Thesis to obtain the degree Master of Science in Physics
Arbeit zur Erlangung des Grades Master of Science in Physik

Jonas Krehl

Incorporating Fresnel-Propagation into Electron Holographic Tomography

A possible way towards three-dimensional atomic resolution

presented on / eingereicht am

4th April 2016

to the / zu der

Department of Physics
Fachrichtung Physik

Faculty of Science and Mathematics
Fakultät für Naturwissenschaften und Mathematik

Technische Universität Dresden

1st referee / Erstgutachter : Prof. Hannes Lichte (TU Dresden)
2nd referee / Zweitgutachter : Prof. Christian Schroer (Universität Hamburg)

Ein Rationalist ist einfach ein Mensch, dem mehr daran liegt zu lernen, als recht zu behalten; der bereit ist , von anderen zu lernen, nicht etwa dadurch, daß er die fremde Meinung einfach annimmt, sondern dadurch, daß er gerne seine Ideen von anderen kritisieren lässt und gerne die Ideen anderer kritisiert.

Karl R. Popper

Abstract

Tomographic electron holography combines tomography, the reconstruction of three-dimensionally resolved data from multiple measurements with different specimen orientations, with electron holography, an interferometrical method for measuring the complex wave function inside a transmission electron microscope (TEM). Due to multiple scattering and free wave propagation conventional, ray projection based, tomography does perform badly when approaching atomic resolution. This is remedied by incorporating propagation effects into the projection while maintaining linearity in the object potential. Using the Rytov approach an approximation is derived, where the logarithm of the complex wave is linear in the potential. The ray projection becomes a convolution with a Fresnel propagation kernel, which is considerably more computationally expensive. A framework for such calculations has been implemented in Python. So has a multislice electron scattering algorithm, optimised for large fields of view and high numbers of atoms for simulations of scattering at nanoparticles. The Rytov approximation gives a remarkable increase in resolution and signal quality over the conventional approach in the tested system of a tungsten disulfide nanotube. The response to noise seems to be similar as in conventional tomography, so rather benign. This comes at the downside of much longer calculation time per iteration.

Zusammenfassung

Tomographische Elektronenholographie kombiniert Tomographie, die Rekonstruktion dreidimensional aufgelößter Daten aus einem Satz von mehreren Messungen bei verschiedenen Objektorientierungen, mit Elektronenholographie, eine interferrometrische Messung der komplexen Elektronenwelle im Transmissionselektronenmikroskop (TEM). Wegen Mehrfachstreuung und Propagationseffekten erzeugt konventionelle, auf einer Strahlprojektion basierende, Tomography ernste Probleme bei Hochauflösung hin zu atomarer Auflösung. Diese sollen durch ein Modell, welches Fresnel-Propagation beinhaltet, aber weiterhin linear im Potential des Objektes ist, vermindert werden. Mit dem Rytov-Ansatz wird eine Näherung abgeleitet, wobei der Logarithmus der komplexen Welle linear im Potential ist. Die Strahlen-Projektion ist dann eine Faltung mit dem Fresnel-Propagations-Faltungskernel welche rechentechnisch wesentlich aufwendiger ist. Ein Programm-Paket für solche Rechnungen wurde in Python implementiert. Weiterhin wurde ein Multislice Algorithmus für große Gesichtsfelder und Objekte mit vielen Atomen wie Nanopartikel optimiert. Die Rytov-Näherung verbessert sowohl die Auflösung als auch die Signalqualität immens gegenüber konventioneller Tomographie, zumindest in dem getesteten System eines Wolframdisulfid-Nanoröhrchens. Das Rauschverhalten scheint ähnlich der konventionallen Tomographie zu sein, also eher gutmütig. Im Gegenzug braucht die Tomographie basierend auf der Rytov-Näherung wesentlich mehr Rechenzeit pro Iteration.

Contents

Contents

1. Introduction

The interest in the research of nanoparticles of all flavours and sources is currently on the rise. The time may be over where finding nanoparticles themselves was a sensation in themselves but now the true breadth and capabilities of the field are being explored. With growing ability to synthesize nanoparticles of specified structure, shape, size and composition, in order to achieve more precisely defined and exotic properties, comes a need for better analytical tools to determine whether or not the set goal has been achieved. In the reverse direction do better analytical tools also lead to a better understanding of complicated properties by revealing the underlying object more directly. Transmission electron microscopy will play an important role in this development beacuse of its (in optical microscopy) unparalleled spatial resolution, reaching down to the atomic level. It is, however, currently not possible to extend the capability of atomic resolution into the three-dimensional so the position and type of each atom in a complex nanostructure could be determined.

In this work one route towards this goal is developed and tested preliminarily. The already established combination of electron holography and tomography[1][2] has shown proven its ability to reconstruct (apart from other quantities) the electric and magnetic potentials of a specimen but when approaching high (i.e. atomic) resolution the commonly assumed model used in the tomographic reconstruction has failed. The phase grating approximation is a blunt approximation, which is incapable to model wave propagation and, by extend, a lot of the multiple scattering effects as well. A new and better approximation will be developed, which incorporates wave propagation but still allows the use of linear tomography for reconstruction, because there efficient reconstruction techniques are readily available. Nonetheless, computational complexity of this new approach will be considerably different and higher than that of conventional tomography so a dedicated implementation of the reconstruction will have to be developed. This will then be tested at a simulated holographic tilt series of a realistic sample system (a tungsten disulfide multiwall nanotube).

For scattering simulation of large non-periodic samples several optimisations are added to the established electron scattering simulation algorithm, the multislice algorithm, and implemented. This large field-of-view multislice algorithm will provide the data for the reconstruction tests, that shall show, whether or nor the newly developed approximation gives a real improvement in reconstruction quality.

1.1. Electron Microscopy

For microscopy three things are needed: a source, a detector and a focussing optic. All of these things were available at the beginning of the 20th century when the young engineer Ernst Ruska started his work in the group of Ernst Knoll trying to develop a high-speed oscilloscope with a focussed electron beam[3]. Against contrary assignments he designed a strong and thin magnetic lens which realised the lens theory of Hans Busch and built a setup for the characterisation of the optical qualities of his lenses. Intrigued by the passable imaging characteristics of his lens he added a second one and

---

[1][36] Wolf et al. 2013 . "Electron holographic tomography".
[2][20] Midgley and Dunin-Borkowski. 2009 . "Electron tomography and holography in materials science".
[3]this historic account is based on [27] Ruska. 1987 . "Das Entstehen des Elektronenmikroskops und der Elektronenmikroskopie".

made this apparatus in 1931 the first transmission electron microscope. Slightly disheartened at the implication of de Broglie's work that an electron microscope is still diffraction limited he was greatly relieved that his electrons, with $75\,\mathrm{keV}$, would still allow a resolution of $2.2\,\text{Å}$ (using his aperture of $20\,\mathrm{mrad}$) many times better than any optical microscope could hope for. It took nevertheless about 40 years for this value to be reached due to the persistence of the spherical aberration (C_S) which is inherent in every rotationally symmetric, time-invariant and space-charge-free electric or magnetic lens as Scherzer proved in 1936[4]. This established itself as the eternal problem of TEM to the point that the hope and funding for the development of an aberration-free microscope was lost. Only the advent of computers made the precise control of non-rotationally symmetric lenses practical and in 1998[5] a working C_S-corrected microscope was demonstrated. This vastly improved both the resolution of electron microscopes and the signal transfer to the detector. While resolution is certainly a very important metric, TEM has developed a huge range of different optical setups (most of which are poached from light optics) and signals, which allow the analysis of numerous specimen characteristics (e.g. morphology, atomic structure or intrinsic fields) by their interaction with an electron beam or by-products of that interaction (e.g. secondary electrons or X-rays).

In the scope of high-resolution microscopy the search for ever better resolution and better signal transfer is ongoing and the frontier currently sits at the image spread, where small thermal-fluctuation-induced currents in the liner randomly deflect the beam which accumulates to a blurring on the detector[6]. All other aberrations are corrected to a sufficient degree and this includes most (isoplanatic and non-isoplanatic) geometric aberrations as well as chromatic aberrations. There are, however, only two achromatic TEMs in the world today so they are still a niche product while C_S-corrected microscopes are the standard today.

1.2. Electron Holography

Holography describes any technique capable of retrieving the complex-valued wave function from one or more detected images. This may be achieved by different methods, such as Focal-Series Holography, where a set of images is taken at different defocus values and a method similar to Kaczmarz' is used for the reconstruction[7]. Focal-Series Holography, however, suffers from ambiguity and instability in the reconstruction as well as poor information transfer of the lower spatial frequencies[8]. Similar problems plague other iterative reconstruction techniques, like Ptychography. But these problems are steadily explored and treated, so at some point these methods will be able to reliably reconstruct the wave function in every scenario, maybe at the cost of a difficult to measure dataset.

Off-Axis Holography offers a different approach: an electron-optical biprism is used to overlay two regions of the beam with a certain angle between them[9]. Given sufficient coherence between the two partial beams, they produce an interference pattern of fringes whose frequency is determined by the angle between the partial beams. Given, that one of the beams is sufficiently well known (most often it is simply a plane wave), the other one can be reconstructed from the interference pattern. This procedure gives the amplitude and phase exclusively by linear operations on the recorded data; the result is therefore unique and stable, which also makes for a well-understood noise-transfer. This method requires the addition of a biprism to the microscope, which is usually done by exchanging an aperture holder for a special one.

---

[4] [29] Scherzer. 1936 . "Über einige Fehler von Elektronenlinsen".
[5] [12] Haider et al. 1998 . "A spherical-aberration-corrected 200 kV transmission electron microscope".
[6] [31] Uhlemann et al. 2013 . "Thermal Magnetic Field Noise Limits Resolution in Transmission Electron Microscopy".
[7] [9] Gerchberg and Saxton. 1972 . "A practical algorithm for the determination of phase from image and diffraction plane pictures".
[8] [8] Fienup and Wackerman. 1986 . "Phase-retrieval stagnation problems and solutions".
[9] [21] Möllenstedt and Wahl. 1968 . "Elektronenholographie und Rekonstruktion mit Laserlicht".

For the question of this thesis, the wave function is the starting point for the tomographic reconstruction algorithm; the exact origin of the wave function is largely irrelevant. It has, however, some relevance in the question how the thermal motion of the atoms appears in the averaged wave function where we will use the mechanism from off-axis holography[10] (also see Section 2.4).

1.3. Electron Tomography

Originally, tomography was a name applied to a technique, developed by G. N. Hounsfield[11] and A. M. Cormack[12] in the 60s, for reconstructing the local attenuation coefficient from the attenuation of X-rays, modelled as the integral of the local attenuation along the rays path through the sample. Over time it has shown itself that the technique and its theoretical framework are much more extensive and tomography now refers to a vast field of mathematical models and a multitude of signals, probes, geometries, … for which these models are applicable. The basic principle is the reconstruction of an object characteristic, not determinable from a single measurement, from a set of multiple measurements [13]. Therefore, tomography is a so called inverse problem: the model for the mapping of the object characteristics to the measurable signal is given but its inversion is not, or does not even exist in any strict sense.

Many of the multitude of signals from TEM have been used for tomographic reconstruction of object characteristics so trivial as the outer shape to the highly sophisticated complete quantum state of the electron wave, and everything in between, such as electric and magnetic potentials, chemical composition or the crystal structure. But, while aberration correction has made atomic resolution a convenience, electron tomography has not yet been able to reliably retrieve the atomic structure of the specimen. Several forays have been made but they rely heavily on the structure of the specimen, such as it being only one atom layer thick[14,15] or being purely crystalline[16,17,18]. Another route is to use novel reconstruction techniques, such as compressive sensing (a representation of the specimen is defined where it should be sparse and the reconstruction optimised for a sparsity in that representation)[19] or non-linear inversion techniques[20]. These forays currently linger at the proof-of-concept stage and considering that they rely in assumptions about the specimen or on algorithms of unproven reliability it is debatable which of these will yield an applicable method. Or rather, considering the individual limitations, difficulties and uncertainties, which will be trustworthy for what kind of specimen.

In that scope the ansatz of this work is a rather conservative one, by replacing the ray projection in the conventional tomography with a Fresnel propagation kernel. This is similar to diffraction tomography in light, ultrasound and X-ray tomography with semi-transparent objects[21] although the specimen is not weakly scattering here. As this is a linear technique it will be unable to model multiple scattering adequately which, ultimately, will give a limit to the thickness of the specimen that can be reconstructed without significant artefacts.

---

[10][26] Rother et al. 2009 . "The statistics of the thermal motion of the atoms during imaging process in transmission electron microscopy and r

[11][14] Hounsfield. 1973 . "Computerized transverse axial scanning (tomography): Part 1. Description of system".

[12][6] Cormack. 1963 . "Representation of a Function by Its Line Integrals, with Some Radiological Applications".

[13]The terminus tomography itself is defined by convention: Those techniques, which are commonly called tomographic, are tomographic. Over the time no encompassing but sharp definition has been proposed and debated.

[14][33] Van Dyck et al. 2012 . "'Big Bang' tomography as a new route to atomic-resolution electron tomography".

[15][1] Bar Sadan et al. 2008 . "[...] Bright-Field Electron Tomography [...] of Fullerene-like Nanostructures".

[16][5] Chen et al. 2015 . "3D reconstruction of nanocrystalline particles from a single projection".

[17][11] Goris et al. 2012 . "Atomic-scale determination of surface facets in gold nanorods".

[18][32] Van Aert et al. 2011 . "Three-dimensional atomic imaging of crystalline nanoparticles".

[19][2] Broek et al. 2009 . "A model based atomic resolution tomographic algorithm".

[20]such as artificial neural networks in [3] Broek and Koch. 2012 . "[...] Inversion of Dynamical Electron Scattering".

[21][37] Wolf. 1969 . "Three-dimensional structure determination of semi-transparent objects from holographic data".

2. Electron Scattering

In most applications of transmission electron microscopy (TEM) some approximations can be made for the electrons at hand. First of which is the neglection of spin which allows the use of the Klein-Gordon equation to describe the electrons motion and its interaction via the introduction of minimal coupling to the electric potential (the magnetic potential shall be neglected as the atomic structure is wholly electric). Assuming the electrons energy is large compared to the potential $e\phi \ll E$ which holds everywhere except at an insignificant small area at the atoms' cores a relativistic correction term is neglected and this perturbed Helmholtz equation [1] follows:

$$\left(\Delta + k^2\right)\Psi(\mathbf{r}) = \underbrace{\frac{2Ee\phi(\mathbf{r})}{\hbar^2 c^2}}_{=:-V(\mathbf{r})}\Psi(\mathbf{r}) \qquad \text{perturbed Helmholtz equation} \qquad (2.1)$$

The used symbols are:

$$\Delta = \text{Laplace operator} \qquad\qquad k = \text{angular wavenumber}$$
$$\Psi = \text{wave function} \qquad\qquad e = \text{unit charge}$$
$$E = \text{total energy of electron} \qquad\qquad \phi = \text{electrostatic potential}$$
$$V = \text{effective potential}$$

Since the perturbation V is just the potential with a constant prefactor it will be called effective potential in spite of the fact that the physical meaning of the perturbation is an "interaction effect" on the wave which is dimensionless and not energy per charge like a potential.

For imaging in a TEM it is usually sufficient to solve the problem in a paraxial approximation, where the angular extension of the beam is small all along the column, therefore, angles of individual rays or phase-gradients are small as well. The chosen axis is z and the lateral component is called ρ. The Laplace operator is split accordingly in an axial and a lateral component $\Delta = \Delta_\rho + \partial_z^2$. Furthermore, the rapidly oscillating phase is separated from the wave function $\Psi(\mathbf{r}) = \psi(\mathbf{r})e^{ikz}$ and subsequently eliminated (only the ∂_z^2-term acts on it) yielding:

$$\left(\Delta_\rho + 2ik\partial_z + \partial_z^2\right)\psi(\mathbf{r}) = -V(\mathbf{r})\psi(\mathbf{r}) \qquad\qquad (2.2)$$

Since the scattering angles are small, the changes of the wave function along z are very small compared to the rapid phase oscillation: $|\partial_z\psi(\mathbf{r})| \ll k$ and the ∂_z^2-term can be neglected:

$$\left(\Delta_\rho + 2ik\partial_z\right)\psi(\mathbf{r}) = -V(\mathbf{r})\psi(\mathbf{r}) \qquad \text{perturbed paraxial Helmholtz equation} \qquad (2.3)$$

The complete neglection of the lateral term in Eq. (2.3) then gives the most simple approximation which shall be called axial approximation here.

$$2ik\partial_z\psi(\mathbf{r}) = -V(\mathbf{r})\psi(\mathbf{r}) \qquad \text{perturbed axial Helmholtz equation} \qquad (2.4)$$

---

[1]The name is chosen to distinguish this equation from an inhomogeneous Helmholtz equation since the perturbation is very different to an inhomogenity.

These three equations (Eq. (2.1), Eq. (2.3) and Eq. (2.4)) are solved in the same manner by transferring the problem to a Lippmann-Schwinger equation (Ψ stands for ψ in the case of the paraxial and the axial approximation):

$$\Psi(\mathbf{r}) = \psi(\mathbf{r}) + \int d\mathbf{r}'\, G(\mathbf{r} - \mathbf{r}')V(\mathbf{r}')\Psi(\mathbf{r}') \qquad \text{Lippmann-Schwinger equation} \qquad (2.5)$$

Where $\psi(\mathbf{r})$ solves the unperturbed equation $\hat{D}\Psi(\mathbf{r}) = 0$ under the given boundary conditions, and the Green's Operator $G(\mathbf{r} - \mathbf{r}')$ solves the unperturbed problem in a distributive sense:

$$\hat{D}G(\mathbf{r} - \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \qquad (2.6)$$

The Green's operator can be used to solve this elliptical differential equation as a boundary condition problem, where $\Psi(\mathbf{r}) = u(\mathbf{r})$ is given on the boundary ∂M of a connected region M:

$$\Psi(\mathbf{r}) = \int_{\partial M} d\mathbf{r}'\, G(\mathbf{r} - \mathbf{r}')V(\mathbf{r}')u(\mathbf{r}') \qquad (2.7)$$

Any standing wave solutions (those solving $\Psi(\mathbf{r}) = 0$ on the boundary ∂G) are neglected so the solution $\Psi(\mathbf{r})$ solves the boundary condition problem uniquely.

Although the Lippmann-Schwinger equation 2.0Eq. (2.5) has a solution in the strict sense, it cannot be evaluated for all but the simplest problems. As such, it is not feasible to solve it completely for the here relevant use-cases and it is necessary to resort to further approximations (see Section 2.1 and Section 2.2).

The Green's functions (also called propagators), the solutions of Eq. (2.6), for the different approximations can be expressed analytically as:

$$G(\mathbf{r}) = \frac{e^{ik|\mathbf{r}|}}{4\pi|\mathbf{r}|} \qquad \text{Huygens-Fresnel propagator} \qquad (2.8)$$

$$G_p(z, \rho) = \frac{-ik}{2\pi z}e^{i\frac{k|\rho|^2}{2z}}\Theta(z) \qquad \text{Fresnel propagator} \qquad (2.9)$$

$$G_a(z, \rho) = \frac{-i}{2k}\delta(\rho)\Theta(z) \qquad \text{axial propagator} \qquad (2.10)$$

The Fresnel and the axial propagator contain a Heaviside Theta-function, which imposes forward-scattering. In the reconstruction the reference plane is midway through the specimen; below this plane the propagator is mirrored to the purely backward-scattering one.

As the geometry of the problems is similar throughout imaging simulations it should be clarified beforehand. The beam travels along z (also called height) from the given initial wave $\Psi_i(\rho) = \Psi(z_i, \rho)$ at height z_i to the final wave $\Psi_f(\rho) = \Psi(z_f, \rho)$ at height z_f. Laterally, the problem is infinitely extended (as the beam is much thinner than the inner tube of the microscope, except in some special cases). For a given problem between z_i and z_f, Ψ_i is given and
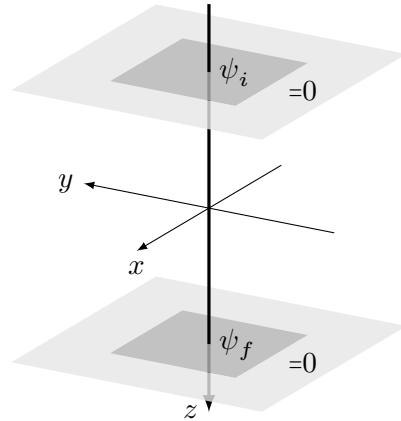

Figure 2.1.: The common scattering geometry, with the incoming wave Ψ_i and the outgoing wave Ψ_f which are zero outside a finite area.

Ψ_f searched for; the boundary conditions are therefore $\Psi_i(\rho)$ in the plane $z = z_i$ and Neumann boundary conditions of $\nabla \Psi(z, \rho) = 0$ for $|\rho| \to \infty$.

2.1. Phase-Grating or Axial Rytov Approximation

A very blunt approximation is the phase-grating approximation (PGA) which is just another name for the axial approximation, where the propagation of the wave is neglected while travelling through the specimen. The perturbed problem can be solved directly in the above mentioned geometry by integration along z:

$$2ik\partial_z\psi(z, \rho) = -V(z, \rho)\psi(z, \rho) \tag{2.11}$$

$$\psi_f(\rho) = \exp\left(\int_{z_i}^{z_f} \mathrm{d}z\, \frac{i}{2k}V(z, \rho)\right)\psi_i(\rho) \tag{2.12}$$

This is the Lambert-Beer law for complex valued functions and can be transferred into an expression for the phase by defining $\Lambda = \log(\psi)$:

$$\Lambda_f(\rho) = \int_{z_i}^{z_f} \mathrm{d}z\, \frac{i}{2k}V(z, \rho)\Lambda_i(\rho) \tag{2.13}$$

With that, it corresponds to the Rytov approximation (Section 2.2) for the axial propagator. The same considerations apply to the meaning of the complex-valued "phase" Λ.

It is a straightforward integration and yields an expression linear in the potential, which makes this approximation suited for tomography. The error is the neglection of any kind of propagation behaviour. This error is on the order of $\int_{z_i}^{z_f} \mathrm{d}z\, \Delta_\rho\psi(z, \rho)$ which, evaluated in lateral Fourier space, gives an upper bound for the error of $\propto |\mathbf{k}_\rho\|^2 |z_f - z_i|$. This indicates that there is a region of small $|\mathbf{k}_\rho\|$ and $|z_f - z_i|$ where the approximation holds. It is not said how big this region is and this simple criterion cannot be applied in any meaningful way, since the structure of the potential (e.g. orientation in a Bragg reflection) has dominating influence on the trust region of this approximation.

Usually the phase-grating only considers the phase of the wave function, i.e. the imaginary component of the Rytov phase, but since this approximation is separable in both real and imaginary component considering the real component does not affect the imaginary one. So we might as well treat the complex problem. That is not to say that an interplay between the two components can be introduced by the reconstruction technique via its regularisation characteristics. The common reconstruction techniques, such as any variant of the Landweber iteration and the conjugate gradient methods, preserve the separability of the projection model and the objective function.

2.2. Rytov Approximation

The ansatz of the, in the field of electron scattering unusual but rather practical, Rytov approximation is to transform the problem via a logarithm into a domain where the real component equates to the logarithm of the amplitude and the imaginary component to the phase of the original problem. Thus, the logarithm Λ (we will call Rytov phase) of the wave function Ψ is defined $\Psi = e^\Lambda$ and inserted

in the perturbed paraxial Helmholtz equation Eq. (2.3) (again, the paraxial approximation can be foregone as it only changes the form of the propagator).

$$\left(\Delta_\rho + 2ik\partial_z\right)e^{\Lambda(\mathbf{r})} = -V(\mathbf{r})e^{\Lambda(\mathbf{r})} \tag{2.14}$$

$$e^{\Lambda(\mathbf{r})}\left(\Delta_\rho\Lambda(\mathbf{r}) + \left(\nabla_\rho\Lambda(\mathbf{r})\right)^2 + 2ik\partial_z\Lambda(\mathbf{r})\right) = -V(\mathbf{r})e^{\Lambda(\mathbf{r})} \tag{2.15}$$

$$\left(\Delta_\rho\Lambda(\mathbf{r}) + \left(\nabla_\rho\Lambda(\mathbf{r})\right)^2 + 2ik\partial_z\Lambda(\mathbf{r})\right) = -V(\mathbf{r}) \tag{2.16}$$

This equation is surprisingly similar to the one for the wave Ψ, only deviating in the existence of the quadratic gradient term and the fact that the perturbation is now an inhomogenity. The perturbed linear differential equation has been transformed to an inhomogeneous quadratic equation, which cannot be solved directly. It is difficult to assess the importance of the gradient-term but it shall be neglected here and discussed for a specific problem in Section 6.4.1. With that, the result is an inhomogeneous linear equation which can be solved directly with a solution Λ_0 for the homogeneous problem in the boundary conditions $\Lambda(\mathbf{r}) = u(\mathbf{r}) : \mathbf{r} \in \partial M$ (M being a connected region) and an explicit solution for the inhomogenity:

$$\left(\Delta_\rho\Lambda(\mathbf{r}) + 2ik\partial_z\Lambda(\mathbf{r})\right) = -V(\mathbf{r}) \tag{2.17}$$

$$\Lambda_{ih}(z,\rho) = \int\limits_M \mathrm{d}z' \, \mathrm{d}\rho' \, G_p(z-z', \rho-\rho')V(z',\rho') \qquad \text{inhomogeneous solution} \tag{2.18}$$

$$\Lambda_h(z,\rho) = \int\limits_{\partial M} \mathrm{d}z' \, \mathrm{d}\rho' \, G_p(z-z', \rho-\rho')\left(u(z',\rho') - \Lambda_{ih}(z',\rho')\right) \qquad \text{homogeneous solution} \tag{2.19}$$

In the geometry of interest the forward-scattering propagator sets $\Lambda_{ih} = 0$ at $z = z_i$ (the only boundary condition of relevance) so the homogeneous solution is independent from the inhomogeneous one. It is furthermore unique via the exclusion of standing wave solutions (those which fulfil the zero boundary conditions).

$$\Lambda(z_f,\rho) = \int \mathrm{d}\rho' \, G_p(z_f - z_i, \rho-\rho')\Lambda(z_i,\rho') + \int\limits_{z_i}^{z_f} \mathrm{d}z' \int \mathrm{d}\rho' \, G_p(z_f - z', \rho-\rho')V(z',\rho') \tag{2.20}$$

The core of this approximation is the neglection of the $\left(\nabla_\rho\Lambda\right)^2$-term, apart from this it is a straightforward approach for a solution allowing direct evaluation. Furthermore, it can be transferred to the non-paraxial case by exchanging the propagator (the neglected term is then $(\nabla\Lambda)^2$).

The neglected term makes this a linear expression, which is per se a single scattering approximation: the at one point scattered wave is not scattered again. Hence, the neglected term must contain the multiple scattering effects in itself. This is remarkable as it only is a kind of quadratic self-interaction of the Rytov phase and does not contain any direct reference to the object.

It should be noted that the Rytov phase Λ is not well defined as the imaginary component (which represents the phase of the wave function) has to be contained in a 2π-interval since the phase is only defined there. However, the Rytov approximation represents the phase of the wave in the imaginary component of Λ which is not bounded or periodical. The mapping between the wave and Λ is therefore

only surjective but not injective, this can be mended by unwrapping the phase of the wave by means of some continuity assumption[2].

2.3. Multislice Algorithm

The Multislice algorithm is a split operator method, whereby a problem is split into different sub-problems which can be solved independently and these solutions are used to approximate a solution to the original problem. In the paraxial approximation in Eq. (2.3) the two operators that are to be separated are the propagation and the perturbation. This can also be done using the Huygens-Fresnel propagator as well only the propagation operator would change. The same goes for the axial propagator but this approach would be meaningless then.

This will be done here in the paraxial approximation but could also be done without it, only the propagator would change.

Without the perturbation (i.e. only propagation) the Green's function can be used to solve the unperturbed boundary condition problem with Eq. (2.7). Here, the wave function Ψ had its rapidly oscillating phase removed (as above in the paraxial and axial approximation).

$$\Psi_f(\rho) = \int \mathrm{d}\rho' \, G_p(z_f - z_i, \rho - \rho') \Psi_i(\rho') \tag{2.21}$$

$$= \frac{-ik}{2\pi(z_f - z_i)} \int \mathrm{d}\rho' \, e^{i\frac{k|\rho - \rho'|^2}{2(z_f - z_i)}} \Psi_i(\rho') \tag{2.22}$$

$$=: \hat{P}(z_f - z_i)\Psi_i(\rho) \tag{2.23}$$

Thus, the propagation operator $\hat{P}(z)$, which propagates a wave the distance z, is defined.

For the influence of the effective potential the axial Rytov approximation is used, which directly yields the interaction operator $\hat{V}(z_f, z_i)$:

$$\Psi_f(\rho) = \exp\left(\int_{z_i}^{z_f} \mathrm{d}z \, \frac{-i}{2k} V(z, \rho)\right) \Psi_i(\rho) \tag{2.24}$$

$$=: \hat{V}(z_f, z_i)\Psi_f(\rho) \tag{2.25}$$

Here, the influence of thermal motion are included as per Section 2.4.

A solution can be constructed using these solutions to partial problems, such as:

$$\Psi_f = \hat{V}(z_f, z_i)\hat{P}(z_f - z_i)\Psi_i \tag{2.26}$$

The error introduced by solving the partial problems separately scales with the distance $z_f - z_i$ so the z-interval is split up into a large number of subintervals. In the infinitesimal limit for the length of those subintervals this procedure would solve the original problem and yield the exact solution. In our case it is beneficial that the potential $V(z)$ is sparse in z as the atoms are essentially point-like in z (the propagation effect is negligible for the width of an atom's potential). So it is appropriate to split the interval (z_i, z_f) at the heights of the atoms z_i giving a chain of subintervals $(z_0 = z_i, z_1, ..., z_N, z_{N+1} = z_f)$ (with N the number of atoms). The solution can then be expressed

---

[2][10] Ghiglia and Pritt. 1998 . *Two-Dimensional Phase Unwrapping*.

as the interaction $\hat{V}_n = \hat{V}_n(-\infty + \infty)$ of atom n only and the propagation between the atoms:

$$\Psi_f = \hat{P}(z_{N+1} - z_N)\hat{V}_N\hat{P}(z_N - z_{N-1})\hat{V}_{N-1}...\hat{V}_1\hat{P}(z_1 - z_0)\Psi_i \tag{2.27}$$

$$= \hat{P}(z_{N+1} - z_N) \overset{N}{\underset{n=1}{\bigcirc}} \hat{V}_n\hat{P}(z_n - z_{n-1})\ \Psi_i \tag{2.28}$$

Here the iterative chaining operator $\bigcirc$ is introduced, which is defined akin to other iterative operators, such as $\sum$ or $\prod$, as the iterated application of the chaining operation to the term and the previous result.

2.4. Influence of Thermal Motion

In high-resolution microscopy the question of the influence of the thermal motion of the atoms has to be considered. The electron itself sees the atoms as stationary due to its speed and it's wave function is therefore the result of a configuration of fixed atoms, i.e. of a frozen lattice. Two main approaches are used in electron microscopy. Firstly, to simulate the scattering at many different frozen lattices and average the resulting waves or, in case of bright field microscopy, average the image intensities. This becomes prohibitively time-consuming for larger systems. Secondly, to smear out the potential of the atoms by a model such as the Debye-Waller factor, which is justified if the wave functions are averaged for the signal (as in holography).

For the motion itself a simple Einstein model is adapted and a harmonic binding potential, so the probability distribution of an atom is a Gaussian. But not the potential is convolved with this Gaussian distribution but the phase shift that potential exerts on the wave function. The blurred phase shift gets an additional amplitude dampening contribution, this is however not an absorbing effect but rather an incoherence effect which transfers intensity from the reconstructable wave function into the incoherent background. The transmission function t from the effective potential V with the Debye-Waller factor B in lateral direction is given by

$$t = \underbrace{\sqrt{\frac{4\pi}{B}}e^{-\frac{2}{B}(x^2+y^2)} *}_{\text{Thermal Diffuse Scattering}} \underbrace{\exp\left(\frac{-i}{2k}V\right)}_{\text{phase shift}} \tag{2.29}$$

Which can be evaluated in Fourier space:

$$t = \mathcal{F}_{xy}^{-1}\left[e^{-\frac{B}{8}\frac{k_x^2+k_y^2}{4\pi^2}}\mathcal{F}_{xy}\left[\exp\left(\frac{-i}{2k}V\right)\right]\right] \tag{2.30}$$

The amplitude-dampening effect can be attributed to the potential by defining an thermally averaged potential[3] $\tilde{V}$ as the logarithm of the convolution of the probability distribution p with the phase shift due to the potential V.

$$\tilde{V} = -i\log\left(p * e^{iV}\right) \qquad\qquad \text{thermally averaged potential} \tag{2.31}$$

---

[3]For the purposes of brevity, the $\frac{-i}{2k}$-factor is pushed into the potentials here.

3. Tomography

3.1. Tomography as Linear Inverse Problem

For the purposes of this work only linear tomography shall be considered, because there is a general framework for solving linear inverse problems, but not for non-linear problems. This framework gives criteria for a unique solution and algorithms for finding said solution. Furthermore it gives a theory, by the name of regularisation, which controls the "reconstruction" of contradictory data (e.g. noise) and represents a trade-off between a better fit and the amplification of errors.

Here, the following notation will be used:

$$\mathbf{v} = T\mathbf{u} \tag{3.1}$$

Where $\mathbf{u} \in U$ is the searched for and $\mathbf{v} \in V$ the measured quantity, which are (Hilbert space) vectors with the capital letters their respective Hilbert spaces and the projection operator T is a linear operator, mapping U into V and therefore $T \in V \times U^*$. More general definitions are certainly conceivable but are not necessary here as the concept of a Hilbert space is sufficient for next to all physical formalisms. This includes the representation of images (functions over two-dimensional space), tilt series of images (functions over three-dimensional space, two are spatial dimensions and one is the tilt angle) or their discrete counterparts (functions over discrete sets of coordinates).

In the continuous form the vectors $\mathbf{v}$, $\mathbf{u}$ and the tensor T can be projected onto position bases yielding $v(\mathbf{s}), \mathbf{s} \in S$, $u(\mathbf{r}), \mathbf{r} \in R$ and $T(\mathbf{s}, \mathbf{r})$ with the following product:

$$v(\mathbf{s}) = \int_R \mathrm{d}\mathbf{r}\, T(\mathbf{s}, \mathbf{r})u(\mathbf{r}) \tag{3.2}$$

In solving the problem of finding a $\mathbf{u}$ whose projection $T\mathbf{u}$ most closely matches a given $\mathbf{v}$, all tomographic reconstruction techniques correspond to solving a linear equation system in the least-squares sense:

$$\|T\mathbf{u} - \mathbf{v}\|_2 \to \min \tag{3.3}$$

In this case there is a linear solution space for $\mathbf{u}$ made up of the product of a singular solution and the kernel (the space of all vectors T maps to 0). By demanding that the solution is orthogonal to the kernel the solution becomes unique (and is called Moore-Penrose inverse). An exact solution is here neither possible nor wanted for, since it would include catastrophic noise induced artefacts. It is therefore the most feasible way to only approximate the least-squares solution, which is what all reconstruction algorithms do in the end. Here, a conjugate-gradient method is used which determines the steepest-descent direction of the norm of the residual, restricts that to the be orthogonal to the current intermediate solution and minimises the norm of the residual in that direction. It is optimal in the sense of convergence speed per iteration[12] for any general linear method and therefore uniquely suited for computationally expensive problems. The regularisation behaviour is harder to predict

---

[1]p 112ff [19] Louis. 1989 . *Inverse und schlecht gestellte Probleme.*
[2]p 177ff [7] Engl et al. 1996 . *Regularization of inverse problems.*

(although it can be done with great effort, e.g. by calculating the singular value decomposition), in general more iterations mean less regularisation (i.e. a higher danger of over-fitting)[3].

In this work the LSQR algorithm of Paige and Saunders[4] is used, which is an implementation of the conjugate gradient method optimised for sparse and ill-conditioned problems.

3.2. Tomography inside a TEM

The case which is here of interest is broadly similar to the original one: A beam is shone upon a specimen and the transmitted part is detected; this procedure is repeated for different incident directions of the beam. The images are then used to reconstruct the three-dimensional distribution of a quantity. In electron tomography there are several different object quantities, which can be reconstructed via tomography. This only depends whether an applicable model for the creation of the signal and a reconstruction method for that model exist. This case (which is widely accepted to be the standard one) will be called Ray Projection Tomography (Section 3.3).

Vectors of the object coordinate system shall be denoted by $\mathbf{r} = (x, y, z)$. The coordinate system of the detector shall be transferred into the object plane, therefore compensating the scaling of the magnification of the imaging system. Aberrations and similar effects are not wholly neglected but rather calculated back into the object plane rather than into the detector plane. The exact focal plane in relation to the object (i.e. the focus/defocus) can be set at will here since the complex wave is available. For the wave itself only the lateral coordinates $(x', y') =: \mathbf{e}$ of the detector are relevant, but here the degrees of freedom of the orientation of the object coordinate systems in relation to the microscope coordinate system are important as well. This relation is given by a shift (which will be set to 0 by alignment) and a rotation of the two coordinate systems against each other, which would mean an element out of $\mathcal{SO}(3)$ (the group of orthogonal 3×3-matrices with determinant 1) but considering that the in-plane rotation of the detector (i.e. around z') does not change the signal in any meaningful way, the relevant part of the orientation is the direction of z' in the object coordinate system and this is $\theta \in S^3$ in the three-dimensional unit sphere.

In reality the geometry is often further simplified by aligning an axis of the object with one of the detector, so only one degree of freedom remains, i.e. the angle $\theta \in$
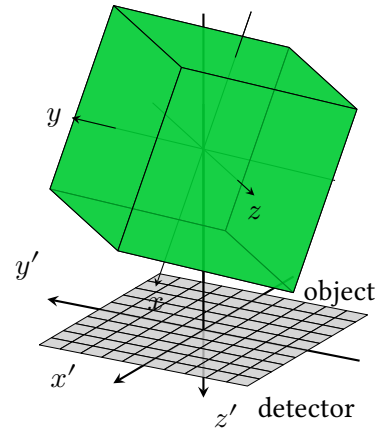

Figure 3.1.: Very general form of the geometry inside a TEM, the relation between the object coordinate system $(\mathbf{e_x}, \mathbf{e_y}, \mathbf{e_z})$ and the microscope (detector) coordinate system $(\mathbf{e_{x'}}, \mathbf{e_{y'}}, \mathbf{e_{z'}})$ can be altered by the goniometer of the microscope (or optically).
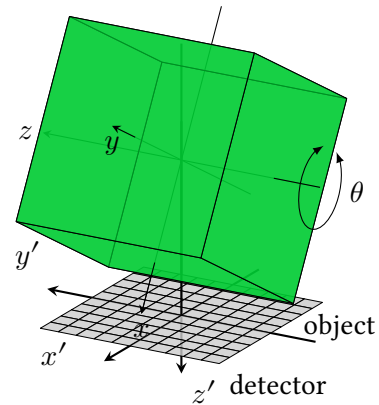

Figure 3.2.: Single tilt axis geometry of electron tomography, whereby the unit vectors $\mathbf{e_z}$ and $\mathbf{e_{y'}}$ of the microscope and the specimen are aligned. The remaining degree of freedom is the angle between $\mathbf{e_x}$ and $\mathbf{e_{z'}}$.

---

[3]p 103ff[19] Louis. 1989 . *Inverse und schlecht gestellte Probleme.*
[4][23] Paige and Saunders. 1982 . "LSQR".

$[-\pi, \pi[$ [5]. Experimentally this equates to only using one
tilt mechanism of the goniometer; although the TEM goniometers allow a large range of tilting, it will pose problems regarding for example shading, so it is easiest and usually sufficient to only use the tilt around the long axis of the holder.

3.3. Ray Projection Tomography

The traditional form[6] of tomography describes the signal in a pixel as the line-integral through the specimen parallel to the beam's direction that intersects the detector pixel in question, which is the same as setting the initial wave to 1 in Eq. (2.13). The coordinates in the projection operator are $\mathbf{r}$ as the three-dimensional position-vector inside the specimen and $\mathbf{s}$ as the coordinate of a pixel in the tilt-series, where it can be decomposed into the detector coordinate ρ' and the specimen orientation $\theta \in S_3$ relative to the beam:

$$T(\mathbf{s}; \mathbf{r}) = T(\rho', \theta; \mathbf{r}) \qquad\qquad v(\mathbf{s}) = v(\rho', \theta) \qquad (3.4)$$

In terms of the axial Rytov approximation Eq. (2.13) the coordinates are transformed in a lateral and axial displacement:

$$\mathbf{e} = \rho' - (\mathbf{r} - \langle \mathbf{r}, \theta \rangle \, \theta) \qquad\qquad \text{lateral coordinate} \qquad (3.5)$$
$$d = \langle \mathbf{r} - \rho', \theta \rangle \, \theta + d_0 \qquad\qquad \text{axial coordinate} \qquad (3.6)$$

Now it is easy to formulate the tensor:

$$T(d(\rho', \theta, \mathbf{r}), \mathbf{e}(\rho', \theta, \mathbf{r})) = \delta(\mathbf{e}) \qquad (3.7)$$

$$(3.8)$$

Thus the application of the tensor to the object vector is just a convolution with a delta-function in the lateral coordinates (from the perspective of the detector). The convolution kernel is however extended in the axial coordinate d and in the tilt coordinate θ.

3.4. Fresnel Propagator Tomography

A new form of electron tomography is proposed, which instead of the ray projector uses a Fresnel propagation convolution kernel as model for the signal. It is an adaptation of Diffraction Tomography[7][8], a method used in visible light microscopy to reconstruct the diffractive indices of semi-transparent objects, a condition not fulfilled in electron scattering. It relies directly on the Rytov approximation (Section 2.2, more specifically Eq. (2.20)) incorporating the Fresnel propagation of the wave (although neglecting the multiple scattering effects). Since the spatial frequencies of TEM are small, it is justified to use Fresnel diffraction rather than Huygens-Fresnel diffraction.

---

[5] Please note the difference between the vector θ and the angle θ, which stands for the restriction of the vector θ to a great circle of S^3

[6] The name Radon is avoided because the Radon transform in its strict mathematical formulation concerns hyperplanes, which would be planes in three dimensional space. In single tilt axis tomography the ray transform can be reduced to the Radon transform but this is only relevant later on.

[7] [15] Kak et al. 1988 . *Principles of Computerized Tomographic Imaging*.

[8] [22] Müller et al. 2015 . "The Theory of Diffraction Tomography".

3.4.1. Rytov Approximation Tomography

In Eq. (2.20) the initial wave is set to 1 (so the initial Rytov phase becomes 0) resulting in an expression linear in V.

$$\Lambda(z_f, \rho) = \underbrace{\int \mathrm{d}\rho'\, G_p(z_f - z_i, \rho - \rho')\Lambda(z_i, \rho')}_{=0} + \int_{z_i}^{z_f} \mathrm{d}z' \int \mathrm{d}\rho'\, G_p(z_f - z', \rho - \rho')V(z', \rho') \quad (3.9)$$

  Very much like Section 3.3 the problem is effectively transformed into a space where the lateral displacement $\mathbf{e}$ and the axial displacement d between a point in the reconstruction volume $\mathbf{r}$ and a pixel in the tilt series $\mathbf{s} = (\rho', \theta)$, and the tensor, is again a convolution (although not a delta function any more). The axial component must be normalised by the effective height d_0 of the detector relative to the specimen coordinate system, i.e. the focus plane in the specimen. In these new coordinates the kernel is the Fresnel convolution kernel:

$$T(d(\rho', \theta, \mathbf{r}), \mathbf{e}(\rho', \theta, \mathbf{r})) = G_p(d, \rho) = \frac{-ik}{2\pi d} e^{i\frac{k|\mathbf{e}|^2}{2d}} \quad (3.10)$$

Since back-propagation is allowed, as the detector is well below of the specimen but the reference height (the focus plane) may be inside the specimen, the Heaviside Theta-function is omitted. Till now, this may seem trivial, but the efficient numerical implementation is rather more difficult than the ray projection.

4. Computational Implementation

For the reconstruction both ray projection and Fresnel propagator tomography are implemented as tensor products with a precomputed tensor and the multiplication operations for the tensor (or its adjoint) mapping the reconstruction space to the tilt series space (or vice versa). The reason for precomputing the tensors is the optimal speed of the product operations (without using Fourier space methods (which all imply periodicity)) at the cost of large memory requirements (the tensor has the nominal size of the number of tilt series pixels times the number of voxels in the reconstruction volume). Since coordinate shifts are virtually free, it is desirable to use any symmetries of the tensor to reduce its representation. But this needs to be discussed in an explicit geometry.

The detector $\rho' = (x', y')$ and the reconstruction volume $\mathbf{r} = (x, y, z)$ are sampled using an equidistant grid, while the tilt angle θ is sampled at specific (not necessarily equidistant) points:

$$x'_\alpha : \alpha = 1, ..., N_{x'} \qquad\qquad x_\mu : \mu = 1, ..., N_x \qquad (4.1)$$

$$y'_\beta : \beta = 1, ..., N_{y'} \qquad\qquad y_\nu : \nu = 1, ..., N_y \qquad (4.2)$$

$$\theta_\gamma : \gamma = 1, ..., N_\theta \qquad\qquad z_\xi : \xi = 1, ..., N_x \qquad (4.3)$$

As in Fig. 3.2 the axes y' and z are aligned and, crucially sampled equally $y'_\beta = z_\beta$. Furthermore, the vectors for the object $u(x_\mu, y_\nu, z_\xi) = u_{\mu\nu\xi}$ and the tilt series $v(x'_\alpha, y'_\beta, \theta_\gamma) = v_{\alpha\beta\gamma}$ are dense (defined for all combinations of their indices).

The tensor $T(\rho', \theta; \mathbf{r})$ can be written as $T_{\alpha\beta\gamma}^{\mu\nu\xi}$ in tensor notation (Einstein summation convention), so the required multiplications become:

$$v_{\alpha\beta\gamma} = T_{\alpha\beta\gamma}^{\mu\nu\xi} u_{\mu\nu\xi} \qquad\qquad \text{and} \qquad\qquad u_{\mu\nu\xi} = T^{+\alpha\beta\gamma}_{\mu\nu\xi} v_{\alpha\beta\gamma} \qquad (4.4)$$

Since the order of the indices in $\alpha\beta\gamma$ and $\mu\nu\xi$ are irrelevant, they are flattened into the two indices $m = 1, ..., N_{x'} N_{y'} N_\theta$ and $n = 1, ..., N_x N_y N_z$. Instead of numbering each voxel of the cube of data by a unique combination of indices for every dimension it is simply numbered with a single index. For reasons, which will become obvious shortly, the orginial dimensions are sorted so that the y'_β and the z_ξ axes are the outermost and θ_γ is the second outermost (which will be important for the way the tensors are reduced in complexity and represented in memory):

$$m = \beta N_\theta N_{x'} + \gamma N_{x'} + \alpha \qquad (4.5)$$

$$n = \xi N_x N_y + \nu N_y + \mu \qquad (4.6)$$

$$v_m = T_m^n u_n \qquad (4.7)$$

The nominal size of the tensor is $N_{x'} N_{y'} N_\theta N_x N_y N_z$ and thereby extremely large (assuming $360\,\mathrm{px}$ for every dimension, 120 tilt angles and the complex single precision datatype the resulting tensor would be $5\,\mathrm{PiB}$).

4.1. Ray Projection

The tensor Eq. (3.7) is not extended along y' and z (since they are lateral and pixel-aligned), so they can be collapsed into a delta symbol $\delta_{\beta\xi}$ and the tensor can be simplified:

$$T^{\mu\nu\xi}_{\alpha\beta\gamma} = \widetilde{T}^{\mu\nu}_{\alpha\gamma}\,\delta^{\xi}_{\beta} \tag{4.8}$$

Hence, it is sufficient to calculate $\widetilde{T}$ and apply it to the blocks of m and n that correspond to equal β and ξ. Geometrically this means that the problem can be separated along the tilt axis and the slices treated independently from each other. The reduced tensor $\widetilde{T}$ cannot be simplified further, even though it is just a delta-function, since the lateral and axial coordinates and are not aligned and the kernel needs to be interpolated. This, however, can be done analytically and reasonably fast.

4.2. Fresnel Propagation

Here, the kernel (Eq. (3.10)) the kernel has a lateral extension. But it is still possible to reduce the complexity of the tensor, because it is a convolution, which only depends on $y' - z$. Since they are aligned and equidistant, this is equivalent to a dependence on $\beta - \xi =: \tau$, so the reduced tensor is:

$$T^{\mu\nu\xi}_{\alpha\beta\gamma} = \widetilde{T}^{\mu\nu\tau}_{\alpha\gamma}\,\delta^{\xi}_{(\beta-\tau)} \tag{4.9}$$

The relevant range of values of τ is small (compared to the ones of β and ξ) and the tensor can be separated into smaller ones which are applied to the pairs of blocks of m and n fulfilling $\tau = \beta - \xi$. The computation of the reduced tensor is more complicated than the one from the ray projection. After several approaches the following implementation crystallised to be the most practical:

1. the propagator is computed in partial Fourier space on a grid equivalent to $\mathbf{k}_\rho, d$ (from Section 3.4)

   - d has the same (or better) sampling as x_μ and y_ν and the necessary extent to include all possible values of d (in the tilt series' geometry)
   - $\mathbf{k}_\rho$ has the extent that corresponds to the sampling of, in one direction the highest out of ρ_{xy}, x_μ and y_ν, and, in the other direction the highest out of ρ_τ and z_ξ. The sampling is chosen high enough to curtail aliasing during the inverse Fourier transform

2. the band-limit is applied as a mask (circular Bessel function mask) in the plane of reciprocal ρ

3. the propagator is transformed into real space

4. insignificantly small values are cut off and the extent of the array is reduced where it is zero, this determines the range of $\tau = N_\tau^-, ..., N_\tau^+$

5. for every τ-layer the ρ_{xy}, d coordinates of α, γ, μ, ν are determined and if the propagator is not zero there, the value is interpolated linearly

The propagator is symmetric in the sign of τ so using that symmetry the tensor could be further simplified and only half of the tensor would have to be memorised (accounting for the $\tau = 0$ slice which is the one with the most non-zero values). This was not done here as the time needed for the tensor products became the hindering factor rather than the size of the tensor in memory.

This procedure for computing the tensor takes a rather long time but certainly less than the application of the tensor afterwards, so the precomputation of the tensor is a net gain. The step of

linearly interpolating the propagator on the different tilt angles has an inherent problem of inaccuracy as the local curvature of the kernel is high in some place. A better approximation of the kernel at additional computational expense could be advisable.

4.3. Sparse Representation and Parallelisation

Even the reduced tensors are largely consisting of zeros, therefore it is beneficial to switch to a sparse representation in memory, where not the values T_m^n are stored for every m and n but rather where for every non-zero entry $l = 1, ..., N_l$ the triplet (m_l, n_l, T_l) is stored. This makes it computationally very expensive to select a single column or row of the array, since the entries follow no particular order in m or n. It is, however, possible to impose such an order, but in light of the size of the tensor it should be already included in the computation of the tensor, reordering the tensor would take very long. In the Fresnel case the tensor was ordered for different τ so the parts of the tensor pertaining to a specific τ can be quickly selected, because they are a contiguous segment in memory.

The tensor product now pertains simply to initialising v_m as zero and iterating through l adding $T_l u_{n_l}$ to v_{m_l} for each l. The product of the adjoint operator $u = T^+ v$ is the same only with the roles of u_n and v_m exchanged and the T_l entries conjugated. To save time not T_l is conjugated but u_n and v_m are (which takes only $\propto N_n + N_m$ time instead of $\propto N_l$, which is much larger and the proportionality factor is about the same).

segmentation of the problem along the y' (or z for the adjoint product) axis of the output array, the reduced tensor is applied with different offsets into u_n and v_m, which makes for N_z completely independent computation blocks. Since they each write exclusively into their own segment of the output array these blocks can be computed in parallel[1] yielding N_z tasks of the same required run-time so load balancing is not a problem.

A performance concern is the near to random memory access of u_{n_l} and v_{m_l} (the items of the reduced tensor are next to unordered) so the CPU cannot guess which parts of the array are needed next and prefetch them into its cache. This may be remedied to some degree by the Hyper-threading of the Intel CPU, where a physical core executes two threads simultaneously (or rather interleaves their execution), the cycles normally wasted by waiting for the data from the RAM are yielded to the other thread, the execution time does benefit but not much and not all the time this was tested, so no qualified statement can be made here.

---

[1]Memory collisions, where two threads try to access a memory cell simultaneously are only critical when writing or doing in-place addition (which is used here). Simple reading instructions cause no problems and are not even slowed down.

5. A Large-Field-of-View Multislice Algorithm

Since the specimen used here are nanoparticles and by their very nature non-periodic, no periodicity can be used in the Multislice algorithm either. Using the rule-of-thumb of a sampling of $\approx 10\,\mathrm{pm}$ this would mean a comparatively small field-of-view of $20\,\mathrm{nm}$ needs to be sampled with $2028\,\mathrm{px} \times 2048\,\mathrm{px}$ wave in the simulation. Also this would only allow a nanoparticle somewhat smaller than $20\,\mathrm{nm}$ as vacuum padding is necessary as long as a Fourier space propagator is used. Nonetheless, such a nanoparticle would still include many thousands of atoms, which, together with the high number of pixels in the wave makes, the simulation very slow.

In a conventional Multislice algorithm the atoms are numerated in ascending order of z with $n = 1, ..., N$ with $\hat{V}_n$ (Eq. (2.25)) being their respective transmission functions; in between the atoms the wave is propagated by $\hat{P}(z_n - z_{n-1})$ (Eq. (2.23)).

$$\Psi_f = \hat{P}(z_{N+1} - z_N) \underset{n=1}{\overset{N}{\bigcirc}} \hat{V}_n \hat{P}(z_n - z_{n-1}) \Psi_i \tag{5.1}$$

It is sometimes (and certainly here) beneficial to combine several atoms into a single layer (which means displacing them into a common height z). The propagator (as a convolution) is evaluated in Fourier space and computationally expensive compared to the transmission function construction for a single atom.

$$\hat{P}(z)\Psi(\rho) = \frac{-ik}{2\pi z} \int \mathrm{d}\rho' \, e^{i\frac{k|\rho|^2}{2z}} \Psi(\rho') \tag{5.2}$$

$$= \mathscr{F}_{xy}^{-1} \left[e^{\frac{iz}{2k}|\mathbf{k}_\rho|^2} \, \mathscr{F}_{xy} \left[\Psi(\rho) \right] \right] \tag{5.3}$$

The transmission function is a bit more complicated as it contains the projected potential of a single atom and the blurring by the thermal motion of the atom described with the, so called, Debye-Waller factor (see Section 2.4). Please note that not the potential itself is blurred but its exponential since this is the quantity effecting the electron wave.

The lateral Fourier transform of the projected potential can be calculated from the atomic potentials calculated from parametrisations such as the one of Weickenmeier and Kohl[1], which was used here. Furthermore, the lateral shift of the atomic potential is performed in Fourier space. The potentials for every type of atom can be precomputed for the sampling of the problem at hand so the computation for each individual atom merely amounts to one inverse Fourier transform. The Fourier transforms are realised as Fast Fourier Transforms (FFT).

---

[1][34] Weickenmeier and Kohl. 1991 . "Computation of absorptive form factors for high-energy electron diffraction".

5.1. Bunching of Atoms

The first optimisation for the simulation of many-atom objects is to bunch atoms of similar height into a common height.

$$\hat{V}_{(n...m)} = \prod_{l=n}^{m} \hat{V}_l \tag{5.4}$$

$$\approx \hat{V}_m \hat{P}(z_m - z_{m-1}) \hat{V}_{m-1} ... \hat{P}(z_{n+1} - z_n) \hat{V}_n \tag{5.5}$$

The maximum of displacement for a single atom should be chosen according to some reasonable criterion. In this case this is:

$$\delta z = \frac{1}{4} 2k/|\mathbf{k}_{max}|^2 \tag{5.6}$$

This equates to a phase shift of $\frac{\pi}{20}$ in the phase term of the Fresnel propagator Eq. (2.9), which was chosen somewhat arbitrarily and yields a δz of $8\,\mathrm{pm}$ for a sampling of one per $10\,\mathrm{pm}$ and an electron energy of $200\,\mathrm{keV}$. For the specimen in question (Section 6.2) $44\,000$ atoms were reduced to ≈ 1300 slices which made the time spent on propagation low enough in comparison with the time needed to create the transmission functions.

5.2. Patching of Transmission Function

As the field-of-view is so large, and certainly much larger than the range of the atom's potential, it is not necessary to calculate transmission functions for the whole region of the wave. The calculation in a region of $2\,\mathrm{nm}$ width is certainly sufficient and reduces the time for the calculation of the Fourier transformed transmission function by about the factor 100. For the case of bunched atoms there is only one inverse FFT needed to calculate the transmission function (the atoms of one bunch can be added in Fourier space). Here, one inverse FFT is needed for each atom, since the patches do are different for each atom. So the reduced size of the FFT is offset by the number of FFTs needed. The pixels of the patch are aligned with the pixels of the full wave, the subpixel shift is then done in the Fourier space of the patch (the subregion is large enough that the implied periodicity causes no problems).
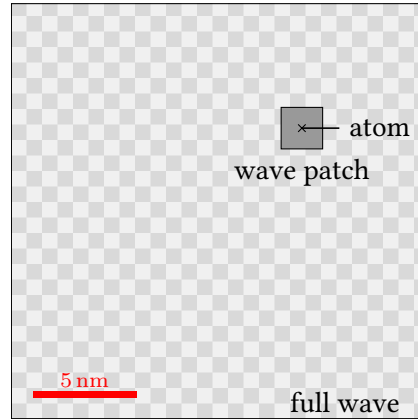
Figure 5.1.: The transmission function for a single atom is calculated in the region around the atom and the patched into the transmission function for the full wave. The proportions are about the same as used in the simulation (Section 6.2).

5.3. Porting onto the GPU

In contrast to Central Processing Units (CPUs), Graphical Processing Units (GPUs) are optimised for massively parallel operations on data using a lot of relatively simple units rather than a CPU which has few but complicated units. In this case the GPU has 1152 CUDA cores (which handle the

low-level calculations but not the control functions, like instruction interpreting and scheduling) and the CPU has 4 physical cores with 2 Floating Point Units per core, so 8 numerical units in total. Both, CPU and GPU have a large instruction set with a lot of specialised instructions (for their respective applications as well as high-performance computing) so there is little that can be said in general about comparing the two. One can only state that the GPU handles very parallel algorithms of simple instructions better, whereas the CPU handles sequential problems with complicated control structures better. Also, whereas the former has a very fast connection to its onboard memory, the latter has a faster connection to the systems working memory (RAM), which is usually larger than the GPU's onboard memory (here, $32\,\text{GiB}$ vs $2\,\text{GiB}$). In GPUs optimised for high performance computing this can look very different and systems of multiple GPUs or CPUs pose even more aspects that need to be considered, e.g. connection and load balancing between the nodes.

The motivation here was that a FFT is generally faster on a GPU than a CPU (it is a simple algorithm of simple instructions). The implementation was straightforward done in the meta-language and on-demand compilation system Reikna (Appendix B).

5.4. Speed Measurements

To evaluate the speed of the different methods a wave of $20\,\text{nm}$ width and height (and sampled with $\approx2000\,\text{px} \times \approx2000\,\text{px}$) was scattered at the specimen of Section 6.2 (with $44\,\text{k}$ atoms) using the above described algorithm. This happened on a system sporting an Intel Xeon E3-1270v3 @ 3.5 GHz quad core (with Hyper-Threading Technology enabled) with $4 \times 8\,\text{GiB}$ (running at DDR3-2400 MHz CL11) and a Nvidia GTX-760 with $2\,\text{GiB}$. The bunching in z was activated the whole time, otherwise it would take too long.

It should be noted that most of the operations where multi-threaded at low-level, thanks to the virtue of Numpy (linked against the Intel MKL BLAS), NumExpr (although not using Intel VML) and PyFFT (which uses the FFTW library).

Scenario	Time taken in s	relative speed	Time spent on FFTs in s	in %
baseline	4249	1	314	7
patching only	480	9	262	55
GPU and patching	166	26	NA	NA

Figure 5.2.: Performance statistics with and without certain performance measures using the above described parameters. The baseline already includes the bunching. The profiling information for the tasks running on the GPU is not available.

The main benefit of the patching is the time saved from calculating the transmission function in Fourier space, the reduction in time spent on FFTs is not that high. The implementation on the GPU is then even faster by a factor of about three. With that and the bunching of atoms, a single simulation with the sampling $(2048\,\text{px})^2$ and $\approx44\,000$ atoms takes $\approx4\,\text{min}$. Compared to the multislice implementation of the SEMI simulation software developed at the Triebenberg this is a speedup of $\approx60^2$. Considering that a tilt series of 120 angles takes $8\,\text{h}$ that speedup is essential.

---

[2]To compare the speed with a previously done simulation[17] the complexity of the SEMI simulation was assumed to scale with the number of atoms and the complexity of the FFTs ($N \log N$, with N the number of pixels).

6. Electron Tomography with a Fresnel Propagator Kernel

In this section, the tomographic reconstruction based on the Rytov approximation is carried out for a number of test cases and compared to the conventional ray projection tomography based on the PGA. While the terms Rytov approximation and PGA per se stand for approximations they shall, by extent, stand for the thereon based reconstruction techniques as well.

First, a short note about the relevant quantities should clarify some points. In both models the effective potential V (see Eq. (2.1)) is connected to the Rytov phase Λ (see Section 2.2) via the tensor T, where all of these three quantities are complex-valued. Accordingly, they all have two components which carry specific physical meaning and will be named accordingly, although the effective Potential V is a scaled form of the "pure" electric potential and the absorbing potential is a statistical result of the thermal diffuse scattering (see Section 2.4). The different meaning of the components of the potential and the Rytov phase are summarised as follows:

electric potential $\mathrm{Re}\,V$**:** phase-shifting part of the potential, in this case the electric interaction between electron wave and atom cores

absorbing potential $\mathrm{Im}\,V$**:** absorbing part of the potential, chiefly due to thermal diffuse scattering

amplitude $\mathrm{Re}\,\Lambda$**:** logarithm of the amplitude of the electron wave

phase $\mathrm{Im}\,\Lambda$**:** phase of the electron wave

The potential is given in Volts but without a detailed discussion of the theoretical expected value the absolute value cannot be vouched for. Both components of the potential are expected to be positive, because of the inherently positive electric potential of the atoms' cores and the amplitude dampening nature of the imaginary part (i.e. when multiplied with i it should being smaller than 0, as it is the logarithm of the amplitude). As a virtual comparison object the PGA reconstruction of Fig. 6.1 is used giving a reference value for tungsten of $V = (700 + i150)\mathrm{V}$.

6.1. Single Atom

It is at first beneficial to consider a very simple system to confirm and illustrate the effect of using a Fresnel propagation kernel instead of a simple ray in electron holographic tomography. To that end, a single atom (of tungsten as a strong scatterer) is simulated in a small volume of $5\,\mathrm{nm} \times 5\,\mathrm{nm} \times 5\,\mathrm{nm}$ at the same acceleration voltage of $200\,\mathrm{keV}$ and with the same sampling and preparation steps as the nanotube specimen (see Section 6.4) considered lateron. Thus, a dataset of 120 waves (using the full tilt range @ 3° increments) with a sampling of $120\,\mathrm{px} \times 120\,\mathrm{px}$ is obtained. The focus of these waves is set at the centre of the volume.

6.1.1. Centred Atom

The atom is positioned exactly at the centre of the volume and therefore always in focus. Thus, propagation should play no role in this particular simulation.
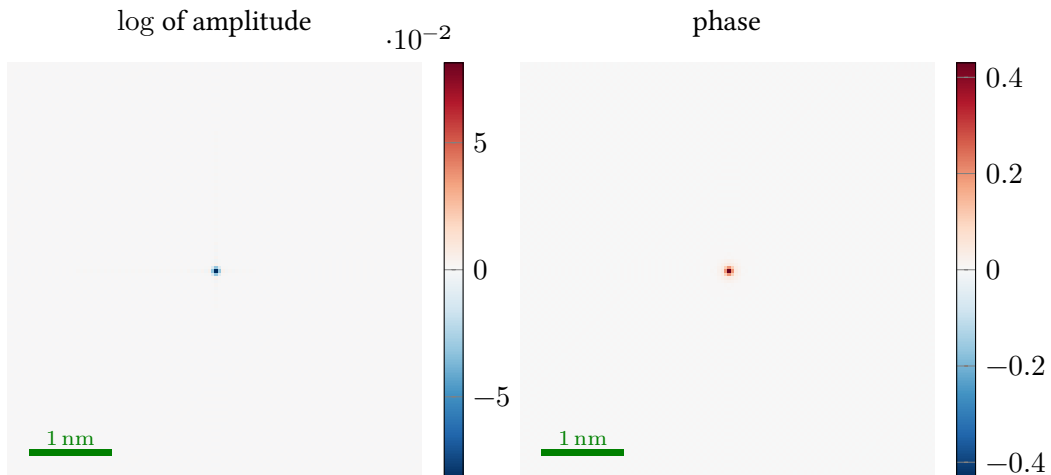
Figure 6.1.: Components of the wave function of a single W atom specimen in focus.

The wave, shown in Fig. 6.1, corresponds to the transmission function (Eq. (2.30)) of the single atom and is exactly the same for all tilt angles. The interaction between the electron wave and the atom core is a purely phase-shifting one, the considerably smaller amplitude contribution is due to the thermal diffuse scattering and weaker by about the factor 5. Since the atom is always centred and quite small, there should be next to no influence of the propagation and henceforth no difference between the two methods.
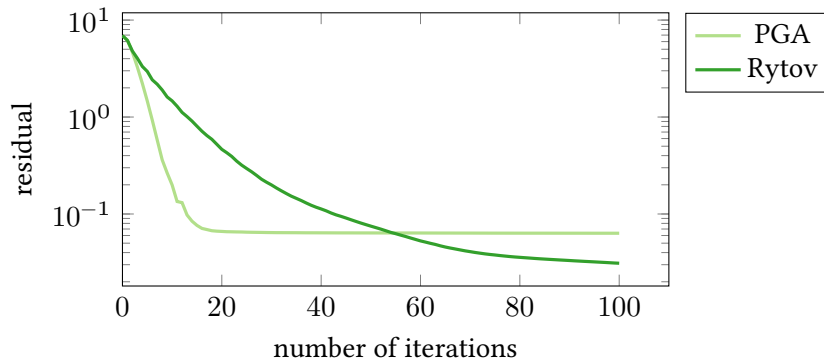
Figure 6.2.: The norm of the residual (i.e. the value of the objective function) during 100 iterations of the LSQR algorithm. The problem does not involve any propagation effects and Rytov in itself is more ill-conditioned so the considerably faster convergence of the PGA is as expected. The ultimately lower residual in the Rytov case stems most likely from artefacts.

The Rytov case is per se more ill-conditioned than the PGA approach because the kernel has a lateral extension. The slower speed of convergence of the Rytov approximation is thus as expected, The PGA case converges quickly to a near constant value, meaning, that it is very slow to overfit the values (even numerical noise can be "reconstructed" into artefacts). The Rytov approximation reduces the residual further and seems to settle at about half an order of magnitude below the PGA.
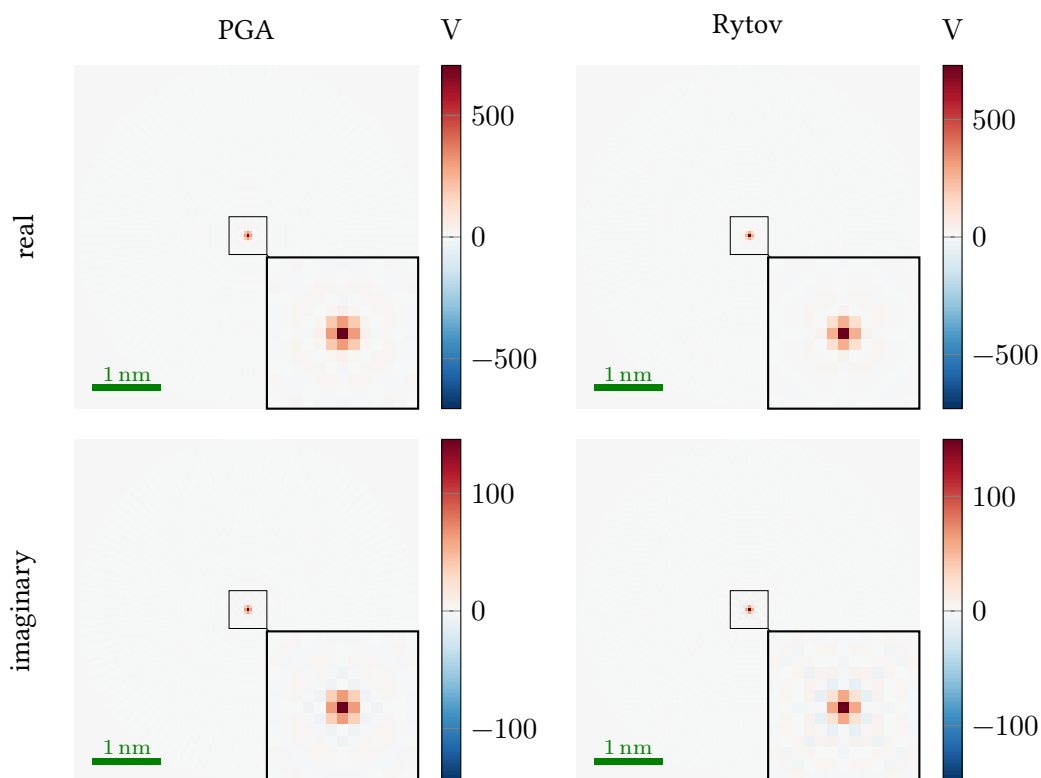
Figure 6.3.: Cross section of the reconstructed potential perpendicular to the tilt after 100 iterations. The atom is reconstructed correctly, with the Rytov approximations throwing some slight artefacts around the peak.

As seen in Fig. 6.3,both methods reconstruct the single atom peak as sharp as it is in the tilt series with nearly the same absolute value. The Rytov approximation causes some slight artefacts (the negative areas in around the peak), most likely a result of overfitting.

This is the best-case scenario for the PGA, without influence of propagation, and the PGA converges very fast to its final value. It is, however, not that much faster than the Rytov approximation and the results are effectively identical, even though the Rytov approximation begins to produce some artefacts.

6.1.2. Off-centre Atom

To introduce propagation effects into the problem the single atom was shifted a mere $2\,\text{nm}$ out of the centre, so its position and the apparent defocus change over the course of the tilt series.
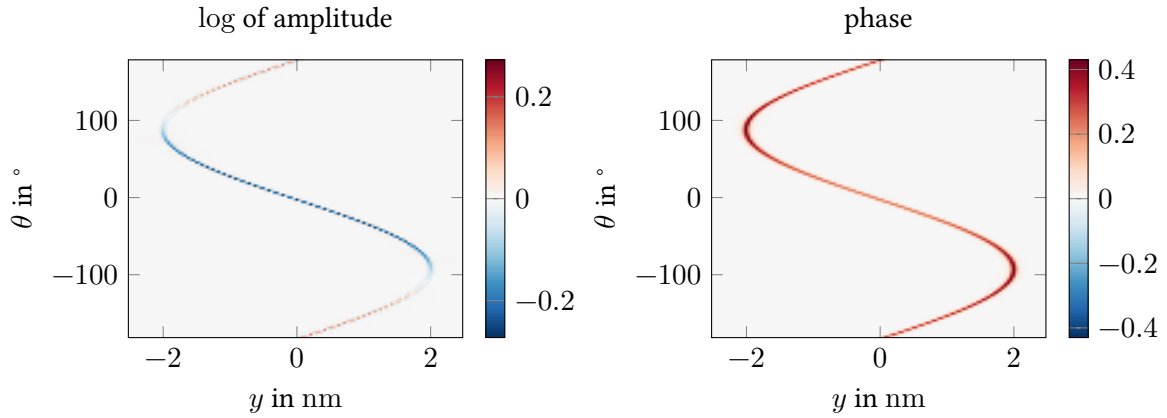
Figure 6.4.: Sinogram of the middle slice ($x = 0$) as the object is rotated about y by the angle θ. The sinusoidal pattern stems from the non-central position of the atom, which is at its maximum of underfocus for $0°$ and at its maximum of overfocus for $-180°$.

The tilt series exhibits the expected propagation effects in the amplitude and the phase. The amplitude is affected strongly in under- and overfocus, where the phase signal is mixed via the imaginary component of the propagator as a positive term for the over- and a negative term for the underfocus. Because the phase is so much stronger than the amplitude modulation (Fig. 6.1). The mixing of phase into the amplitude is much stronger than the other way around, there the phase mainly exhibits the broadening due to the real part of the propagator).
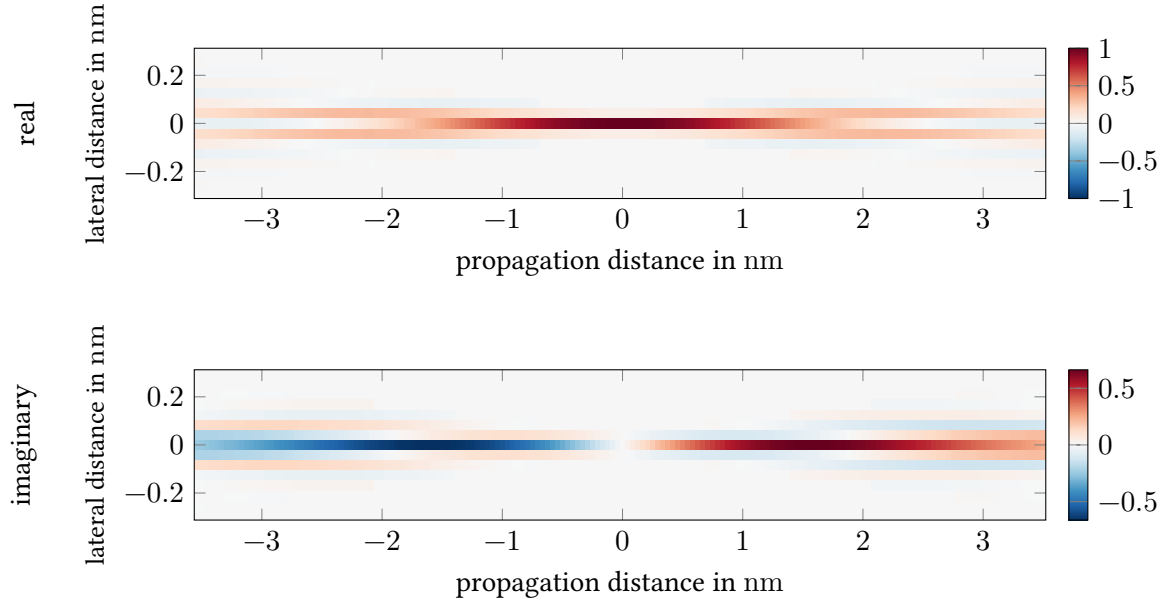
Figure 6.5.: Fresnel propagation convolution kernel resolution limited to the sampling rate of the reconstruction grid. The middle of the these images corresponds to the middle of the xy-plane of the reconstruction volume. This kernel was used in the reconstruction, the simulation used the Fourier space form of the propagation.

The real part of the propagator (see Fig. 6.5) represents, how each component of the logarithm

of the wave is mapped to its direct pendant (real to real, imaginary to imaginary). The imaginary part maps each component to its opposite. The former is symmetric around the focal plane, whereas the latter is anti-symmetric. This propagator is normalised and does not include the scaling of the reconstructed potential.
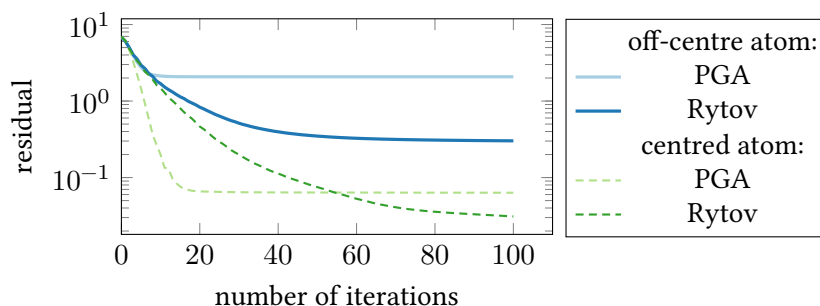
Figure 6.6.: The norm of the residual (i.e. the value of the objective function) during 100 iterations of the LSQR algorithm. Although the Rytov approximation leaves a larger residual when in flattens out, the advantage over the PGA is even larger than with the centred atom.

When analysing the convergence behaviour of the two techniques (see Fig. 6.6) it can be seen that the starting point is nearly the same as in the previous case of a centred atom but the final levels of residual are considerably higher. The Rytov approximation performs even better against the PGA, converging just as fast and fitting the data much closer than the PGA.
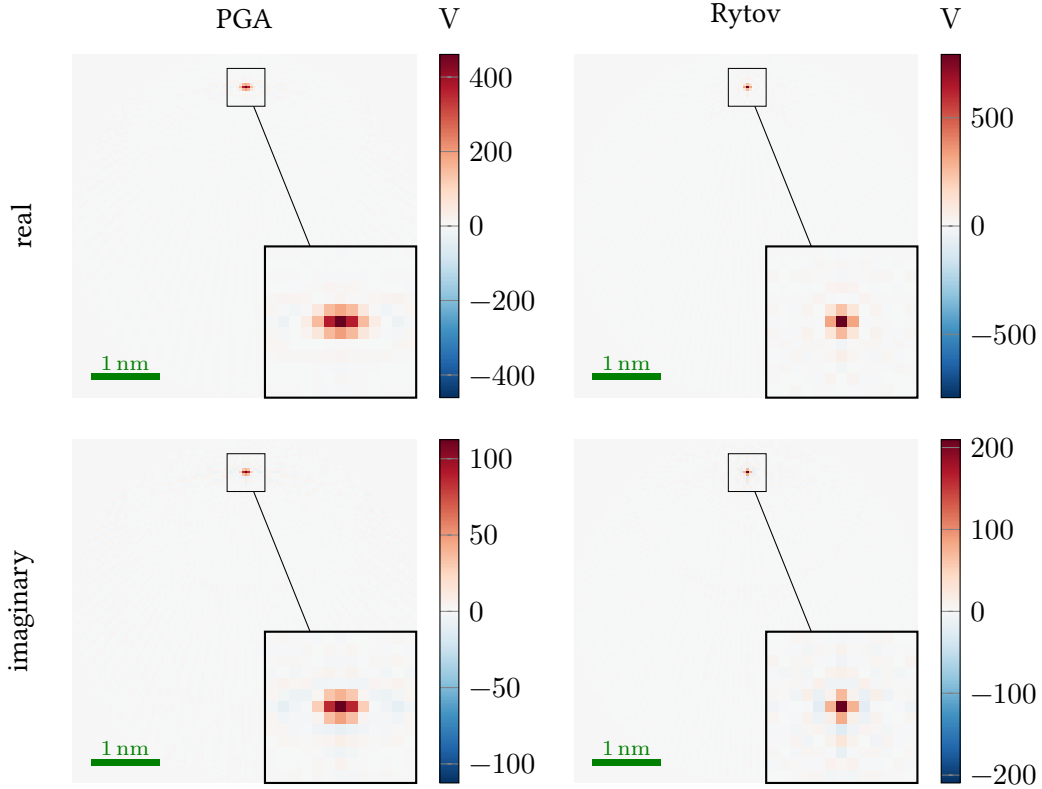
Figure 6.7.: Cross section perpendicular to the tilt axis of the reconstructed, complex valued, potential after 100 iterations. The Rytov approximation reconstructs the atom in a very narrow peak, whereas the PGA exhibits an elongation perpendicular to the direction of the greatest displacement and, therefore, the greatest propagation distances.

While the atom is reconstructed clearly in a single peak with both methods, the PGA has an elongation perpendicular to the direction of displacement from the centre. The PGA is quite good, considering it can not model the strong propagation effects visible in Fig. 6.4. These effects are compensated by the full 360° tiltseries used here. The opposing projections, which the PGA treats as identical, are effectively averaged (their contradicting part is suppressed by regularisation), which partially compensates the propagation effects. Due to the position of the focal planes (at the centre of the specimen) the anti-symmetric part of the propagator is compensated, so, from the view of the PGA, only the real component remains. This component smears out the atom potential mostly perpendicular to the direction of greatest defocus, which explains the elongation.

The Rytov approximation concentrates the information in a very sharp peak, with only a slight elongation in horizontal direction. It is near the reference value for the peak height, with the imaginary component surpassing this value. Whether this stems from an artefact or is the result of a stronger concentration in a single pixel cannot be said considering the peak is narrower than in Fig. 6.4 but there are some slight artefacts around the peak.

6.2. WS$_2$ Nanotube

In the beginning of this work it was envisaged to produce an experimental proof of concept to whether improved tomographic reconstruction models would allow atomic resolution in Electron Holographic

Tomography. To this end, contact was made to Prof. Reshef Tenne of the Weizmann Institute, Revohot, Israel, who contributed greatly to the field of inorganic nanotubes[1]. Nanotubes are very appropriate specimen for high resolution tomography, since they are very stable (mechanically and in terms of bonding of the individual atoms) and offer a suitable size (tens of nanometers) for a relatively small thickness (in terms of atomic layers), furthermore they are easy to mount. When investigating nanoparticles the problem of mounting and surface mobility of the atoms might pose considerable challenges.

The Weizmann Institute graciously supplied us with two powder samples of tungsten disulfide nanotubes. In preliminary studies, the tubes proved to be too large (diameter $\approx$80 nm) to fit in the field of view of electron holography at highest resolution ($\approx$40 nm). While smaller WS$_2$-nanotubes have been synthesised[2], we were unable to obtain an usable sample. The WS$_2$-nanotube shall nonetheless serve as the test specimen for this work. For the moment, the tilt series cannot be recorded experimentally and, therefore, has to be simulated.

6.3. Specimen

Planar tungsten disulfide forms a layer of tungsten between two sulfur layers with a hexagonal cell structure[3]. These WS$_2$-layers are then stacked together with an approximate layer distance of 6.18 Å due to, compared to the in-layer binding, weak van-der-Waals binding. The chirality of the layer is chosen close to the armchair-configuration. The n-m-notation of carbon nanotubes will be used. Going from a 44-44-layer (diameter: 7.7 nm), using the aforementioned layer distance of 6.18 Å, the layers 53-49 and 65-51 (with the diameters 8.9 nm and 10.1 nm) where chosen as can be seen in Fig. 6.8. For a given length of 15 nm this gave $\approx$44 000 atoms in the simulation.
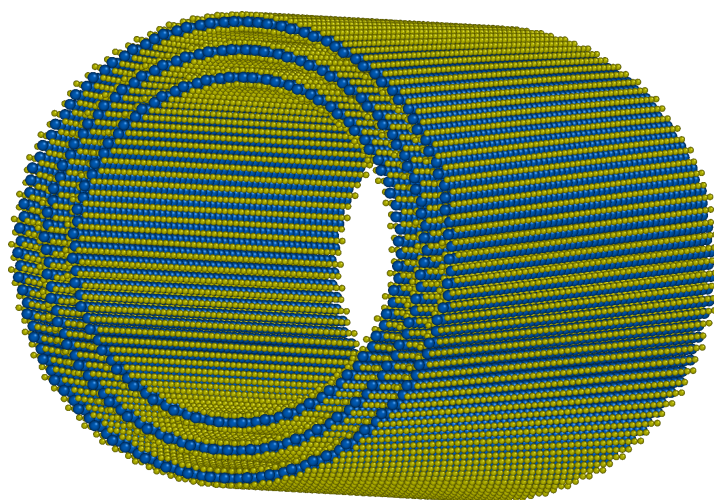

Figure 6.8.: Tungsten disulfide nanotube with three layers of nearly armchair-configuration; the outer diameter is $\approx$10 nm

---

[1][24] Panchakarla et al. 2014 . "Nanotubes from Misfit Layered Compounds".

[2][4] Brüser et al. 2014 . "Single- to Triple-Wall WS2 Nanotubes by [...] Plasma Ablation [...]".

[3][30] Seifert et al. 2000 . "On the electronic structure of WS2 nanotubes".

6.4. Simulation and Preparation

This tube was embedded in a $20\,\text{nm} \times 20\,\text{nm}$ field of view and the scattering of a plane wave (sampled with $2048\,\text{px} \times 2048\,\text{px}$) was simulated using the multislice algorithm for the full tilt range of θ at 3° intervals. The waves are focussed on the middle of the specimen so the overall extent of the Fresnel kernel (which grows in radius with the square root of the propagation distance) is minimal.
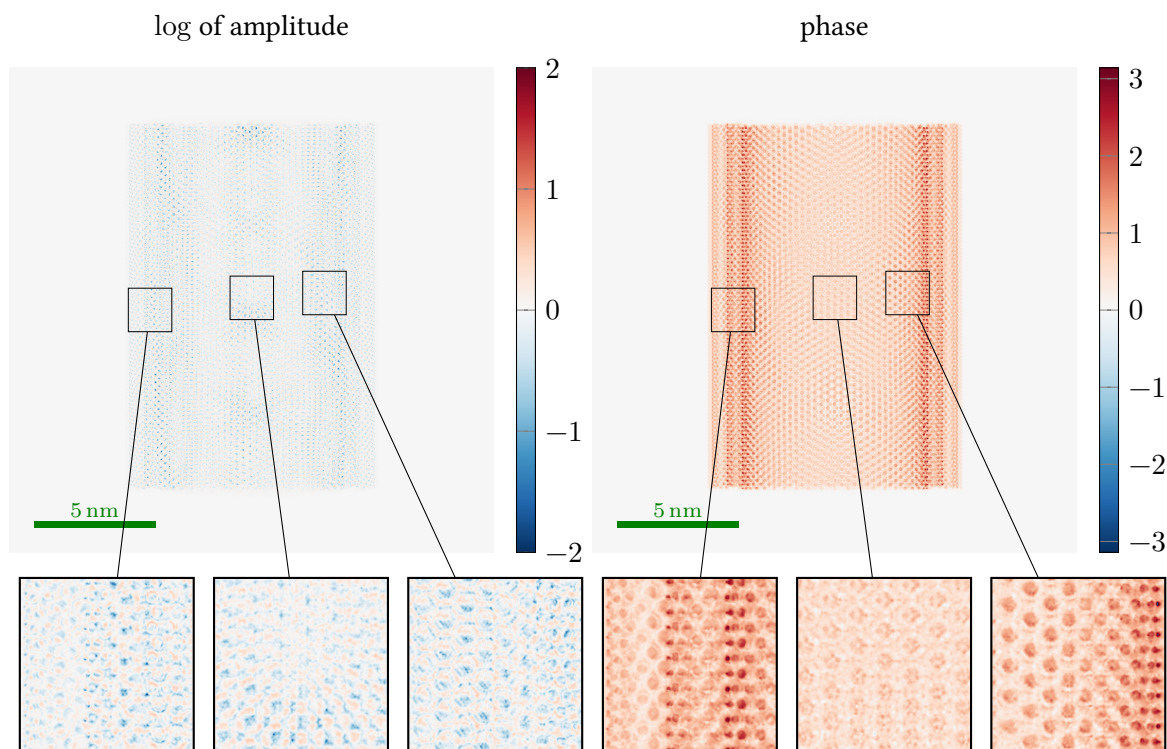

Figure 6.9.: Object exit wave of tungsten disulfide nanotube backpropagated to the middle of the specimen with the full resolution of the multislice simulation. The range of the amplitude has been cropped to exclude some outliers.

The scattering effect of the specimen is rather strong (strong phase and amplitude signal) at the sides, where the walls are viewed edge-on and have the highest effective thickness. This is especially strong at the inner wall, which is also imaged in a systematic direction, where the atoms form a (curved) column. Furthermore, due to the choice of focal plane, some atoms are in-focus (as in the left sub-pictures) and appear quite clearly in the phase. Otherwise, there are long-range variations, especially in the amplitude, which can be attributed to a Moire pattern effect; the lattice structures of the walls are rotated against each other (the outer wall 4° and the middle wall 1.3° against the inner one) so there are regions where the atoms are aligned behind each other and some where they are not. The range of the phase values is inside a 2π-interval so no unwrapping was necessary.
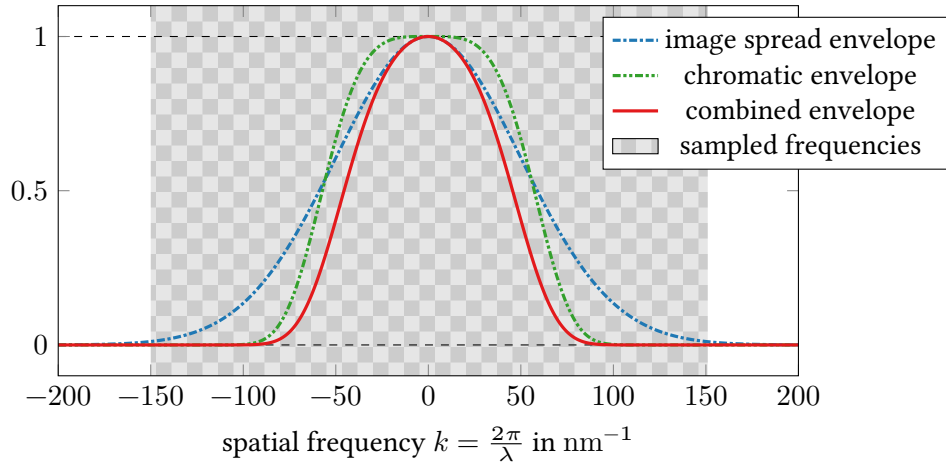
Figure 6.10.: Dampening envelopes for the parameters $C_C = 2.3\,\text{mm}$, $\sigma_E = 0.3\,\text{eV}$ and $\sigma_i = 20\,\text{pm}$. The highlighted area covers the spatial frequencies sampled by the reduced wave.

The resolution of this wave (as implied by the sampling) of $\approx 10\,\text{pm}$, which is well below the limit of any TEM today. The mechanisms in the TEM that limit the resolution are chiefly called aberrations and are modelled as envelope functions for the spectrum of the wave (or the corresponding point spread functions in real space). Today, the fundamentally limiting aberrations are image spread and the chromatic aberration (at least for TEMs without a achromatic corrector, of which there are currently two) and only those will be modelled here. For that, the following, rather generous, microscope parameters are assumed: $C_C = 2.3\,\text{mm}$, $\sigma_E = 0.3\,\text{eV}$[4] and $\sigma_i = 20\,\text{pm}$[5]. Applying these envelopes and selecting the centre $15\,\text{nm} \times 15\,\text{nm}$, the needed sampling (as indicated in Fig. 6.10) can be reduced to $360\,\text{px} \times 360\,\text{px}$.
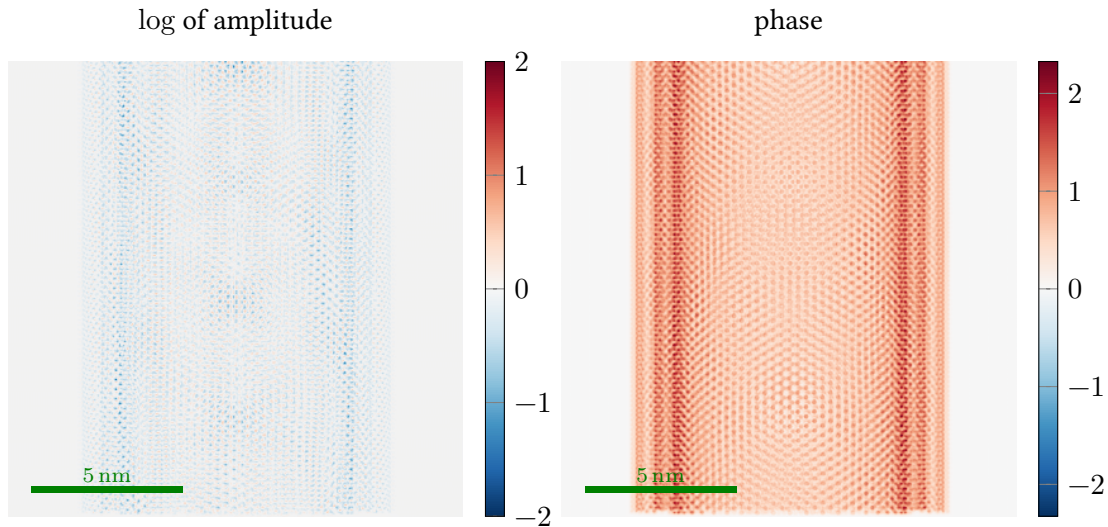
Figure 6.11.: Simulated image wave of tungsten disulfide nanotube with reduced resolution and field of view. No noise was simulated. The focal plane has been set to the middle of the specimen.

---

[4]These are generic values for a new aplanatic corrected TEM and for a cold field emission gun.
[5]as measured in [13] Haigh et al. 2009 . "Optimal tilt magnitude determination for [...] wave function reconstruction".

This is the starting point for the reconstructions. But beforehand we would like to discuss the validity of the Rytov approximation in terms of this problem.

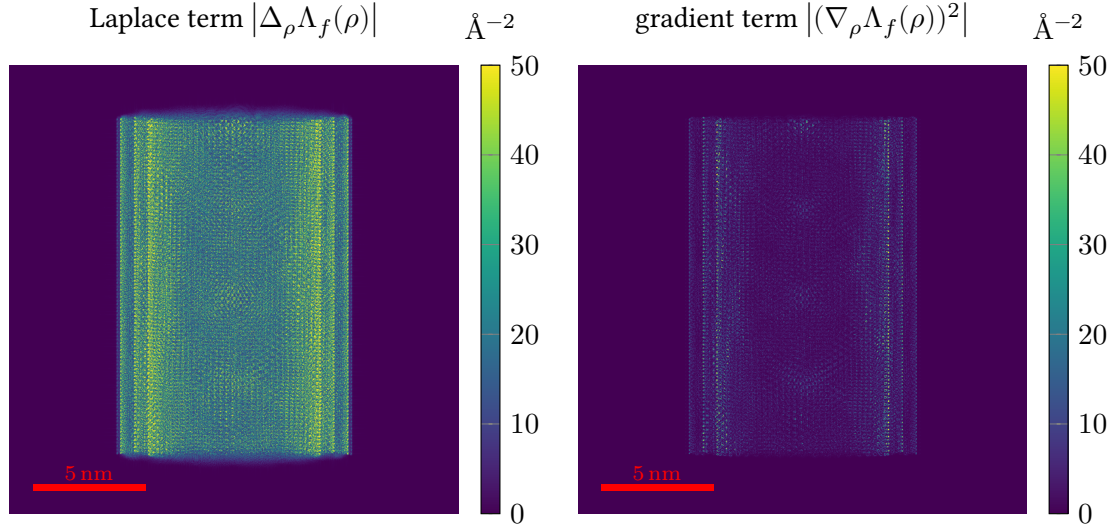### 6.4.1. Validity of the Rytov Approximation

Figure 6.12.: Comparison of the magnitude (in Å^{-2}) of the two relevant terms of the Rytov approximation Eq. (2.16). The terms are evaluated in the image wave, which had been propagated back into the middle of the specimen. The colour scale is cropped to the relevant region.

The rather unmotivated neglection of the quadratic gradient term from Eq. (2.16) should at least be evaluated numerically for the given simulated wave function to give some indication whether the Rytov approximation may be applicable. The Laplace term is overall larger than the gradient term but at the sides, where the atoms are in focus, and especially at the inner wall, which is viewed along atom columns there, the gradient term becomes very large. Because of this, the gradient term becomes larger than the Laplace term for small regions of wave and it is not possible to say at this point whether that has a significant influence. Imaging along atom columns might exacerbate the size of the gradient term, but due to the other walls that cannot be deduced clearly from this image.

The application of the Rytov approximation is, therefore, somewhat justified but not in so far as the presence of artefacts can be excluded categorically.
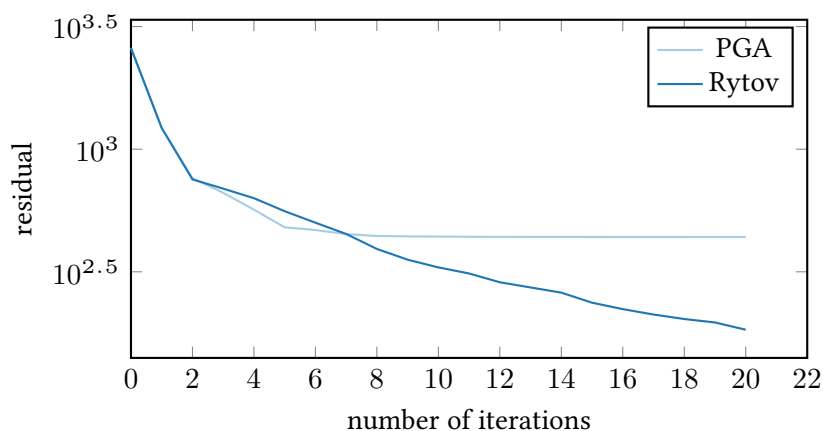
6.5. Reconstruction

Figure 6.13.: Residual (the value of the objective function) during 20 iterations of the LSQR algorithm.

The convergence of the residual Fig. 6.13 shows that, similar to the single atom cases, the Rytov approximation converges in some region slightly slower than the PGA but matches the tiltseries much closer. While he PGA flattens out after just six iterations the Rytov approximation overtakes it at that point and has gained half an order of magnitude in the residual by iteration 20. This is, however, at the expense of over 40 times the time needed per iteration than the PGA (27.3 min vs. 0.6 min, with kernels of the size 21.4 GiB and 0.8 GiB).
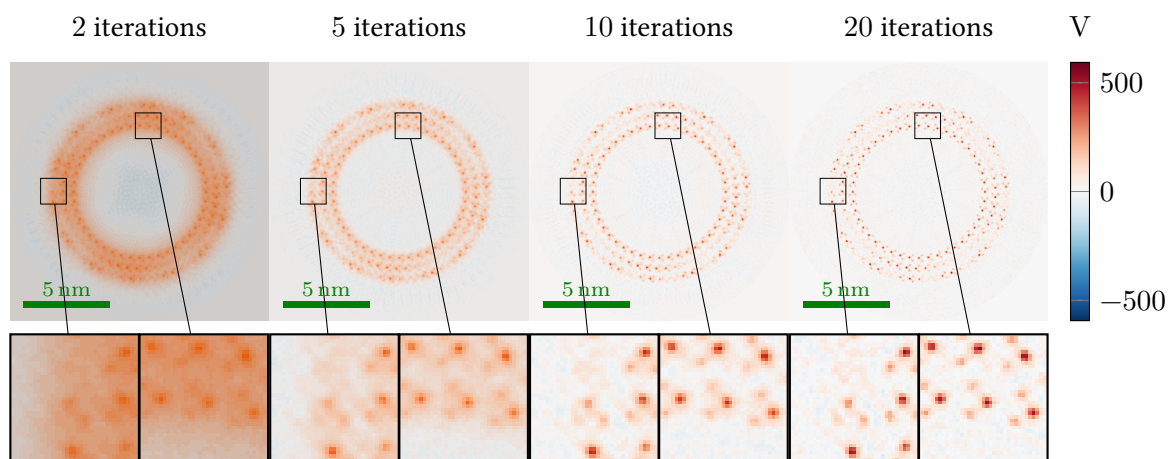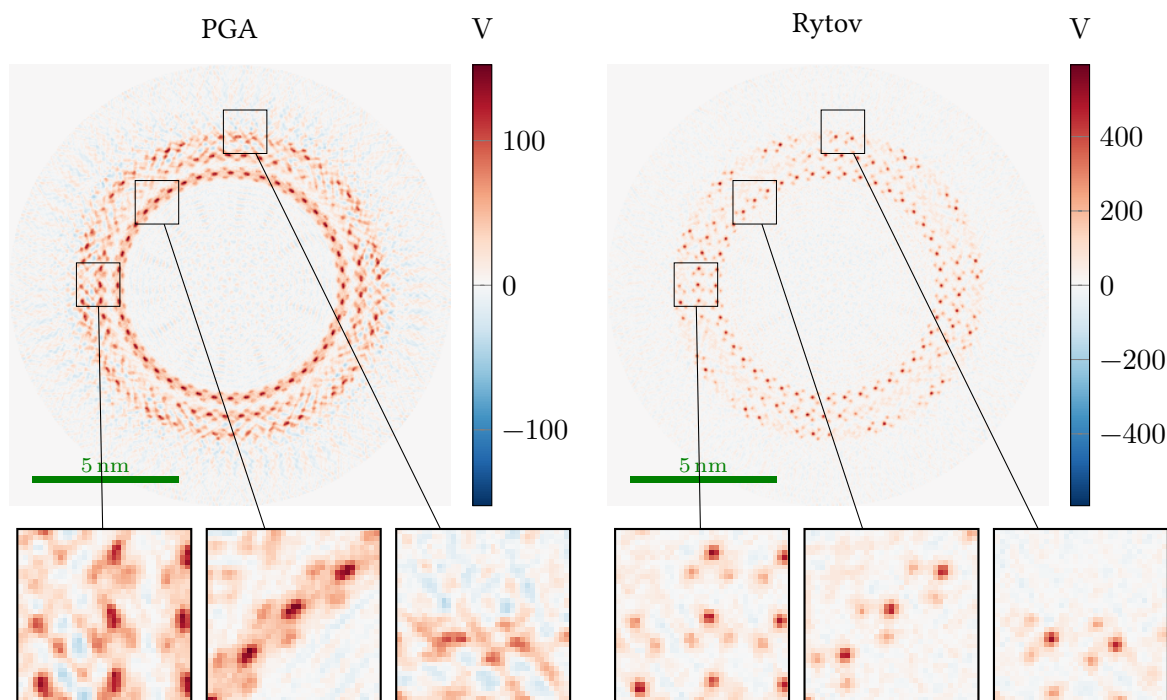

Figure 6.14.: Cross section (perpendicular to the tilt axis) of the real component of the reconstructed (effective) potential of a nanotube after 5, 10, 15 and 20 iterations using the Rytov method. This shows the appearance of the features as the algorithm converges.

In the first iterations the outer shape of the sample is reconstructed and the atomic structures begin to emerge. The constant background is a distinct artefact, which is eliminated by the tenth iteration and and the atoms progressively sharpen till the 20th iteration and possible further on.

6.6. Evaluation

Figure 6.15.: Cross section (perpendicular to the tilt axis) of the real component of the reconstructed (effective) potential of a nanotube after 20 iterations. The Rytov approximation shows much narrower peaks than the PGA and a much lower level of artefacts.

Both methods are capable of resolving the atoms of the tubes, not only the tungsten atoms, but also the sulfur atoms. The Rytov approximation resolves them as near point-like peaks and the PGA as elongated blobs and strong artefacts (especially in the outer wall, see the left and the right subpicture) make some atoms in the PGA not discernible any more. The Rytov approximation clearly fares much better as every single atom peak has roughly equal height (for tungsten and sulfur respectively), which is for tungsten with $\approx 600\,\mathrm{V}$ slightly lower, than the value from the single atom simulations. The PGA is nowhere near that value, but as the peaks are wider they may have the same integrated potential.
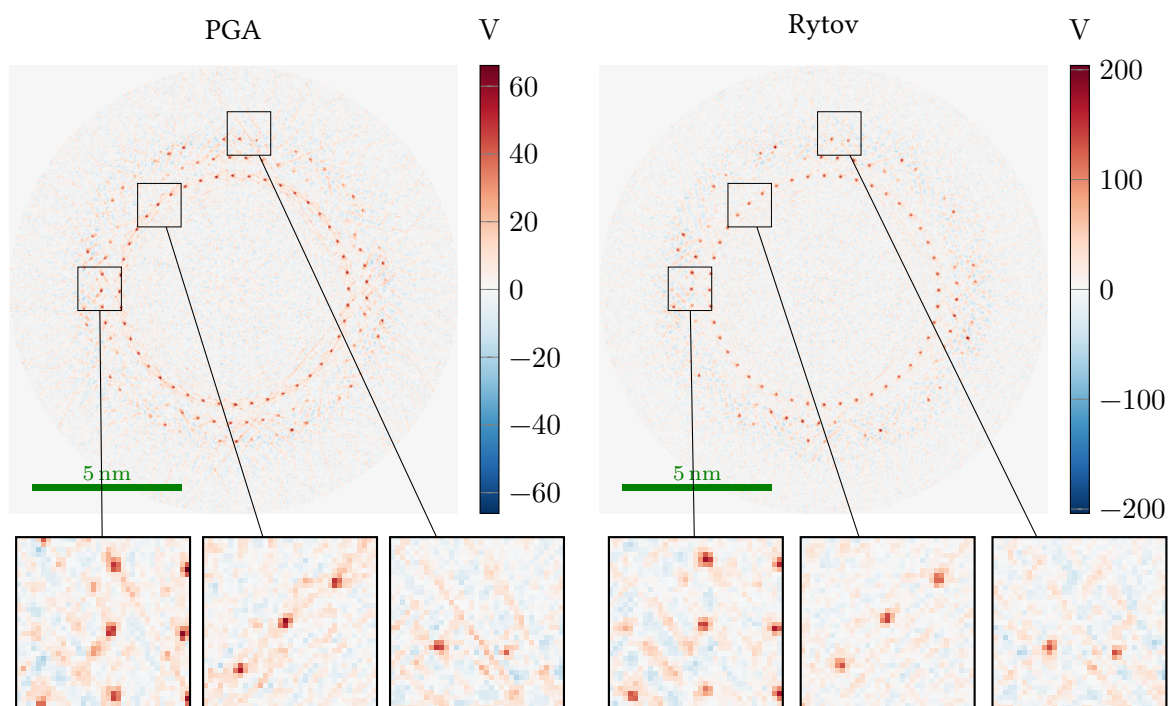
Figure 6.16.: Cross section (perpendicular to the tilt axis) of the imaginary component of the reconstructed (effective) potential of a nanotube after 20 iterations. Both methods perform nearly equally well, apart from a slightly higher level of artefacts in the PGA and a much higher absolute signal in the Rytov approximation.

The imaginary part of the reconstructed potential exhibits narrow tungsten atom peaks for both methods. The PGA has a slightly higher level of artefacts and much lower signal than the Rytov, where some peaks have higher values than the reference case Fig. 6.1 (which could bee seen at the off-centre atom as well).

For a better (or rather a different) evaluation the reconstructed potential is interpolated onto a cylinder surface with a certain radius so the tube is unrolled into a sheet, an azimuthal cross section so to say.
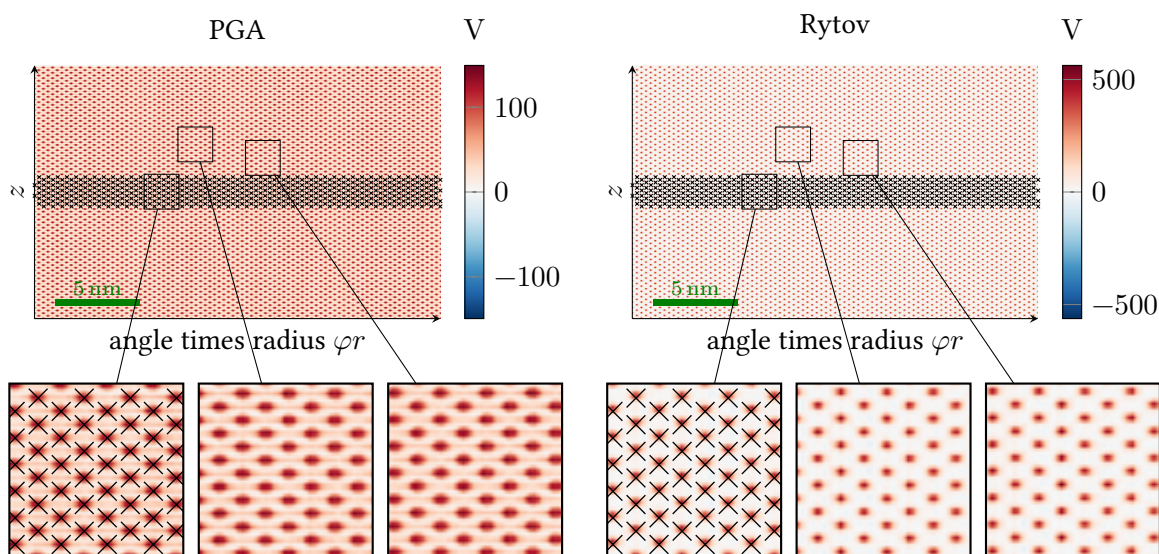
Figure 6.17.: Real component of the reconstructed potential (after 20 iterations) unrolled at the tungsten layer of the inner wall (a diameter of 7.7 nm). The crosses mark the ideal atom positions in a 2 nm wide band.

Again, the Rytov approximation delivers well condensed high peaks for the atoms, whereas the PGA only has considerably larger and elongated peaks. The atoms are at the correct positions and the chirality of the lattice is certainly visible in both.
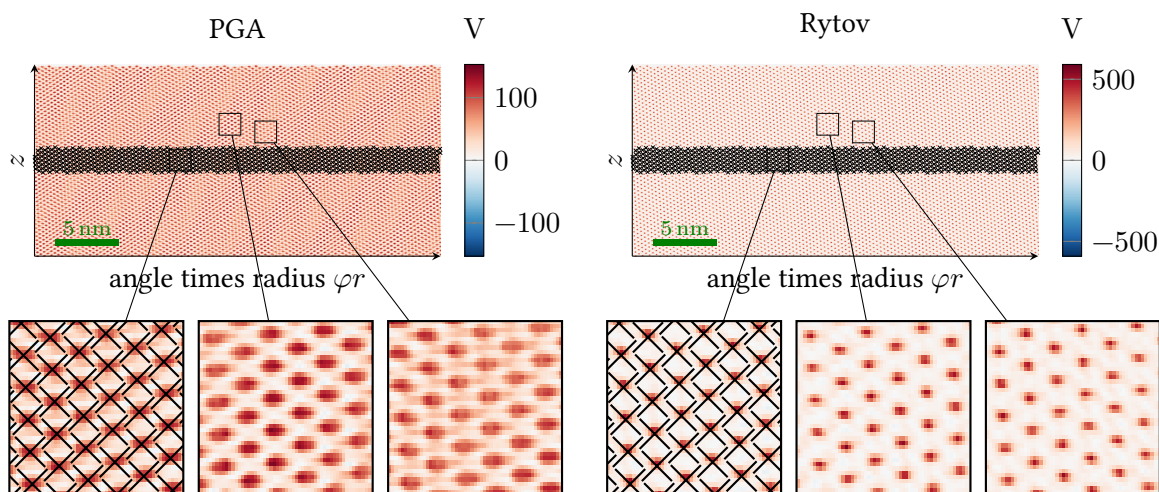

Figure 6.18.: Real component of the reconstructed potential (after 20 iterations) unrolled at the tungsten layer of the outer wall (a diameter of 10.4 nm). The crosses mark the ideal atom positions in a 2 nm wide band.

The PGA raises some problems in the outer wall with long range variations in the reconstructed potential. In some regions (right subpicture) the atoms have considerably lower signal than in others (middle subpicture) and a distorted peak shape. Nonetheless, all atoms are reconstructed at the correct positions. The Rytov approximation has no such problems. Overall, the maximum signal levels are comparable to the inner wall.
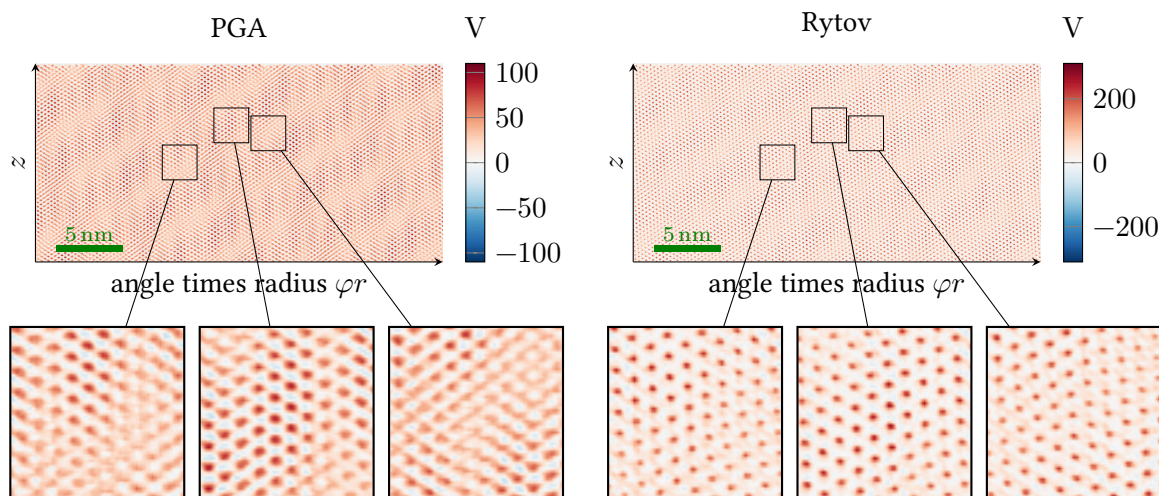
Figure 6.19.: Real component of the reconstructed potential (after 20 iterations) unrolled at the inner sulfur layer of the outer wall (a diameter of $10.1\,\mathrm{nm}$).

The sulfur atoms are much harder to reconstruct than the tungsten atoms, owing to their charge of 16 versus 74 for tungsten and the reduced height of the potential of about 2.5 (using the Wentzel potential). Nevertheless, the Rytov approximation shows some variations in the height of the peaks and some are a little to high compared to the tungsten atoms. The PGA is much worse, with some atoms barely discernible from the background and some much too high compared to the tungsten atoms. The inner layers of sulfur atoms are overall reconstructed better.

6.7. Effects of Noise

To study the reconstruction under near-to-experimental conditions some noise was added to the simulated waves using the Lenz model[67], which comes down to a Gaussian distribution centred around the ideal value (for both the real and the imaginary component) of every pixel with a standard deviation σ that depends on the intensity in that pixel. For these waves an electron dose was assumed that resulted in a mean σ of ≈ 0.09 relative to the mean amplitude.

[6][18] Lichte et al. 1987 . "Electron Noise In Off-Axis Image Plane Holography.".
[7][28] Röder et al. 2014 . "Noise estimation for off-axis electron holography".

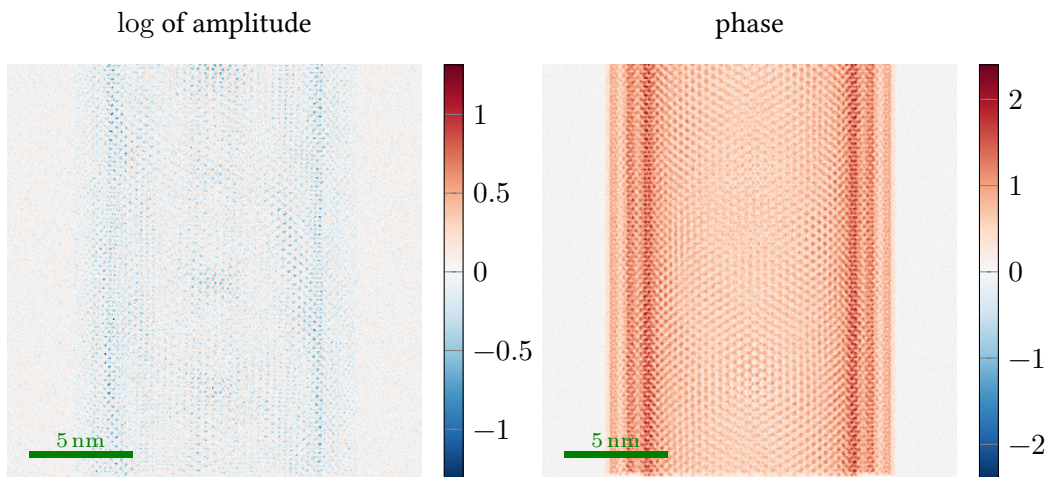log of amplitude   phase

Figure 6.20.: Object exit wave function, focussed on the middle of the nanotube with noise of a mean standard deviation of ≈0.09 using the Lenz model.

The influence of the noise is more noticeable in the amplitude than in the phase. This is partially caused by the logarithmic scaling of the amplitude but mainly because the additive noise is transferred less in the phase and more in the amplitude of the wave function.
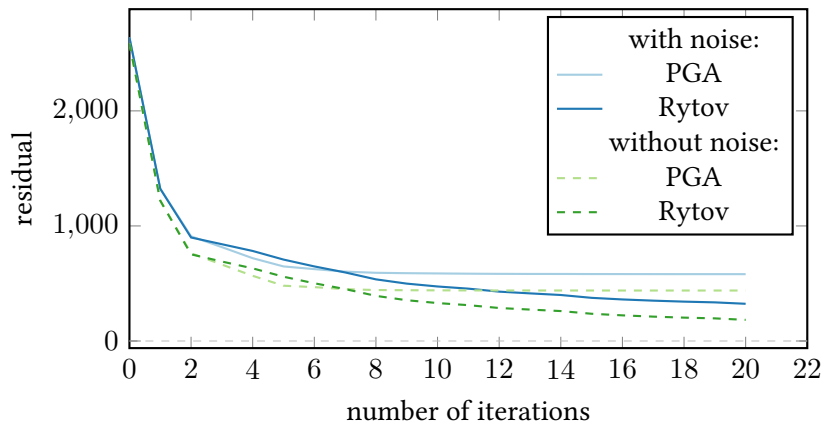

Figure 6.21.: Residual (the value of the objective function) during 20 iterations of the LSQR algorithm for the tilt series with added noise. For comparison the lines from Fig. 6.13 are included as well.

The noise causes a mere displacement of the convergence curves, considering that the distance between the noised and the noise-free reconstructions is the same for the two methods. It can be concluded, that for the moderate noise levels assumed here, the Rytov approximation is roughly equally well conditioned as the PGA and, although noise amplification is always a problem, it will not be worse than with conventional tomography.
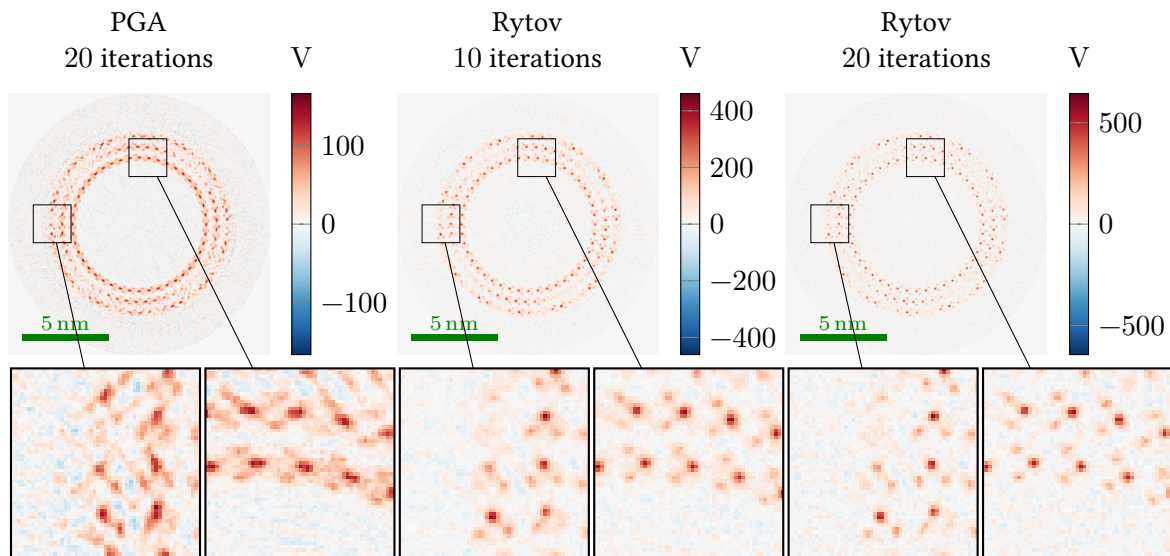
Figure 6.22.: Cross section (perpendicular to the tilt axis) of the real component of the reconstructed (effective) potential of a nanotube after 20 iterations of PGA, 10 iterations and 20 iterations of Rytov.

The reconstruction is seemingly unaffected by the noise, even the height of the atoms peaks is about the same as in the noise-free case.
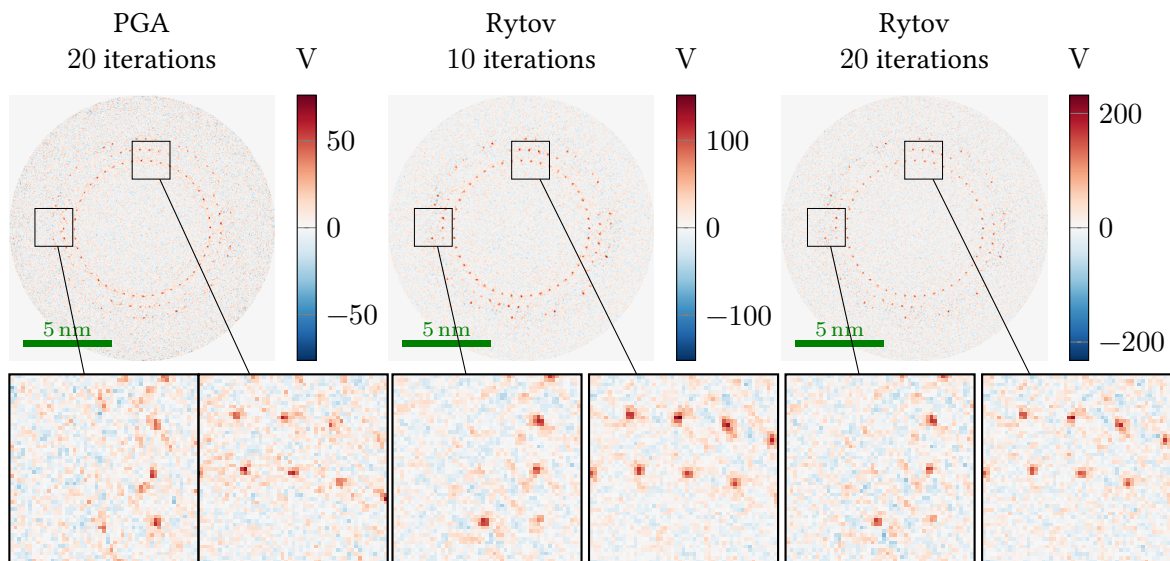

Figure 6.23.: Cross section (perpendicular to the tilt axis) of the imaginary component of the reconstructed (effective) potential of a nanotube after 20 iterations of PGA, 10 iterations and 20 iterations of Rytov.

In the amplitude the noise transfer is noticeable with a higher level of background noise, although there is seemingly little difference between the two methods here as well.

Conclusively, the addition of moderate noise does not affect the reconstruction in any serious way and there is no reason, on these grounds, to expect the noise amplification behaviour of the Rytov approximation based tomography to be much different from that of conventional tomography.

7. Summary and Conclusion

In this work, the inclusion of Fresnel propagation into the model used for electron holography is proposed. In combination with the Rytov ansatz to the perturbation problem for the Helmholtz equation (which adequately describes electron scattering under some assumptions) it yields a linear model between the electron wave and a given object potential. Or rather, it yields three approximations for three scattering regimes: a complex-valued phase grating approximation for axial scattering, Fresnel propagation for paraxial scattering and Huygens propagation for the non-paraxial scattering. The Rytov approximation pertains to an unconventional manifestation of the neglection of multiple scattering (the neglection of the gradient-term in the Rytov approximation, see Section 2.2 and Section 6.4.1), interestingley by neglecting a purely wave-dependent term (independent from the specimen potential). This could give insight into the trust region of the single-scattering approximation, maybe leading to practical a priori criteria. The Rytov approximation itself is not the end for this approach, the neglected term could be approximated further (e.g. by treatment as a perturbation), which may lead to still better approximations.

Numerically efficient algorithms for the computation of the Fresnel convolution kernel, the construction of the projection tensor and the tensor multiplication have been implemented. The size of the tensor is optimised to its minimal size by using translation symmetries of the discrete representation of the convolution kernel. Thus, it was possible to keep it small enough to fit the tensor for a realistic example ($(360\,\mathrm{px})^3 \to (360\,\mathrm{px})^2 \times 120$ tilt angles) into the memory of a conventional desktop machine (with $32\,\mathrm{GiB}$ of RAM). The code itself is as optimised as it can be using generic methods and relying on the automatic optimisation routines of a modern compiler.

For the simulations of non-periodical (rather) large specimen a multislice algorithm for large fields of view was implemented, using a number of adaptations for such problems (bunching the atoms along z and computing the transmission function in subregions only), and, ultimately, moving it onto the graphics card. The focus on parallel computing and the reasonably good general computing performance of even a consumer grade graphics cards makes the speedup over the CPU quite impressive.

For the single atom specimen the general considerations of moving towards a Fresnel kernel for tomography are exemplified. It converges slightly slower than the conventional reconstruction method for both cases but eventually reaches a closer fit. In the case of the centred atom this stems from artefacts, but in the other case the atom is reconstructed very sharp, whereas the conventional tomography only produces a wide and elongated peak.

A tungsten disulfide nanotube with three walls and a diameter of $\approx 10\,\mathrm{nm}$ is simulated and reconstructed both with the phase grating approximation (PGA) and the paraxial Rytov approximation. The latter converges considerably faster and minimises the residual further than the former. In the reconstruction all atoms are appear as near point-like peaks, even the sulfur atoms (though they show some peak-height variations in the outer layers). Thus, it performs much better than the PGA whose atom peaks are much wider and the level of artefacts in the reconstruction is noticeably high, to the point that some atoms can barely be distinguished from the background, especially sulfur atoms.

The noise transfer characteristics of the proposed technique seem to be the same as those of conventional tomography, at least for the tested case of moderate Poisson noise. This bodes well for an experimental realisation; although other problems, such as misalignment, will play a significant

role there.

The Rytov approximation shows a remarkable capability to incorporate propagation effects into a linear model. The step from traditional linear models, like the Born Approximation, which are linear in the wave itself, to a model, which is linear in the logarithm of the wave, seems to be very promising as a linear model for dynamic scattering. The Rytov approximation, particularly the neglected term, should be investigated further to ascertain the characteristics (e.g. how multiple scattering effects look like in the Rytov approach) and the limits of this approximation. Furthermore, the neglected term itself can be approximated further: as a perturbation, in a linearised form or in a non-linear tomographic reconstruction technique.

It was, regrettably not possible to obtain a suitable sample for an experimental proof-of-concept, which would certainly give these gains more weight, so this will certainly be the next step along this path.

8. Acknowledgement

I am glad to have gotten into an environment where I could reap so much benefit from the people around me as I had at the Triebenberg laboratory. I experienced a climate of honest dedication to a further understanding of physics, a will to teach and a will to learn, often simultaneously. A good argument, irrespective of its wielder, was always welcome and a bad argument diligently righted, for it may still contain a valid point or an edifying perspective. In my works I was given the freedom that a scientist wishes for, with all its foibles and traps. This caused some grievances, which may have been net losses to science (of which I am not sure) but have certainly been instructional experiences for me.

My prime thanks goes to my mentor Axel Lubk, whose tireless curiosity and willingness to explain has made this thesis a very educational and strangely humbling experience. The other members of the Triebenberg laboratory, though few in number, offered a wide spectrum from the instruments, their effective use, phenomena, techniques and theories, and were always readily available to help, explain or discuss the physics of this world. This is, in no small parts, the achievement of Prof. Hannes Lichte, always focussed on helping "his" scientists to reach their scientific best and kindly took on the part of supervising this work. Additionally I would like to thank Prof. Christian Schroer for readily consenting to writing the second review of this masters thesis.

I would like to thank the members of the group of Prof. Reshef Tenne, who kindly supplied us with some tungsten disulfide nanotubes which were unfortunately too large.

In a more less direct sense and as almost all of the software used was free I want to express my appreciation for all the developers of free software and all the research grants, donations and companies that fund their work, if it is paid at all.
(

Bibliography

[1] Maya Bar Sadan et al. "Toward Atomic-Scale Bright-Field Electron Tomography for the Study of Fullerene-Like Nanostructures". In: *Nano Lett.* 8.3 (Mar. 1, 2008). ISSN: 1530-6984. DOI: 10.1021/nl073149i. (Visited on 09/19/2014).

[2] W. Van den Broek et al. "A model based atomic resolution tomographic algorithm". In: *Ultramicroscopy* 109.12 (Nov. 2009). ISSN: 0304-3991. DOI: 10.1016/j.ultramic.2009.08.003. (Visited on 09/22/2014).

[3] Wouter Van den Broek and Christoph T. Koch. "Method for Retrieval of the Three-Dimensional Object Potential by Inversion of Dynamical Electron Scattering". In: *Phys. Rev. Lett.* 109.24 (Dec. 10, 2012). DOI: 10.1103/PhysRevLett.109.245502. (Visited on 09/22/2014).

[4] Volker Brüser et al. "Single- to Triple-Wall WS2 Nanotubes Obtained by High-Power Plasma Ablation of WS2 Multiwall Nanotubes". In: *Inorganics* 2.2 (Apr. 29, 2014). DOI: 10.3390/inorganics2020177. (Visited on 01/14/2016).

[5] F. R. Chen et al. "3D reconstruction of nanocrystalline particles from a single projection". In: *Micron* 68 (Jan. 2015). ISSN: 0968-4328. DOI: 10.1016/j.micron.2014.08.009. (Visited on 10/13/2014).

[6] A. M. Cormack. "Representation of a Function by Its Line Integrals, with Some Radiological Applications". In: *Journal of Applied Physics* 34.9 (Sept. 1, 1963). ISSN: 0021-8979, 1089-7550. DOI: 10.1063/1.1729798. (Visited on 09/22/2014).

[7] Heinz W. Engl et al. *Regularization of inverse problems*. Kluwer Acad. Publ., 1996. ISBN: 0-7923-4157-0.

[8] J. R. Fienup and C. C. Wackerman. "Phase-retrieval stagnation problems and solutions". In: *J. Opt. Soc. Am. A* 3.11 (Nov. 1986).

[9] R. W. Gerchberg and W. O. Saxton. "A practical algorithm for the determination of phase from image and diffraction plane pictures". In: *Optik* 35.2 (1972).

[10] Dennis C. Ghiglia and Mark D. Pritt. *Two-Dimensional Phase Unwrapping*. Wiley, 1998.

[11] Bart Goris et al. "Atomic-scale determination of surface facets in gold nanorods". In: *Nat Mater* 11.11 (Nov. 2012). ISSN: 1476-1122.

[12] Max Haider et al. "A spherical-aberration-corrected 200 kV transmission electron microscope". In: *Ultramicroscopy* 75.1 (Oct. 1, 1998). ISSN: 0304-3991. DOI: 10.1016/S0304-3991(98)00048-5. (Visited on 12/15/2014).

[13] S. J. Haigh et al. "Optimal tilt magnitude determination for aberration-corrected super resolution exit wave function reconstruction". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1903 (Sept. 28, 2009). ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2009.0124. (Visited on 03/18/2016).

[14] G. N. Hounsfield. "Computerized transverse axial scanning (tomography): Part 1. Description of system". In: *The British Journal of Radiology* 46.552 (1973). DOI: 10.1259/0007-1285-46-552-1016.

Bibliography

[15] A.C. Kak et al. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988. ISBN: 978-0-87942-198-4.

[16] J. Kirkland. *Advanced Computing in Electron Microscopy*. Journal for the Study of the. Springer US, 1998. ISBN: 978-0-306-45936-8.

[17] Jonas Krehl and Axel Lubk. "Prospects of linear reconstruction in atomic resolution electron holographic tomography". In: *Ultramicroscopy* 150 (Mar. 2015). ISSN: 0304-3991. DOI: 10.1016/j.ultramic.2014.11.026. (Visited on 03/31/2015).

[18] H. Lichte et al. "Electron Noise In Off-Axis Image Plane Holography." In: *Optik* 77.3 (1987).

[19] Alfred Karl Louis. *Inverse und schlecht gestellte Probleme*. Teubner, 1989. ISBN: 3-519-02084-X.

[20] Paul A. Midgley and Rafal E. Dunin-Borkowski. "Electron tomography and holography in materials science". In: *Nat Mater* 8.4 (Apr. 2009). ISSN: 1476-1122. DOI: 10.1038/nmat2406. (Visited on 09/22/2014).

[21] Gottfried Möllenstedt and Herbert Wahl. "Elektronenholographie und Rekonstruktion mit Laserlicht". In: *Naturwissenschaften* 55 (1968).

[22] P. Müller et al. "The Theory of Diffraction Tomography". In: *ArXiv e-prints* (July 2015).

[23] Christopher C. Paige and Michael A. Saunders. "LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares". In: *ACM Trans. Math. Softw.* 8.1 (Mar. 1982). ISSN: 0098-3500. DOI: 10.1145/355984.355989. (Visited on 09/22/2014).

[24] Leela S. Panchakarla et al. "Nanotubes from Misfit Layered Compounds: A New Family of Materials with Low Dimensionality". In: *J. Phys. Chem. Lett.* 5.21 (Nov. 6, 2014). ISSN: 1948-7185. DOI: 10.1021/jz5016845. (Visited on 01/14/2016).

[25] L. M. Peng et al. "Robust Parameterization of Elastic and Absorptive Electron Atomic Scattering Factors". In: *Acta Crystallographica Section A Foundations of Crystallography* 52.2 (Mar. 1, 1996). ISSN: 01087673. DOI: 10.1107/S0108767395014371. (Visited on 03/18/2016).

[26] Axel Rother et al. "The statistics of the thermal motion of the atoms during imaging process in transmission electron microscopy and related techniques". In: *Ultramicroscopy* 109.2 (Jan. 2009). ISSN: 0304-3991.

[27] E. Ruska. "Das Entstehen des Elektronenmikroskops und der Elektronenmikroskopie". In: *Phys. Bl.* 43.7 (July 1, 1987). ISSN: 1521-3722. DOI: 10.1002/phbl.19870430729. (Visited on 01/19/2016).

[28] Falk Röder et al. "Noise estimation for off-axis electron holography". In: *Ultramicroscopy* 144 (Sept. 2014). ISSN: 0304-3991. DOI: 10.1016/j.ultramic.2014.04.002. (Visited on 10/10/2014).

[29] O. Scherzer. "Über einige Fehler von Elektronenlinsen". In: *Z. Physik* 101.9-10 (Sept. 1936). ISSN: 0044-3328. DOI: 10.1007/BF01349606. (Visited on 01/27/2016).

[30] Gotthard Seifert et al. "On the electronic structure of WS2 nanotubes". In: *Solid State Communications* 114.5 (Apr. 11, 2000). ISSN: 0038-1098. DOI: 10.1016/S0038-1098(00)00047-8. (Visited on 01/14/2016).

[31] Stephan Uhlemann et al. "Thermal Magnetic Field Noise Limits Resolution in Transmission Electron Microscopy". In: *Phys. Rev. Lett.* 111.4 (July 2013).

[32] Sandra Van Aert et al. "Three-dimensional atomic imaging of crystalline nanoparticles". In: *Nature* 470.7334 (Feb. 17, 2011). ISSN: 0028-0836. DOI: 10.1038/nature09741. (Visited on 09/22/2014).

[33] Dirk Van Dyck et al. "'Big Bang' tomography as a new route to atomic-resolution electron tomography". In: *Nature* 486.7402 (June 14, 2012). ISSN: 0028-0836. DOI: 10.1038/nature11074. (Visited on 10/22/2014).

[34] A. Weickenmeier and H. Kohl. "Computation of absorptive form factors for high-energy electron diffraction". In: *Acta Crystallographica Section A Foundations of Crystallography* 47.5 (Sept. 1, 1991). ISSN: 01087673. DOI: 10.1107/S0108767391004804. (Visited on 09/22/2014).

[35] Andreas Weickenmeier. *FSCATT*.

[36] D. Wolf et al. "Electron holographic tomography". In: *Current Opinion in Solid State and Materials Science*. Electron Tomography 17.3 (June 2013). ISSN: 1359-0286. DOI: 10.1016/j.cossms.2013.05.002. (Visited on 09/22/2014).

[37] Emil Wolf. "Three-dimensional structure determination of semi-transparent objects from holographic data". In: *Optics Communications* 1.4 (Sept. 1, 1969). ISSN: 0030-4018. DOI: 10.1016/0030-4018(69)90052-2. (Visited on 03/21/2016).

A. Programme Overview

The program is contained as a set of interpreted scripts with a few segments of non-interpreted source code, which is compiled on demand by NumExpr, Cython or Reikna. The main structure is provided by Python and the whole programme or parts of it can be imported as a simple module. At the highest level it is split into four parts Scattering, Tomography, Mathematics and Utilities.

The source code is openly available at the GitHub repository `https://github.com/JKrehl/Electrons` and licensed under the MIT License.

Scattering Scattering simulation framework for multislice and multislice-like techniques.

> **Algorithms** The three implemented techniques (multislice, projection and single scattering) are implemented rather similar: the cloud of atoms is ordered in z direction, a list of operators (propagators, transmission functions, ...) is created, ordered in z and then successively applied to the given entry wave. They are abstracted against the choice of propagator, transmission function and atom potential parametrisation.

> **Operators** Implementations of both propagators and transmission functions as well as basic classes for the chaining of the operators.

>> **Propagators** Four different flavours of Fresnel propagation are implemented: three in Fourier space (one normal, one with padding and one normal on the GPU) and one in real space (via the convolution with a Laplace kernel).

>> **TransferFunctions** All transmission function generators model the phase shift of the atom's core and the thermal diffuse scattering smearing of that potential. The four flavors are the combinations of normal and the patched transmission function computation both on CPU and GPU.

Potentials base class for an atomically sparse potential (essentially a list of atoms) and atom potential parametrisations

> **WeickenmeierKohl** the parametrisation proposed by A. Weickenmeier and H. Kohl[1]

> **WeickenmeierKohl_FSCATT** the parametrisation implemented by A. Weickenmeier and H. Kohl in the FSCATT[2]

> **Kirkland** the parametrisation from E. Kirkland[3]

> **PengDudarev** the parametrisation of L.-M. Peng, G. Ren, S.L. Dudarev and M.J. Whelan[4]

Tomography A framework for tomographic reconstruction using a model of (symmetry reduced) precomputed tensors. A kernel of minimal size is computed and then inserted into a projector object, which mimics the reduced symmetries. This can then be inserted into a solver for linear least-squares problems.

---

[1][34] Weickenmeier and Kohl. 1991 . "Computation of absorptive form factors for high-energy electron diffraction".

[2][35] Weickenmeier . *FSCATT*.

[3]p 203ff [16] Kirkland. 1998 . *Advanced Computing in Electron Microscopy*.

[4][25] Peng et al. 1996 . "Robust Parameterization of Elastic and Absorptive Electron Atomic Scattering Factors".

Kernels The kernel class implements a memory managing object which handles arrays (and objects of any kind) in memory and on disk (and loaded on demand). Two kernels are implemented, the normal ray kernel for conventional tomography and a Fresnel propagator convolution kernel for more Fresnel propagator tomography. While the former is in-plane (4-dimensional) the latter is semi-in-plane (reduced 5-dimensional) as described in Chapter 4.

Projectors expansion of the kernels of minimal size to the full extent (in-plane or three-dimensional) by looping over the coordinates

Solvers adapted copys of the LSQR and LSMR implementations of SciPy

Mathematics collection of mathematical functions, some of which became superfluous during development

FourierTransforms abstraction layer to the pyFFTW package including functions for forward and inverse transforms, both in informatics and mathematical convention concerning centre position, calculation of inverse coordinates and centre shifting routines

CoordinateTrafos classes for 2D and 3D coordinate base transforms in many different notations (e.g. quarternions, $(n + 1)$-matrices, axis and angle)

LaplaceKernel creation of Laplacian convolution kernels with rotationally uniform response

VolumeTrafo volumetric coordinate transformation with linear interpolation

Interpolator2D linear interpolation of values from a regular gridded dataset

Utilities AtomsViewer simple OpenGL-based viewer for atom clouds

CompressedSparse first experimental steps towards converting the coordinate format of the sparse matrices into a partially compressed format

Physics constants and expressions for e.g. converting between energy and wave number

SlicePlayer view 3D array as 2D slices

Magic functions of mainly programming interest

Progress jupyter widgets progress bar

Colourmap symmetric norm for diverging colourmaps

ImExport export matplotlib images with colourbar, colour limits and extent for use in PGFPlots

B. Compendium of used Software

This shall serve as a short overview of software used in this work, to illustrate the range of necessary components, maybe inspire the reader searching for a tool (he could not imagine existing), a problem (he could not imagine solved) or a stroke of genius (he took for granted) and last but not least to acknowledge the work of their developers. The following are mentionable in their own right but not essential to this thesis: Jupyter, PyCharm, Eclipse, git, Gentoo Linux, Zotero, matplotlib, VisPy, the Intel Math Kernel Library, LaTeX, PGF, PGFPlots.

The Coding Stack

From high-level to low-level these tools allow to write fast, what is not worth optimising (as its complicated, abstract or run seldom) and to optimise what takes much time. This added versatility makes it much faster to write powerful code.

Python In the field of interpreted scripting languages Python has established itself as the mainstay in scientific and performance computing. It has an easy yet powerful syntax, easy interoperability down to its base-layer (C) and a huge collection of available libraries.

Cython An language extension for Python including C/C++-like directives including an automated build system (building libraries on demand). Interopability with both Python and native C/C++ makes it the perfect glue.

C/C++ De-facto standard for low-level programming (although others are persistent) very powerful in abstract programming (templates, the Standard Template Library, compiler directives, …). Here only needed to expose certain special capabilities to Cython.

Automatically Optimised Libraries

One main advantage of the higher calculation power of modern computers (from a scientific programmers perspective) is the sudden feasibility of self-optimisation. A lot of calculations can be expended on the computer optimising libraries for itself and for each other.

NumPy An extension of Python for working with n-dimensional arrays featuring a large collection of high-level, but highly-optimised, functions. It is a must for scientific computation, although it is interoperable with almost any other library using arrays thanks to the abstracted array interface of Python.

SciPy Collections of low-level-optimised functions and algorithms which where deemed to be of too narrow an interest to be included in NumPy.

NumExpr Automatic build system for short numerical expressions, which are compiled in the background and parallelised.

GCC The Gnu Compiler Collection includes compilers for, among others, C and C++, which are far developed and automatically optimise code on a machine level.

Interfaces

Interfaces to both the GPU and the data storage.

Reikna An automatic build and binding system for GPU-code as well as an abstraction layer coupled with a templating system for meta-programming for both PyOpenCL and PyCUDA, the Python bindings for the two big GPU computation systems OpenCL and Nvidia CUDA. While rather inelegant it allows uncomplicated access to the capabilities of the GPU, even the full capabilities when foregoing the interopability between PyOpenCL and PyCUDA.

h5py Interface to the hdf5-implementation of the Hierarchical Data Format, a general-purpose and widespread file format for storing complicated structures of large data sets.

pyFFTW Python bindings for the FFTW library for self-optimising Fast Fourier Transforms.