabi, P.-H. Lu, R. Balboni, F. Venturi, C. Menozzi, G. C. Gazzadi, S. Frabboni, A. Sit, R. E. Dunin-Borkowski, E. Karimi, and V. Grillo, *Realization of Electron Vortices with Large Orbital Angular Momentum Using Miniature Holograms Fabricated by Electron Beam Lithography*, Appl. Phys. Lett. **110**, 093113 (2017).

[114] A. M. Blackburn and J. C. Loudon, *Vortex Beam Production and Contrast Enhancement from a Magnetic Spiral Phase Plate*, Ultramicroscopy **136**, 127 (2014).

[115] A. Béché, R. Van Boxem, G. Van Tendeloo, and J. Verbeeck, *Magnetic Monopole Field Exposed by Electrons*, Nat. Phys. **10**, 26 (2014).

[116] A. Béché, R. Juchtmans, and J. Verbeeck, *Efficient Creation of Electron Vortex Beams for High Resolution STEM Imaging*, Ultramicroscopy **178**, 12 (2017).

[117] P. A. M. Dirac, *Quantised Singularities in the Electromagnetic Field,* Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character **133**, 60 (1931).

[118] Y. M. Shnir, *Magnetic Monopoles* (Springer Science & Business Media, 2006).

[119] A. Tonomura, *Applications of Electron Holography*, Rev. Mod. Phys. **59**, 639 (1987).

[120] G. Pozzi, P.-H. 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 **181**, 191 (2017).

[121] A. H. Tavabi, H. Larocque, P.-H. Lu, M. Duchamp, V. Grillo, E. Karimi, R. E. Dunin-Borkowski, and G. Pozzi, *Generation of Electron Vortices Using Nonexact Electric Fields*, Phys. Rev. Res. **2**, 013185 (2020).

[122] A. H. Tavabi, V. Migunov, C. Dwyer, R. E. Dunin-Borkowski, and G. Pozzi, *Tunable Caustic Phenomena in Electron Wavefields*, Ultramicroscopy **157**, 57 (2015).

[123] G. M. Vanacore, G. Berruto, I. Madan, E. Pomarico, P. Biagioni, R. J. Lamb, D. McGrouther, O. Reinhardt, I. Kaminer, B. Barwick, H. Larocque, V. Grillo, E. Karimi, F. J. García de Abajo, and F. Carbone, *Ultrafast Generation and Control of an Electron Vortex Beam via Chiral Plasmonic near Fields*, Nat. Mater. **18**, 573 (2019).

[124] G. M. Vanacore, A. W. P. Fitzpatrick, and A. H. Zewail, *Four-Dimensional Electron Microscopy: Ultrafast Imaging, Diffraction and Spectroscopy in Materials Science and Biology*, Nano Today **11**, 228 (2016).

[125] L. Piazza, D. J. Masiel, T. LaGrange, B. W. Reed, B. Barwick, and F. Carbone, *Design and Implementation of a Fs-Resolved Transmission Electron Microscope Based on Thermionic Gun Technology*, Chem. Phys. **423**, 79 (2013).

[126] V. A. Lobastov, R. Srinivasan, and A. H. Zewail, *Four-Dimensional Ultrafast Electron Microscopy*, Proc. Natl. Acad. Sci. **102**, 7069 (2005).

[127] A. T. O'Neil, I. MacVicar, L. Allen, and M. J. Padgett, *Intrinsic and Extrinsic Nature of the Orbital Angular Momentum of a Light Beam*, Phys. Rev. Lett. **88**, 053601 (2002).

[128] A. M. Amaral, E. L. Falcão-Filho, and C. B. de Araújo, *Characterization of Topological Charge and Orbital Angular Momentum of Shaped Optical Vortices*, Opt. Express **22**, 30315 (2014).

[129] K. Saitoh, Y. Hasegawa, K. Hirakawa, N. Tanaka, and M. Uchida, *Measuring the Orbital Angular Momentum of Electron Vortex Beams Using a Forked Grating*, Phys. Rev. Lett. **111**, 074801 (2013).

[130] G. Guzzinati, L. Clark, A. Béché, and J. Verbeeck, *Measuring the Orbital Angular Momentum of Electron Beams*, Phys. Rev. A **89**, 025803 (2014).

[131] L. Clark, G. Guzzinati, A. Béché, A. Lubk, and J. Verbeeck, *Symmetry-Constrained Electron Vortex Propagation*, Phys. Rev. A **93**, 063840 (2016).

[132] L. Clark, A. Béché, G. Guzzinati, and J. Verbeeck, *Quantitative Measurement of Orbital Angular Momentum in Electron Microscopy*, Phys. Rev. A **89**, 053818 (2014).

[133] H. Larocque, F. Bouchard, V. Grillo, A. Sit, S. Frabboni, R. E. Dunin-Borkowski, M. J. Padgett, R. W. Boyd, and E. Karimi, *Nondestructive Measurement of Orbital Angular Momentum for an Electron Beam*, Phys. Rev. Lett. **117**, 154801 (2016).

[134] P. Schattschneider, M. Stöger-Pollach, and J. Verbeeck, *Novel Vortex Generator and Mode Converter for Electron Beams*, Phys. Rev. Lett. **109**, 084801 (2012).

[135] C. Kramberger, S. Löffler, T. Schachinger, P. Hartel, J. Zach, and P. Schattschneider, *π/2 Mode Converters and Vortex Generators for Electrons*, Ultramicroscopy **204**, 27 (2019).

[136] G. C. G. Berkhout, M. P. J. Lavery, J. Courtial, M. W. Beijersbergen, and M. J. Padgett, *Efficient Sorting of Orbital Angular Momentum States of Light*, Phys. Rev. Lett. **105**, 153601 (2010).

[137] M. Mirhosseini, M. Malik, Z. Shi, and R. W. Boyd, *Efficient Separation of the Orbital Angular Momentum Eigenstates of Light*, Nat. Commun. **4**, 2781 (2013).

[138] B. J. McMorran, T. R. Harvey, and M. P. J. J. Lavery, *Efficient Sorting of Free Electron Orbital Angular Momentum*, New J. Phys. **19**, 023053 (2017).

[139] A. H. Tavabi, P. Rosi, E. Rotunno, A. Roncaglia, L. Belsito, S. Frabboni, G. Pozzi, G. C. Gazzadi, P.-H. 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**, 094802 (2021).

[140] O. Bryngdahl, *Geometrical Transformations in Optics**, J. Opt. Soc. Am. **64**, 1092 (1974).

[141] Y. Saito, S. Komatsu, and H. Ohzu, *Scale and Rotation Invariant Real Time Optical Correlator Using Computer Generated Hologram*, Opt. Commun. **47**, 8 (1983).

[142] W. J. Hossack, A. M. Darling, and A. Dahdouh, *Coordinate Transformations with Multiple Computer-Generated Optical Elements*, J. Mod. Opt. **34**, 1235 (1987).

[143] G. Pozzi, V. Grillo, P.-H. Lu, A. H. Tavabi, E. Karimi, and R. E. Dunin-Borkowski, *Design of Electrostatic Phase Elements for Sorting the Orbital Angular Momentum of Electrons*, Ultramicroscopy **208**, 112861 (2020).

[144] G. Pozzi, P. Rosi, A. Tavabi, E. Karimi, R. Dunin-Borkowski, and V. Grillo, *A Sorter for Electrons Based on Magnetic Elements*, (2021).

[145] J. Verbeeck, P. Schattschneider, S. Lazar, M. Stöger-Pollach, S. Löffler, A. Steiger-Thirsfeld, and G. Van Tendeloo, *Atomic Scale Electron Vortices for Nanoresearch*, Appl. Phys. Lett. **99**, 203109 (2011).

[146] J. C. Idrobo and S. J. Pennycook, *Vortex Beams for Atomic Resolution Dichroism*, Microscopy **60**, NP (2011).

[147] J. Rusz, S. Bhowmick, M. Eriksson, and N. Karlsson, *Scattering of Electron Vortex Beams on a Magnetic Crystal: Towards Atomic-Resolution Magnetic Measurements*, Phys. Rev. B **89**, 134428 (2014).

[148] P. Schattschneider, S. Löffler, M. Stöger-Pollach, and J. Verbeeck, *Is Magnetic Chiral Dichroism Feasible with Electron Vortices?*, Ultramicroscopy **136**, 81 (2014).

[149] A. P. Hitchcock, *Near Edge Electron Energy Loss Spectroscopy: Comparison to X-Ray Absorption*, Jpn. J. Appl. Phys. **32**, 176 (1993).

[150] J. C. Fuggle and J. E. Inglesfield, editors , *Unoccupied Electronic States*, Vol. 69 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1992).

[151] H. Kohl and H. Rose, *Theory of Image Formation by Inelastically Scattered Electrons in the Electron Microscope*, in (1985), pp. 173–227.

[152] D. Pohl, S. Schneider, P. Zeiger, J. Rusz, P. Tiemeijer, S. Lazar, K. Nielsch, and B. Rellinghaus, *Atom Size Electron Vortex Beams with Selectable Orbital Angular Momentum*, Sci. Rep. **7**, 934 (2017).

[153] R. Van Boxem, B. Partoens, and J. Verbeeck, *Inelastic Electron-Vortex-Beam Scattering*, Phys. Rev. A **91**, 032703 (2015).

[154] J. Rusz, J.-C. Idrobo, and S. Bhowmick, *Achieving Atomic Resolution Magnetic Dichroism by Controlling the Phase Symmetry of an Electron Probe*, Phys. Rev. Lett. **113**, 145501 (2014).

[155] J. Rusz, J.-C. Idrobo, and L. Wrang, *Vorticity in Electron Beams: Definition, Properties, and Its Relationship with Magnetism*, Phys. Rev. B **94**, 144430 (2016).

[156] P. Schattschneider, B. Schaffer, I. Ennen, and J. Verbeeck, *Mapping Spin-Polarized Transitions with Atomic Resolution*, Phys. Rev. B **85**, 134422 (2012).

[157] J. C. Idrobo, J. Rusz, J. Spiegelberg, M. A. McGuire, C. T. Symons, R. R. Vatsavai, C. Cantoni, and A. R. Lupini, *Detecting Magnetic Ordering with Atomic Size Electron Probes*, Adv. Struct. Chem. Imaging **2**, 5 (2016).

[158] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Springer US, Boston, MA, 2010).

[159] S. Löffler and P. Schattschneider, *Elastic Propagation of Fast Electron Vortices through Crystals*, Acta Crystallogr. Sect. A Found. Crystallogr. **68**, 443 (2012).

[160] B. G. Mendis, *Dynamic Scattering of Electron Vortex Beams – A Bloch Wave Analysis*,

Ultramicroscopy **149**, 74 (2015).

[161] T. R. Harvey, V. Grillo, and B. J. McMorran, *Proposal for Magnetic Dichroism With a Standard STEM Probe Beam*, Microsc. Microanal. **22**, 1708 (2016).

[162] T. Schachinger, S. Löffler, A. Steiger-Thirsfeld, M. Stöger-Pollach, S. Schneider, D. Pohl, B. Rellinghaus, and P. Schattschneider, *EMCD with an Electron Vortex Filter: Limitations and Possibilities*, Ultramicroscopy **179**, 15 (2017).

[163] L. B. Lesem, P. M. Hirsch, and J. A. Jordan, *Scientific Applications: Computer Synthesis of Holograms for 3-D Display*, Commun. ACM **11**, 661 (1968).

[164] V. Grillo and E. Rotunno, *STEM_CELL: A Software Tool for Electron Microscopy: Part I—Simulations*, Ultramicroscopy **125**, 97 (2013).

[165] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, *Exact Solution to Simultaneous Intensity and Phase Encryption with a Single Phase-Only Hologram*, Opt. Lett. **38**, 3546 (2013).

[166] R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope* (Springer Science & Business Media, 2011).

[167] K. Iakoubovskii, K. Mitsuishi, Y. Nakayama, and K. Furuya, *Mean Free Path of Inelastic Electron Scattering in Elemental Solids and Oxides Using Transmission Electron Microscopy: Atomic Number Dependent Oscillatory Behavior*, Phys. Rev. B **77**, 104102 (2008).

[168] V. Grillo, G. Carlo Gazzadi, E. Karimi, E. Mafakheri, R. W. Boyd, and S. Frabboni, *Highly Efficient Electron Vortex Beams Generated by Nanofabricated Phase Holograms*, Appl. Phys. Lett. **104**, 043109 (2014).

[169] T. R Harvey, J. S Pierce, A. K Agrawal, P. Ercius, M. Linck, and B. J. McMorran, *Efficient Diffractive Phase Optics for Electrons*, New J. Phys. **16**, 093039 (2014).

[170] S. Bhattacharyya, C. T. Koch, and M. Rühle, *Projected Potential Profiles across Interfaces Obtained by Reconstructing the Exit Face Wave Function from through Focal Series*, Ultramicroscopy **106**, 525 (2006).

[171] L. Grünewald, D. Gerthsen, and S. Hettler, *Fabrication of Phase Masks from Amorphous Carbon Thin Films for Electron-Beam Shaping*, Beilstein J. Nanotechnol. **10**, 1290 (2019).

[172] A. Auslender, M. Halabi, G. Levi, O. Diéguez, and A. Kohn, *Measuring the Mean Inner Potential of Al2O3 Sapphire Using Off-Axis Electron Holography*, Ultramicroscopy **198**, 18 (2019).

[173] M. Schowalter, J. T. Titantah, D. Lamoen, and P. Kruse, *Ab Initio Computation of the Mean Inner Coulomb Potential of Amorphous Carbon Structures*, Appl. Phys. Lett. **86**, 112102 (2005).

[174] M. Wanner, D. Bach, D. Gerthsen, R. Werner, and B. Tesche, *Electron Holography of Thin Amorphous Carbon Films: Measurement of the Mean Inner Potential and a Thickness-Independent Phase Shift*, Ultramicroscopy **106**, 341 (2006).

[175] A. Harscher and H. Lichte, *Inelastic Mean Free Path and Mean Inner Potential of Carbon Foil and Vitrified Ice Measured with Electron Holography*, ICEM14, Cancun, Mex. **31**, 553 (1998).

[176] A. Sanchez and M. A. Ochando, *Calculation of the Mean Inner Potential*, J. Phys. C Solid State Phys. **18**, 33 (1985).

[177] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy* (Springer US, Boston, MA, 2009).

[178] V. Grillo, E. Karimi, R. Balboni, G. C. Gazzadi, F. Venturi, S. Frabboni, J. S. Pierce, B. J. McMorran, and R. W. Boyd, *Electron Holograms Encoding Amplitude and Phase for the Generation of Arbitrary Wavefunctions*, Microsc. Microanal. **21**, 503 (2015).

[179] W.-H. Lee, *Binary Computer-Generated Holograms*, Appl. Opt. **18**, 3661 (1979).

[180] W.-H. Lee, *Binary Synthetic Holograms*, Appl. Opt. **13**, 1677 (1974).

[181] T. Schachinger, A. Steiger-Thirsfeld, S. Löffler, M. Stöger-Pollach, S. Schneider, D. Pohl, B. Rellinghaus, and P. Schattschneider, *Preparation of High Fidelity Holographic Vortex Masks Using Advanced FIB Milling Strategies*, in *European Microscopy Congress 2016: Proceedings* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2016), pp. 717–718.

[182] W. Hu, K. Sarveswaran, M. Lieberman, and G. H. Bernstein, *High-Resolution Electron Beam Lithography and DNA Nano-Patterning for Molecular QCA*, IEEE Trans. Nanotechnol. **4**, 312 (2005).

[183] M. Häffner, A. Haug, A. Heeren, M. Fleischer, H. Peisert, T. Chassé, and D. P. Kern, *Influence of Temperature on HSQ Electron-Beam Lithography*, J. Vac. Sci. Technol. B Microelectron. Nanom. Struct. **25**, 2045 (2007).

[184] S. Hettler, L. Radtke, L. Grünewald, Y. Lisunova, O. Peric, J. Brugger, and S. Bonanni, *Phase Masks for Electron Microscopy Fabricated by Thermal Scanning Probe Lithography*, Micron **127**, 102753 (2019).

[185] L. Clark, A. Béché, G. Guzzinati, A. Lubk, M. Mazilu, R. Van Boxem, and J. Verbeeck, *Exploiting Lens Aberrations to Create Electron-Vortex Beams*, Phys. Rev. Lett. **111**, 064801 (2013).

[186] J. Verbeeck, A. Béché, K. Müller-Caspary, G. Guzzinati, M. A. Luong, and M. Den Hertog, *Demonstration of a 2 × 2 Programmable Phase Plate for Electrons*, Ultramicroscopy **190**, 58 (2018).

[187] P. Thakkar, V. A. Guzenko, P.-H. Lu, R. E. Dunin-Borkowski, J. P. Abrahams, and S. Tsujino, *Fabrication of Low Aspect Ratio Three-Element Boersch Phase Shifters for Voltage-Controlled Three Electron Beam Interference*, J. Appl. Phys. **128**, 134502 (2020).

[188] F. VENTURI, *Nuovi Metodi per La Manipolazione e La Caratterizzazione Della Fase Nel Microscopio Elettronico in Trasmissione*, (2018).

[189] C. W. Johnson, D. H. Bauer, and B. J. McMorran, *Improved Control of Electron Computer-Generated Holographic Grating Groove Profiles Using Ion Beam Gas-Assisted Etching*, Appl. Opt. **59**, 1594 (2020).

[190] E. Karimi, G. Zito, B. Piccirillo, L. Marrucci, and E. Santamato, *Hypergeometric-Gaussian Modes*, Opt. Lett. **32**, 3053 (2007).

[191] R. D. Guenther, *Modern Optics* (Wiley, 1990).

[192] T. D. Visser and E. Wolf, *The Origin of the Gouy Phase Anomaly and Its Generalization to Astigmatic Wavefields*, Opt. Commun. **283**, 3371 (2010).

[193] A. M. Yao and M. J. Padgett, *Orbital Angular Momentum: Origins, Behavior and Applications*, Adv. Opt. Photonics **3**, 161 (2011).

[194] F. Venturi, M. Campanini, G. C. Gazzadi, R. Balboni, S. Frabboni, R. W. Boyd, R. E. Dunin-Borkowski, E. Karimi, and V. Grillo, *Phase Retrieval of an Electron Vortex Beam Using Diffraction Holography*, Appl. Phys. Lett. **111**, (2017).

[195] V. Grillo, B. J. McMorran, R. W. Boyd, G. C. Gazzadi, F. Venturi, S. Frabboni, R. Balboni, E. Karimi, and J. S. Pierce, *Electron Holograms Encoding Amplitude and Phase for the Generation of Arbitrary Wavefunctions*, Microsc. Microanal. (2016).

[196] C. W. Johnson, J. S. Pierce, R. C. Moraski, A. E. Turner, A. T. Greenberg, W. S. Parker, and B. J. McMorran, *Exact Design of Complex Amplitude Holograms for Producing Arbitrary Scalar Fields*, Opt. Express **28**, 17334 (2020).

[197] M. Zanfrognini, E. Rotunno, S. Frabboni, A. Sit, E. Karimi, U. Hohenester, and V. Grillo, *Orbital*

Angular Momentum and Energy Loss Characterization of Plasmonic Excitations in Metallic Nanostructures in TEM, ACS Photonics **6**, 620 (2019).

[198] J. H. McLeod, *The Axicon: A New Type of Optical Element*, J. Opt. Soc. Am. **44**, 592 (1954).

[199] Z. Bin and L. Zhu, *Diffraction Property of an Axicon in Oblique Illumination*, Appl. Opt. **37**, 2563 (1998).

[200] T. Tanaka and S. Yamamoto, *Comparison of Aberration between Axicon and Lens*, Opt. Commun. **184**, 113 (2000).

[201] A. Thaning, Z. Jaroszewicz, and A. T. Friberg, *Diffractive Axicons in Oblique Illumination: Analysis and Experiments and Comparison with Elliptical Axicons*, Appl. Opt. **42**, 9 (2003).

[202] A. Vasara, J. Turunen, and A. T. Friberg, *Realization of General Nondiffracting Beams with Computer-Generated Holograms*, J. Opt. Soc. Am. A **6**, 1748 (1989).

[203] J. A. Davis, E. Carcole, and D. M. Cottrell, *Intensity and Phase Measurements of Nondiffracting Beams Generated with a Magneto-Optic Spatial Light Modulator*, Appl. Opt. **35**, 593 (1996).

[204] J. A. Davis, E. Carcole, and D. M. Cottrell, *Nondiffracting Interference Patterns Generated with Programmable Spatial Light Modulators*, Appl. Opt. **35**, 599 (1996).

[205] K. Uehara and H. Kikuchi, *Generation of Nearly Diffraction-Free Laser Beams*, Appl. Phys. B Photophysics Laser Chem. **48**, 125 (1989).

[206] A. J. Cox and D. C. Dibble, *Nondiffracting Beam from a Spatially Filtered Fabry–Perot Resonator*, J. Opt. Soc. Am. A **9**, 282 (1992).

[207] V. Grillo, J. Harris, G. C. Gazzadi, R. Balboni, E. Mafakheri, M. R. Dennis, S. Frabboni, R. W. Boyd, and E. Karimi, *Generation and Application of Bessel Beams in Electron Microscopy*, Ultramicroscopy **166**, 48 (2016).

[208] K. Saitoh, K. Hirakawa, H. Nambu, N. Tanaka, and M. Uchida, *Generation of Electron Bessel Beams with Nondiffractive Spreading by a Nanofabricated Annular Slit*, J. Phys. Soc. Japan **85**, 043501 (2016).

[209] C. Zheng, T. C. Petersen, H. Kirmse, W. Neumann, M. J. Morgan, and J. Etheridge, *Axicon Lens for Electrons Using a Magnetic Vortex: The Efficient Generation of a Bessel Beam*, Phys. Rev. Lett. **119**, 174801 (2017).

[210] P. A. Midgley and R. E. Dunin-Borkowski, *Electron Tomography and Holography in Materials Science*, Nat. Mater. **8**, 271 (2009).

[211] G. Guzzinati, W. Ghielens, C. Mahr, A. Béché, A. Rosenauer, T. Calders, and J. Verbeeck, *Electron Bessel Beam Diffraction for Precise and Accurate Nanoscale Strain Mapping*, Appl. Phys. Lett. **114**, 243501 (2019).

[212] S. Hettler, L. Grünewald, and M. Malac, *Quasi Non-Diffractive Electron Bessel Beams Using Direct Phase Masks with Applications in Electron Microscopy*, New J. Phys. **21**, 033007 (2019).

[213] V. Grillo, A. H. Tavabi, E. Yucelen, P.-H. Lu, F. Venturi, H. Larocque, L. Jin, A. Savenko, G. C. Gazzadi, and R. Balboni, *Towards a Holographic Approach to Spherical Aberration Correction in Scanning Transmission Electron Microscopy*, Opt. Express **25**, 21851 (2017).

[214] M. Linck, P. A. Ercius, J. S. Pierce, and B. J. McMorran, *Aberration Corrected STEM by Means of Diffraction Gratings*, Ultramicroscopy **182**, 36 (2017).

[215] R. Shiloh, R. Remez, P.-H. Lu, L. Jin, Y. Lereah, A. H. Tavabi, R. E. Dunin-Borkowski, and A. Arie, *Spherical Aberration Correction in a Scanning Transmission Electron Microscope Using a Sculpted*

Thin Film, Ultramicroscopy **189**, 46 (2018).

[216] J. Verbeeck, H. Tian, and A. Béché, *A New Way of Producing Electron Vortex Probes for STEM*, Ultramicroscopy **113**, 83 (2012).

[217] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, *Electron Ptychography of 2D Materials to Deep Sub-Ångström Resolution*, Nature **559**, 343 (2018).

[218] R. Courtland, *The Microscope Revolution That's Sweeping through Materials Science*, Nature **563**, 462 (2018).

[219] L. Jin, J. Barthel, C.-L. Jia, and K. W. Urban, *Atomic Resolution Imaging of YAlO3: Ce in the Chromatic and Spherical Aberration Corrected PICO Electron Microscope*, Ultramicroscopy **176**, 99 (2017).

[220] E. Yücelen, I. Lazić, and E. G. T. Bosch, *Phase Contrast Scanning Transmission Electron Microscopy Imaging of Light and Heavy Atoms at the Limit of Contrast and Resolution*, Sci. Rep. **8**, 2676 (2018).

[221] S. de Graaf, J. Momand, C. Mitterbauer, S. Lazar, and B. J. Kooi, *Resolving Hydrogen Atoms at Metal-Metal Hydride Interfaces*, Sci. Adv. **6**, eaay4312 (2020).

[222] S. Lopatin, B. Cheng, W.-T. Liu, M.-L. Tsai, J.-H. He, and A. Chuvilin, *Optimization of Monochromated TEM for Ultimate Resolution Imaging and Ultrahigh Resolution Electron Energy Loss Spectroscopy*, Ultramicroscopy **184**, 109 (2018).

[223] D. Rossouw and G. A. Botton, *Plasmonic Response of Bent Silver Nanowires for Nanophotonic Subwavelength Waveguiding*, Phys. Rev. Lett. **110**, 066801 (2013).

[224] R. Henderson, *The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules*, Q. Rev. Biophys. **28**, 171 (1995).

[225] R. Henderson, *Avoiding the Pitfalls of Single Particle Cryo-Electron Microscopy: Einstein from Noise*, Proc. Natl. Acad. Sci. **110**, 18037 (2013).

[226] H. H. Rose, *Historical Aspects of Aberration Correction*, J. Electron Microsc. (Tokyo). **58**, 77 (2009).

[227] N. Voloch-Bloch, Y. Lereah, Y. Lilach, A. Gover, and A. Arie, *Generation of Electron Airy Beams*, Nature **494**, 331 (2013).

[228] V. Grillo, E. Karimi, G. C. Gazzadi, S. Frabboni, M. R. Dennis, and R. W. Boyd, *Generation of Nondiffracting Electron Bessel Beams*, Phys. Rev. X **4**, 011013 (2014).

[229] G. Guzzinati, A. Béché, D. McGrouther, and J. Verbeeck, *Prospects for Out-of-Plane Magnetic Field Measurements through Interference of Electron Vortex Modes in the TEM*, J. Opt. **21**, 124002 (2019).

[230] 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. **8**, 5592 (2018).

[231] F. S. Hage, R. J. Nicholls, J. R. Yates, D. G. McCulloch, T. C. Lovejoy, N. Dellby, O. L. Krivanek, K. Refson, and Q. M. Ramasse, *Nanoscale Momentum-Resolved Vibrational Spectroscopy*, Sci. Adv. **4**, eaar7495 (2018).

[232] R. Senga, K. Suenaga, P. Barone, S. Morishita, F. Mauri, and T. Pichler, *Position and Momentum Mapping of Vibrations in Graphene Nanostructures*, Nature **573**, 247 (2019).

[233] C. A. Guido, E. Rotunno, M. Zanfrognini, S. Corni, and V. Grillo, *Exploring the Azimuthal Symmetries of Electronic Transitions in Molecular and Biomolecular Systems by Swift Electrons*, (2020).

[234] G. Bertoni, E. Rotunno, A. Tavabi, M. Zanfrognini, P. Rosi, S. Frabboni, E. Karimi, R. Dunin-Borkowski,

and V. Grillo, *Combination of Electron Energy-Loss Spectroscopy and Orbital Angular Momentum Spectroscopy. Applications to Electron Magnetic Chiral Dichroism, Plasmon-Loss, and Core-Loss*, Microsc. Microanal. **26**, 1752 (2020).

[235] K. X. Nguyen, Y. Jiang, M. C. Cao, P. Purohit, A. K. Yadav, P. García-Fernández, M. W. Tate, C. S. Chang, P. Aguado-Puente, J. Íñiguez, F. Gomez-Ortiz, S. M. Gruner, J. Junquera, L. W. Martin, R. Ramesh, and D. A. Muller, *Transferring Orbital Angular Momentum to an Electron Beam Reveals Toroidal and Chiral Order*, (2020).

[236] A. Lubk, G. Guzzinati, F. Börrnert, and J. Verbeeck, *Transport of Intensity Phase Retrieval of Arbitrary Wave Fields Including Vortices*, Phys. Rev. Lett. **111**, 173902 (2013).

[237] R. Shiloh, Y. Tsur, R. Remez, Y. Lereah, B. A. Malomed, V. Shvedov, C. Hnatovsky, W. Krolikowski, and A. Arie, *Unveiling the Orbital Angular Momentum and Acceleration of Electron Beams*, Phys. Rev. Lett. **114**, 096102 (2015).

[238] T. R. Harvey, V. Grillo, and B. J. McMorran, *Stern-Gerlach-like Approach to Electron Orbital Angular Momentum Measurement*, Phys. Rev. A **95**, 021801 (2017).

[239] Y. Noguchi, S. Nakayama, T. Ishida, K. Saitoh, and M. Uchida, *Efficient Measurement of the Orbital-Angular-Momentum Spectrum of an Electron Beam via a Dammann Vortex Grating*, Phys. Rev. Appl. **12**, 064062 (2019).

[240] G. Ruffato, E. Rotunno, and V. Grillo, *A General Framework for Conformal Transformations in Electron Optics*, ArXiv (2020).

[241] J. Y. Wang, *Optical Resolution through a Turbulent Medium with Adaptive Phase Compensations**, J. Opt. Soc. Am. **67**, 383 (1977).

[242] M. Haider, H. Müller, S. Uhlemann, J. Zach, U. Loebau, and R. Hoeschen, *Prerequisites for a Cc/Cs-Corrected Ultrahigh-Resolution TEM*, Ultramicroscopy **108**, 167 (2008).

[243] B. Kabius, P. Hartel, M. Haider, H. Muller, S. Uhlemann, U. Loebau, J. Zach, and H. Rose, *First Application of Cc-Corrected Imaging for High-Resolution and Energy-Filtered TEM*, J. Electron Microsc. (Tokyo). **58**, 147 (2009).

[244] P. C. Tiemeijer, *Section 5: Analytical Electron Microscopy-Operation Modes of a TEM Monochromator*, in *Institute of Physics Conference Series*, Vol. 161 (Bristol [England]; Boston: Adam Hilger, Ltd., c1985-, 1999), pp. 191–194.

[245] O. L. Krivanek, T. C. Lovejoy, N. Dellby, T. Aoki, R. W. Carpenter, P. Rez, E. Soignard, J. Zhu, P. E. Batson, M. J. Lagos, R. F. Egerton, and P. A. Crozier, *Vibrational Spectroscopy in the Electron Microscope*, Nature **514**, 209 (2014).

[246] S. Morishita, M. Mukai, K. Suenaga, and H. Sawada, *Resolution Enhancement in Transmission Electron Microscopy with 60-KV Monochromated Electron Source*, Appl. Phys. Lett. **108**, 013107 (2016).

[247] G. Guzzinati, P. Schattschneider, K. Y. Bliokh, F. Nori, and J. Verbeeck, *Observation of the Larmor and Gouy Rotations with Electron Vortex Beams*, Phys. Rev. Lett. **110**, 093601 (2013).

[248] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann, *Coma-Free Alignment of High Resolution Electron Microscopes with the Aid of Optical Diffractograms*, Ultramicroscopy **3**, 49 (1978).

[249] N. Dellby, O. L. Krivanek, P. D. Nellist, P. E. Batson, and A. R. Lupini, *Progress in Aberration-Corrected Scanning Transmission Electron Microscopy*, Microscopy **50**, 177 (2001).

[250] H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka,

Y. Oshima, Y. Tanishiro, and K. Takayanagi, *Measurement Method of Aberration from Ronchigram by Autocorrelation Function*, Ultramicroscopy **108**, 1467 (2008).

[251] A. R. Lupini, P. Wang, P. D. Nellist, A. I. Kirkland, and S. J. Pennycook, *Aberration Measurement Using the Ronchigram Contrast Transfer Function*, Ultramicroscopy **110**, 891 (2010).

[252] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet Classification with Deep Convolutional Neural Networks*, Adv. Neural Inf. Process. Syst. **25**, 1097 (2012).

[253] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, (2014).

[254] Y. Jin, Y. Zhang, L. Hu, H. Huang, Q. Xu, X. Zhu, L. Huang, Y. Zheng, H.-L. Shen, W. Gong, and K. Si, *Machine Learning Guided Rapid Focusing with Sensor-Less Aberration Corrections*, Opt. Express **26**, 30162 (2018).

[255] L. Möckl, P. N. Petrov, and W. E. Moerner, *Accurate Phase Retrieval of Complex 3D Point Spread Functions with Deep Residual Neural Networks*, Appl. Phys. Lett. **115**, 251106 (2019).

[256] H. Guo, Y. Xu, Q. Li, S. Du, D. He, Q. Wang, and Y. Huang, *Improved Machine Learning Approach for Wavefront Sensing*, Sensors **19**, 3533 (2019).

[257] Y. Xu, D. He, Q. Wang, H. Guo, Q. Li, Z. Xie, and Y. Huang, *An Improved Method of Measuring Wavefront Aberration Based on Image with Machine Learning in Free Space Optical Communication*, Sensors **19**, 3665 (2019).

[258] B. P. Cumming and M. Gu, *Direct Determination of Aberration Functions in Microscopy by an Artificial Neural Network*, Opt. Express **28**, 14511 (2020).

[259] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, *Deep Learning Microscopy*, Optica **4**, 1437 (2017).

[260] M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, *Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy*, Nat. Methods **15**, 1090 (2018).

[261] F. Troiani, E. Rotunno, S. Frabboni, R. B. G. Ravelli, P. J. Peters, E. Karimi, and V. Grillo, *Efficient Molecule Discrimination in Electron Microscopy through an Optimized Orbital Angular Momentum Sorter*, Phys. Rev. A **102**, 043510 (2020).

[262] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (2014).

[263] A. Gulli and S. Pal, *Deep Learning with Keras* (Packt Publishing Ltd, 2017).

[264] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin, *Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, ArXiv Prepr. ArXiv1603.04467 (2016).

[265] V. Grillo, G. C. Gazzadi, E. Mafakheri, S. Frabboni, E. Karimi, and R. W. Boyd, *Holographic Generation of Highly Twisted Electron Beams*, Phys. Rev. Lett. **114**, 034801 (2015).

[266] H. Qassim, F. M. Miatto, J. P. Torres, M. J. Padgett, E. Karimi, and R. W. Boyd, *Limitations to the Determination of a Laguerre–Gauss Spectrum via Projective, Phase-Flattening Measurement*, J. Opt. Soc. Am. B **31**, A20 (2014).

[267] M. P. J. Lavery, D. J. Robertson, G. C. G. Berkhout, G. D. Love, M. J. Padgett, and J. Courtial, *Refractive Elements for the Measurement of the Orbital Angular Momentum of a Single Photon*, Opt. Express **20**, 2110 (2012).

[268] V. Grillo, L. Marrucci, E. Karimi, R. Zanella, and E. Santamato, *Quantum Simulation of a Spin Polarization Device in an Electron Microscope*, New J. Phys. **15**, 093026 (2013).

[269] N. K. Fontaine, R. Ryf, H. Chen, D. T. Neilson, K. Kim, and J. Carpenter, *Laguerre-Gaussian Mode Sorter*, Nat. Commun. **10**, 1865 (2019).

[270] W. T. Scott and T. C. Hardy, *The Physics of Electricity and Magnetism*, J. Electrochem. Soc. **106**, 271C (1959).

[271] E. Durand, *Electrostatique*, (1964).

[272] M. Beleggia, P. F. Fazzini, and G. Pozzi, *A Fourier Approach to Fields and Electron Optical Phase-Shifts Calculations*, Ultramicroscopy **96**, 93 (2003).

[273] V. Mathematica, *4, Stephen Wolfram, The Mathematica Book, Wolfram Media*.

[274] C. Shukla and A. Das, *Ion Vortex Beam*, (2017).

[275] K. Nagayama, *Another 60 Years in Electron Microscopy: Development of Phase-Plate Electron Microscopy and Biological Applications*, Microscopy **60**, S43 (2011).

[276] R. M. Glaeser, *Invited Review Article: Methods for Imaging Weak-Phase Objects in Electron Microscopy*, Rev. Sci. Instrum. **84**, 111101 (2013).

[277] R. Cambie, K. H. Downing, D. Typke, R. M. Glaeser, and J. Jin, *Design of a Microfabricated, Two-Electrode Phase-Contrast Element Suitable for Electron Microscopy*, Ultramicroscopy **107**, 329 (2007).

[278] J. M. Perry-Houts, B. Barton, A. K. Schmid, N. Andresen, and C. Kisielowski, *Novel Long-Lived Electrostatic Work Function Phase Plates for TEM*, Microsc. Microanal. **18**, 476 (2012).

[279] H. Tamaki, H. Kasai, K. Harada, Y. Takahashi, and R. Nishi, *Development of a Contact-Potential-Type Phase Plate*, Microsc. Microanal. **19**, 1148 (2013).

[280] Y. Aharonov and D. Bohm, *Significance of Electromagnetic Potentials in the Quantum Theory*, Phys. Rev. **115**, 485 (1959).

[281] Y. Aharonov, *Non-Local Phenomena and the Aharonov-Bohm Effect*, in *Proceedings of the International Symposium Foundations of Quantum Mechanics in the Light of New Technology* (1984).

[282] S. Olariu and I. I. Popescu, *The Quantum Effects of Electromagnetic Fluxes*, Rev. Mod. Phys. **57**, 339 (1985).

[283] T. H. Boyer, *Classical Electromagnetic Deflections and Lag Effects Associated with Quantum Interference Pattern Shifts: Considerations Related to the Aharonov-Bohm Effect*, Phys. Rev. D **8**, 1679 (1973).

[284] T. H. Boyer, *Semiclassical Explanation of the Matteucci–Pozzi and Aharonov–Bohm Phase Shifts*, Found. Phys. **32**, 41 (2002).

[285] G. Matteucci, G. F. Missiroli, and G. Pozzi, *A New Electrostatic Phase-Shifting Effect*, Ultramicroscopy **10**, 247 (1982).

[286] G. Matteucci and G. Pozzi, *New Diffraction Experiment on the Electrostatic Aharonov-Bohm Effect*, Phys. Rev. Lett. **54**, 2469 (1985).

[287] G. Matteucci, F. F. Medina, and G. Pozzi, *Electron-Optical Analysis of the Electrostatic Aharonov-Bohm Effect*, Ultramicroscopy **41**, 255 (1992).

[288] E. Ball, *Potential from a Ring of Charge*, Proc. Inst. Electr. Eng. **124**, 664 (1977).

[289] F. R. Zypman, *Off-Axis Electric Field of a Ring of Charge*, Am. J. Phys. **74**, 295 (2006).

[290] O. Ciftja, A. Babineaux, and N. Hafeez, *The Electrostatic Potential of a Uniformly Charged Ring*, Eur. J. Phys. **30**, 623 (2009).

[291] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists 6th Ed.*, Math. Methods Phys. 6th Ed. by Georg. B. Arfken Hans J. Weber. Publ. Amsterdam; Bost. Elsevier (2005).

[292] M. Beleggia, T. Kasama, R. E. Dunin-Borkowski, S. Hofmann, and G. Pozzi, *Direct Measurement of the Charge Distribution along a Biased Carbon Nanotube Bundle Using Electron Holography*, Appl. Phys. Lett. **98**, 243101 (2011).

[293] C. Gatel, A. Lubk, G. Pozzi, E. Snoeck, and M. Hÿtch, *Counting Elementary Charges on Nanoparticles by Electron Holography*, Phys. Rev. Lett. **111**, 025501 (2013).

[294] M. Beleggia, L. C. Gontard, and R. E. Dunin-Borkowski, *Local Charge Measurement Using Off-Axis Electron Holography*, J. Phys. D. Appl. Phys. **49**, 294003 (2016).

[295] H. B. Michaelson, *The Work Function of the Elements and Its Periodicity*, J. Appl. Phys. **48**, 4729 (1977).

[296] J. R. Rumble, D. R. Lide, and T. J. Bruno, *CRC Handbook of Chemistry and Physics [2019-2020] : A Ready-Reference Book of Chemical and Physical Data*.

[297] G. Pozzi, M. Beleggia, T. Kasama, and R. E. Dunin-Borkowski, *Interferometric Methods for Mapping Static Electric and Magnetic Fields*, Comptes Rendus Phys. **15**, 126 (2014).

[298] S. Fürhapter, A. Jesacher, C. Maurer, S. Bernet, and M. Ritsch-Marte, *Spiral Phase Microscopy*, in (2007), pp. 1-59e.

[299] R. Juchtmans, L. Clark, A. Lubk, and J. Verbeeck, *Spiral Phase Plate Contrast in Optical and Electron Microscopy*, Phys. Rev. A **94**, 023838 (2016).

[300] R. Juchtmans and J. Verbeeck, *Local Orbital Angular Momentum Revealed by Spiral-Phase-Plate Imaging in Transmission-Electron Microscopy*, Phys. Rev. A **93**, 023811 (2016).

[301] R. Danev, R. M. Glaeser, and K. Nagayama, *Practical Factors Affecting the Performance of a Thin-Film Phase Plate for Transmission Electron Microscopy*, Ultramicroscopy **109**, 312 (2009).