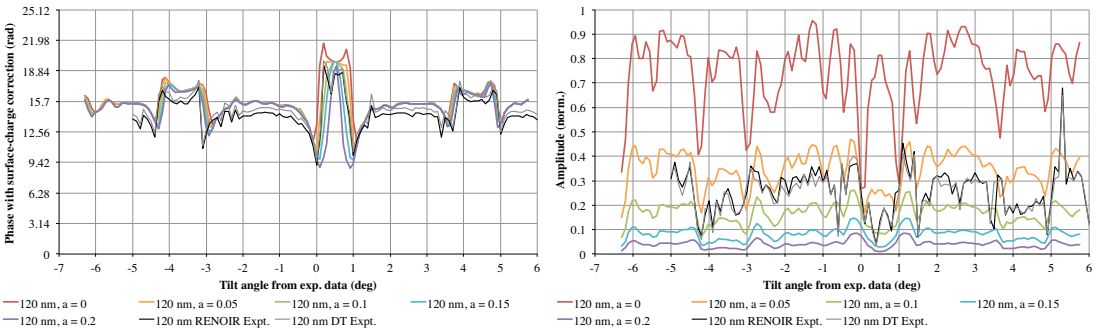
Factors Influencing Mean Inner Potentials As Studied Using Electron Holography and Density Functional Theory

Pennington, Robert Scott; Hansen, Jørn Otto Bindslev; Dunin-Borkowski, Rafal E.; Boothroyd, Chris; Wagner, Jakob Birkedal

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Factors Influencing Mean Inner Potentials As Studied Using Electron Holography and Density Functional Theory



Robert Scott Pennington

PhD Thesis

February 2012

Technical University of Denmark

Thoughts and Acknowledgements

It's a beautiful day.

I have watched the Center for Electron Nanoscopy grow and flourish, and it is an honor and a privilege to be the first PhD student to emerge – hopefully, Cen feels the same way, and hopefully I'll pass! From a handful of people and a new building and frustrating microscopes, I have watched us become and remain an established group with deep interests, a functioning building…and frustrating microscopes. Some things don't change, I suppose – at least the frustrations are less frequent. It's been easier than we thought, sometimes, and harder than we thought, sometimes, but I am proud, very proud, to have had a place here these past four years.

Much of the past four years of my life has been devoted to the production of this dissertation. There are many people I owe various debts to, for those myriad things they have done that I am thankful for. I am truly humbled by the support of multitudes, and I know I will inevitably omit someone – I ask his or her forgiveness.

First, I'd like to thank my family for their support and love. They inspired me to follow this path, many years ago, and their inspiration continues to shine brightly. I would be nowhere near where I am today without them. Dad, thank you - I know now what you meant about the PhD – some things don't change, although mine had more atoms, fewer stars. Mom, thank you – your sense of humor and support have been vital, and the stories from home are always welcome. Victoria, thank you – you've turned out pretty well, sister – I'm proud of you. And, grandparents, I'm very lucky to have you and your insight and wisdom.

Professionally, while I have many people to thank, the first words should be for my four supervisors. Rafal, it has truly been an honor and a privilege working with you – thank you for your supervision, discussions, and guidance throughout the process – you've helped immensely, and you have taught me so much. Jørn, thank you – you've been superlative in your unexpected late role as my thesis supervisor, and your assistance (with Danish – especially checking over the abstract! – and Japanese) is appreciated. Chris, thank you for all that I have learned (especially about the ATEM), for the feedback, the discussions, and the insightful humor. Jakob, you've been both friendly and helpful – thank you, too, for your assistance, especially with the ETEM.

My colleagues and friends at DTU Cen have been along for the ride, helping me retain my sanity – some are still with us, some have moved. Johan, we've shared an office (and musical tastes, and mutual incomprehension) for almost three years – thank you for being and remaining my friend, even after introducing me to the wonders of opera, and even as we wandered through the streets of Tokyo following my stupid guidebook. Linus, thank you for some very fun times, in Paris and Tokyo and Washington, while we set a new decadent baseline in the eternal search for hot chocolate. Filippo and Davide – thank you for the conversation and the food. Shima – thank you for being a voice of reason and sanity. Lionel – thank you for always being sunny. Thank you, Takeshi, for your help with the ATEM and for the laughs. My Cen officemates over the past few years – Alice, Lionel, Sadegh, Johan, Shima – it's been a pleasure – and thank you for putting up with me. Last, but far, far from least: Andy – you're a good and decent man, and thank you for your rock-solid support.

Several other professional colleagues have been helpful. Many of these are thanked in the text, where they contributed. However, others have contributed to my development, or hosted me, over the past few years, and I would be deeply remiss to not thank them. Thank you, Molly McCartney and Dave Smith, for introducing me to TEM and electron holography during my time at Arizona State University and leading me to this place. Thanks to Martin Hÿtch and the group at CEMES-CNRS for hosting me back in 2008. A very big thank you to Dave Cooper and the group at CEA-Grenoble, for hosting me for much longer in 2009 – thank you for both the science and the mountains, and everything in between. Joerg Jinschek, thank you for teaching me about Titan operations (and for your always-good conversations).

I am lucky enough to have a wide variety of friends on both of the continents I have called home – allow me to thank a few. Lindsay – thank you for your friendship and your hospitality and for being genuinely kind. Brandon and Kristin – thank you for the time in Vancouver, and for the insightful, rapid conversation. Cassandra – thank you for being thoughtful. Francesca – thank you for the time-shifted conversations and the kitties (of course). Megan and Miryam – thank you for being friends in a strange land. Ask – thank you for the games and the insights into the finer points of Danish grammar.

Finally, there are two friends that deserve something special, sweet, elaborate, and brilliant in every respect. Let the record show that, without them, my life would be far poorer, and long may their friendship endure. They have both, in their separate ways, shared in the joys and the sorrows of my life, and they deserve fantastic rewards. What I have to give, though, are paragraphs, so paragraphs they will get.

To Annelies, my close and profound friend…you are the pinnacle of understanding and playful wisdom. You have been a wellspring of strength and joy ever since the day we met and started talking – long may it continue. Thank you for an uncountable number of supportive deeds, big and small, born from awesome caring and sweet friendship – you always seem to know the right thing to say, or to put into my inbox – and thank you for letting me try to return the favor. Thank you for pointing out which parts of this dissertation were incomprehensible – hopefully, I've taken care of the "this" problem. One day, I will edit yours, if you'd like me to see it. Thank you (and Marc) for hosting me when I just needed to get away from it all for a while and regain some sanity – it's an oasis of sanity (and thunder) around there! The colors might flow downward, but they do not fade, even in the snowy snow.

To Emily, my close and enduring friend…you are the pinnacle of kindness and grace. Your soul-deep generosity and thoughtfulness is both inspirational and humbling – you have made, and you continue to make, me a better person, I think. Thank you (and Elliott, and the little one) for hosting me - letting me write most of Chapter 3 sitting in your living room, and letting me into your lives. The life you have created is truly beautiful and inspirational. Thank you for the thoughtful conversations, and I most truly hope we keep speaking for a long time. Thank you for the radiant picnic under that blue smiling sky. To the littlest host, I kindly advise you to take care, smile, and listen thoughtfully to your wise parents. Finally, Emily, wishes are powerful – a honeycomb has six sides (see Figure 3.9, and hues in Paper I).

Now, to the science.

English Abstract

In this dissertation, factors affecting electron holographic measurements of the mean inner potential are explored. Electron holography in the transmission electron microscope (TEM) allows for quantitative retrieval of the amplitude and phase of the electron beam. Both the amplitude and the phase reflect properties of the specimen. The phase can yield quantitative measurements of nanoscale electric and magnetic potentials. One such electrostatic potential is called the mean inner potential. The mean inner potential is the average electrostatic potential measured between the bulk of a material and vacuum far from the specimen, and is non-zero for all materials. However, previous mean inner potential measurements have disagreed for the same material measured by different groups. Additionally, experiment and image simulation are known to differ for high-resolution TEM, but not for electron holography, and this difference is known as the Stobbs factor - therefore, a dataset that allows for good comparison between image simulation and experiment might highlight possible improvements in the simulation software.

The factors that affect the mean inner potential are explored through both experiment and simulation. Thickness measurements using different tomographic algorithms (algebraic, geometric, and discrete tomography) and non-tomographic methods are compared on an InAs nanowire. A self-calibrating tilt-series of holograms on the same InAs nanowire is acquired and compared with image simulations to analyze diffraction effects on the amplitude and the phase. There is relatively good comparison between image simulation and experimental data, but the experimental absorption parameter is found to differ between strongly and weakly diffracting conditions. Density functional theory simulations of the mean inner potential are carried out using the GPAW program, allowing for exploration of the surface dependence of the mean inner potential. Factors including surface facet, structure optimization (atomic position relaxation), adsorbates, and fringing fields at corners are all examined. Finally, surface modification is attempted on an InAs/InP nanowire, but leads to nanowire dissolution instead - this dissolution is briefly characterized for GaAs nanowires.

Dansk Resumé

I denne afhandling undersøges de faktorer, der påvirker elektronholografimålinger af det ”indre middelpotential” i materialer. I transmissionelektronmikroskopien (TEM) giver elektronholografi mulighed for kvantitativ udmåling af elektronernes amplitude og fase. Både amplituden og fasen afspejler prøveemnets egenskaber. Fasen giver en kvantitativ bestemmelse af de elektriske og magnetiske potentialer på nanoskala. Et sådant elektrostatisk potential kaldes det ”indre middelpotential”. Det indre middelpotential er det gennemsnitlige elektrostatiske potential målt mellem det indre af materialet og vakuum langt fra prøven; det er forskelligt fra nul for alle materialer. Imidlertid har tidligere målinger af det indre middelpotential udvist uoverensstemmelser for et givet materiale udmålt af forskellige grupper. Derudover er det velkendt at for højopløsning TEM (HRTEM), men ikke for elektronholografi, afviger de eksperimentelle resultater og billedsimuleringsresultaterne fra hinanden med en størrelse, der kaldes Stobbs faktor - derfor ville elektronholografiske datasæt der giver gode muligheder for sammenligninger mellem billedesimulation og eksperiment kunne lede frem til forbedringer i simuleringsalgoritmerne (softwaren).

De faktorer, der påvirker det indre middelpotential undersøges gennem både eksperimenter og simuleringer. Tykkelsesmålinger ved hjælp af forskellige tomografialgoritmer (algebraiske, geometriske, og diskret tomografi) og ikke-tomografiske metoder sammenlignes på en InAs-nanotråd. En selvkalibrerende ”tilt”-serie af hologrammer er optaget på den samme InAs-nanotråd og disse sammenlignet med billedesimuleringer for at analysere diffraktionseffekter på amplituden og fasen. Der er relativt god overensstemmelse mellem resultaterne af billedesimulering og de eksperimentelle data, men den eksperimentelle absorptionsparameter ses at variere mellem tilfælde med stærk og svage diffraktion. Tæthedsfunktionalteori simuleringer af det indre middelpotential er udført ved hjælp af et ”GPAW” program, der giver mulighed for beregninger af det indre middelpotentials overfladeafhængighed. Faktorer, herunder overfladefacet, strukturoptimering (graden af relaksation af de atomare positioner), adsorbater, og feltforvrængning omkring hjørner er alle undersøgt. Endelig er det forsøgt at foretage overflademodifikation af en InAs/InP-nanotråd, men forsøgene fører i stedet til destruktion (opløsning) af nanotråden - der gives en kortfattet beskrivelse af denne destruktion af GaAs-nanotråde.

Table of Contents

Chapter 1: Introduction

Motivation

The transmission electron microscope (TEM) is a tool for measurement and characterization at the microscale, nanoscale, and atomic scale using many different techniques on many different specimens [1]. One such technique is electron holography [2]. However, to properly apply a TEM technique such as electron holography to specimen characterization, the experimental and theoretical factors affecting the technique must be known. The objective of this dissertation is to understand the factors that can influence off-axis electron holography [3] of electric potentials in the TEM, through undertaking a combination of experiments using electron holography combined with other techniques, and also through examining simulations of both TEM images and electronic structure to determine the effect of different parameters, both controllable and intrinsic.

Off-axis electron holography in the TEM records the phase and amplitude of the electron beam. This allows for the quantitative and direct characterization of electric and magnetic potentials at the nanoscale. Quantitative electric and magnetic potential information about specimens at this scale is useful for a variety of applications, such as those seen in [4], [5], and [6].

However, if the phase and amplitude of a specimen are recorded using electron holography, then the recorded phase will also include a phase shift due to the intrinsic average electric potential in the specimen relative to vacuum, even for uncharged specimens. This electric potential is known as the mean inner potential, referred to in the literature as V_0 [7].

The mean inner potential is always present when phase shifts in a specimen relative to vacuum are recorded using electron holography. This would be the case for *e.g.* electron holography characterization of nanostructured *p-n* junctions or of nanowire materials, where the phase shift relative to vacuum could be used. Therefore, both theoretical and experimental influences on the mean inner potential should be considered. These measurements are easily quantifiable, because, if there is no other electric charge or magnetic contribution to the phase ϕ, then the phase is given by the equation $\phi = C_E V_0 t$, where C_E is a calculated factor only dependent on the accelerating voltage of the electrons, but both V_0 (the specimen mean inner potential), and t (the specimen thickness), are parameters that depend on the specimen. Thus, not only should the factors affecting the measurement of the phase be considered, but also the factors affecting the measurement of the specimen thickness, and the dependencies of the mean inner potential itself.

One reason why these factors should be examined is because mean inner potential measurements previously reported in the literature have not always been consistent. In principle, measuring the mean inner potential should be straightforward, but experimental measurements reported in the literature show a wide variation for the same material system, outside of their mutual margins of error – [8] provides a review of several different measurements of *e.g.* Si which do not agree. This disagreement raises multiple questions. What factors influence the precision and accuracy of an

experimental mean inner potential measurement, including the thickness? Given these factors, what is the most precise and most accurate way to measure the mean inner potential? More speculatively, is there possibly an additional factor or factors not considered at length in the literature that might be different between these experiments that makes the measurements not comparable between experiments? If simulations are used, either of the specimen or of the imaging process, how precise and accurate are their results? Is there a way to use phase and amplitude information from electron holography to evaluate the quality of current image simulation results, keeping in mind that sometimes these do not coincide for other imaging techniques (a mismatch called the Stobbs factor, for which an overview is provided in [9])?

Objectives of This Dissertation

Keeping the above questions in mind, the primary objective of this dissertation is the exploration of accuracy and precision in the measurement of the mean inner potential through multiple methods, including a consideration of specimen surface dependence. These methods are elaborated on in more detail in Chapter 2, but briefly outlined in this paragraph. Choosing a specimen that requires minimal preparation can minimize specimen preparation artifacts. The precision and accuracy of the specimen thickness determination, determined using a different TEM technique, can be both optimized and quantified, and different thickness determination techniques compared. Simulations are also possible – both of the imaging process and of the specimen. For the imaging process, there are dynamical diffraction effects on the recorded phase and amplitude that are due to the specimen orientation relative to the electron beam, and these effects can be measured by comparison with simulations, even for diffraction conditions that are nominally weakly diffracting. For the specimen, the effect of realistic electronic structure can be examined quantitatively using density functional theory, including the effect on the mean inner potential of surface reconstruction and modification. Finally, attempts at surface control can be made to try to realize some of these surface modification effects experimentally. These individual considerations comprise the set of experimental, analytical, and theoretical factors considered in this thesis. The technical background for this analysis, as well as a summary of past work in the field, is in Chapter 2.

The quantitative thickness determination of the TEM specimen is the focus of Chapter 3. Semiconductor nanowires are used to reduce specimen preparation artifacts – a single InAs nanowire is used in Chapter 4 as a specimen for electron holography. However, before the electron holography measurement can be analyzed, the cross-sectional thickness profile of this nanowire must be known. Therefore, in Chapter 3, the nanowire cross-section is characterized quantitatively, by applying both tomographic reconstruction algorithms and non-tomographic calculations (the latter using a technique introduced here as the renormalized orthogonal-images routine, or RENOIR) to images taken using high-angle annular dark-field scanning transmission electron microscopy (HAADF STEM). The factors affecting accuracy and precision of these thickness determination methods are discussed, and cross-sectional thickness profiles of the wire are determined.

In Chapter 4, electron holography is performed on the nanowire, and both the phase and amplitude are analyzed using the thickness information from Chapter 3. Electron holograms from different, regularly spaced specimen orientations are acquired from the InAs nanowire specimen. This systematic tilt series of holograms is then analyzed quantitatively for effects on the phase-derived mean inner potential and amplitude-

derived mean free path. There is found to be a systematic effect from specimen orientation. Diffraction effects on the holography measurements of phase and amplitude are studied both in strongly and weakly diffracting specimen orientations. The effect of diffraction contrast is studied quantitatively with reference to simulations to explain results from both the strongly and weakly diffracting conditions.

Then, in Chapter 5, the effects of the surface on the mean inner potential are simulated using density functional theory (DFT). This includes simulations under different surface conditions, for both one-surface-type simulations ("thin-film") and two-surface-type simulations ("nanowires"), for different surface configurations, including bulk-like unmodified surfaces, surface configurations reported in the scientific literature, and minimum-energy surface configurations determined using DFT structure optimization. The amount of mean inner potential shift due to surface modification is compared to the precision of the mean inner potential measurements.

Then, given the surface dependence described in Chapter 5, Chapter 6 starts by focusing on attempting to modify the surface of a semiconductor nanowire in the TEM under observation; while this surface modification does not occur as expected, the nanowire dissolves, and the dissolution behavior is characterized, but only in a preliminary fashion.

Finally, Chapter 7 summarizes the conclusions of the results in chapters 3, 4, 5, and 6, and points towards future work that would further address both the research questions that inspired this dissertation, and the new research questions arising from this dissertation's results. Four appendices (A, B, C, and D) are included, to address peripheral questions arising from the work. Three conference papers (**Paper I, II, and III**) are also included, as their questions are directly addressed in the relevant chapters.

Chapter 2: Technical Background

Overview

In this chapter, the technical background and methods underpinning this dissertation are discussed, including a discussion of previous relevant work reported in the scientific literature. The transmission electron microscope (TEM) is discussed generally, then the specific techniques used in this dissertation – electron holography, high-angle annular dark-field scanning transmission electron microscopy (HAADF STEM), and electron tomography – are described. Then, density functional theory (DFT) is briefly detailed in the context of mean inner potential simulations. Finally, previous experimental and theoretical work on the mean inner potential is discussed, forming a framework for experiment, analysis, and interpretation.

The Transmission Electron Microscope

The TEM, as a commercially available instrument for materials characterization, has been under continual improvement since its earliest days. Since the design and construction of the first TEM by Ernst Ruska in the 1930s, the practical and theoretical development of the instrument has proceeded on multiple fronts, yielding several Nobel prizes: Ernst Ruska in 1986 for TEM development [10], and Denis Gabor in 1971 for holography [11]. Recent developments, including spherical aberration correction (one of the first examples of spherical aberration correction is reported in [12]), extend the spatial resolution to the sub-Ångström length scale (see [13] for more details), with the reference [14] summarizing the past and possible future, and the reference [15], which discusses instrument design. While spatial resolution is only one aspect of the instrument, the continuing efforts to improve the TEM show the drive to improve the practical and theoretical understanding of the instrumentation and its application to answer novel material science questions (*cf.* [16] for a summary up to 1986, and [14] for a discussion up to 2008). At present, there are several competing commercial manufacturers for the TEM system (*e.g.* FEI, JEOL, Zeiss, NION), and additional components, including aberration correctors, detectors, specimen holders, and software, are also commercially produced. As the instrument design itself is outside the scope of this dissertation, the methods of TEM characterization and the results yielded by it are explained in this work from the perspective of the user.

The following section starts with a description of some of the general principles underlying the instrumentation, then narrows to focus on the specific techniques and instrumentation used in this dissertation. The scientific context of these techniques, including applications and ongoing challenges, is also discussed. Several books are recommended for further elaboration outside the scope of this work.

Principles and Overview of the Transmission Electron Microscope

The transmission electron microscope generates and guides a beam of electrons moving at relativistic velocities through a thin specimen, and the signals generated by this interaction are detected. This section discusses the general principles of TEM operation, including several points of practical TEM operation for the FEI Titan 80-300 microscopes used in this dissertation. These principles and practice form the foundation for the experimental results in chapters 3, 4, and 6, and, therefore, must be

explained and discussed. Reference [1] provides the general reference material for this section.[1]

Electrons in the TEM start their paths through the instrument by being emitted from the electron "gun", which is also called a tip. Several different tip designs are possible; the work in this dissertation requires and uses a field-emission gun (FEG) due to the increased electron coherence and source brightness. In a FEG TEM electron source, a strong electric field is applied to the tip of a sharply pointed needle. At the very tip, this electric field lowers the barrier for electrons to tunnel out, and the electrons that tunnel out form the electron beam in the microscope. There are two types of FEG TEM: the Schottky FEG TEM and the cold FEG TEM; this work uses the former.[2]

After emission from the FEG, the electrons are then accelerated to their primary energy (in this work, 120 keV or 300 keV) by a linear acceleration stage, and they are focused using electromagnetic lenses in fixed positions. Many lenses, taken together, comprise the electron-optical system of the TEM. These lenses are coils of wire that generate circularly symmetric ("round lenses") or non-circularly symmetric (multipole) magnetic fields. The round lenses are used for *e.g.* focusing, changing magnification, or selecting either a diffraction pattern or an image to appear at the observation plane. The multipole magnetic fields generated by the non-circularly symmetric lenses are used for aligning the beam and minimizing aberrations such as astigmatism (2-fold or 3-fold) or beam tilt. Different microscopes have different arrangements of lenses.

The differences between a typical TEM and an FEI Titan used in this work are additions to the latter, but both operate on the same general principles. A typical TEM such as an FEI Tecnai, on the left in Figure 2.1, has two condenser lenses and an objective minicondenser lens to shape the beam before the beam reaches the specimen, an objective lens with top-and-bottom polepieces to focus the beam onto the specimen sitting in the gap between the polepieces, and several post-specimen lenses (a diffraction lens, an intermediate lens, and two projector lenses) to allow for a choice of magnification, and a choice between a diffraction pattern or an image, with the result recorded on a charge-coupled device (CCD) camera. On the right, in Figure 2.1, is the roughly equivalent diagram for the FEI Titan 80-300 installed at DTU Cen referred to as the "ATEM"[3]. In addition to all the features seen in the "standard" microscope in the left diagram (including a CCD, omitted from the diagram), the ATEM on the right includes a Lorentz objective lens, a probe spherical-aberration corrector made of multiple multipole lenses, and a monochromator. The Lorentz lens

[1] There are two nomenclatures in the literature for "condenser" and "objective"; one used generally, but especially for TEM work, and one used for STEM work only. This work consistently follows the TEM convention.

[2] The Schottky FEG is used on FEI instruments. It is more stable operationally, but has a lower theoretical brightness, as compared to the cold FEG – see [1] for details.

[3] The differences in capability between the two similar FEI Titan 80-300 instruments installed at DTU Cen – the "ATEM" and the "ETEM" – are worth noting. In total: the ATEM has a spherical aberration corrector before the objective lens to minimize STEM probe spherical aberration, and the ETEM has a spherical aberration corrector after the objective lens to minimize TEM image spherical aberration (both correctors are the same optical elements). The ETEM has a gas-handling system in the objective lens polepiece gap and differential pumping apertures to introduce gases into the specimen region; the ATEM does not. The ATEM has an electrostatic biprism, for electron holography, at the aperture near the image plane with the selected-area aperture.

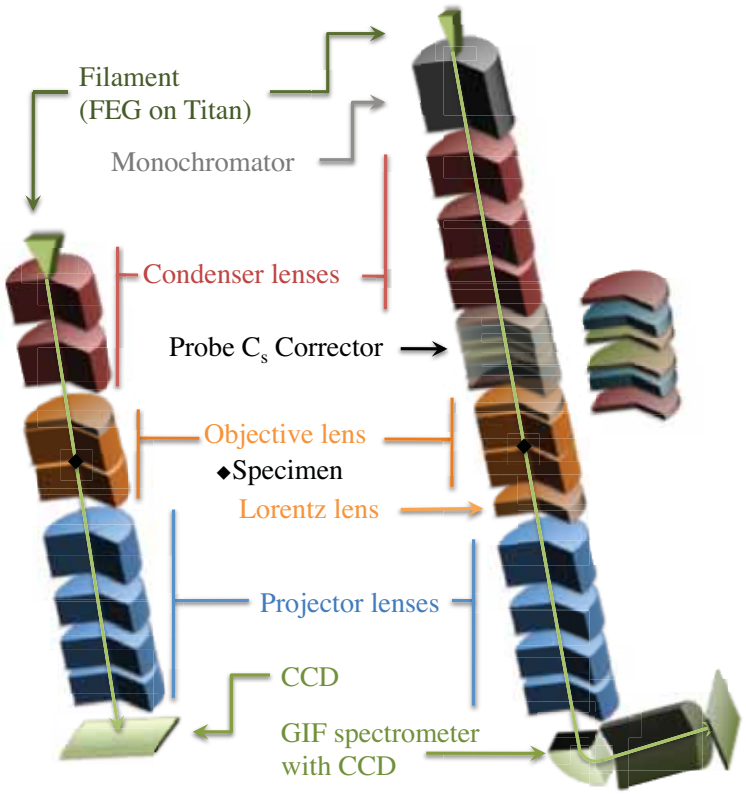


Figure 2.1: Microscope cutaway schematic diagram, showing primary round lenses and special multipole lenses. Left is a diagram of a generally available TEM, such as a FEI Tecnai T20; right is the DTU Cen ATEM - a FEI Titan 80-300 with monochromator, probe spherical-aberration corrector, and Gatan Imaging Filter (GIF). The green line shows the electron path. Note the third condenser lens in the Titan.

is designed primarily for imaging magnetic specimens without influencing their magnetic field; however, in Chapter 4, the Lorentz lens is used for medium-resolution electron holography. The probe spherical aberration corrector is designed for correcting spherical aberration of the objective lens in STEM mode. While the probe corrector is aligned for the STEM tomography carried out in Chapter 3, it is not used to improve the spatial resolution to sub-Ångström levels; it is also manually tweaked slightly for electron holography to remove beam distortions like the ones seen in [7]. The installed Wien-filter monochromator is not used to monochromate electrons in this work, but is a part of the electron-optical path through the TEM. Therefore, the monochromator is only aligned to allow the electron beam to pass through, and used as the gun lens would be used on a non-monochromated TEM[4].

The apertures in the TEM, while not shown in the schematic in Figure 2.1, are important for the function of the instrument. There are a handful of moveable apertures; on the ATEM, this includes three condenser apertures, one objective aperture, and one selected-area aperture. The condenser apertures control the total electron dose on the specimen, the illuminated area, and the probe size in STEM mode. The objective aperture is set in the back focal plane of the objective lens, allowing for selection of a region of reciprocal space (also called diffraction-space). The objective aperture is used during the electron holography experiments to ensure only the [000] beam is recorded; however, the edges of the aperture are blurred instead of sharp in Lorentz mode. The last moveable aperture is the selected-area aperture, at the image-conjugate plane to the specimen during normal ("SA" mode) operation. The use of this aperture allows for a diffraction pattern to be taken from a selected area. However, this aperture is important in this work for a different reason: on the ATEM, the electrostatic biprism used for holography is installed here [17]. By operating the microscope in "Mh" mode instead of the normal "SA" mode, the SA aperture plane is optically displaced from the image plane, allowing the biprism wire in the SA aperture plane to be charged up, making two different regions in the specimen plane overlap in the conjugate image plane. The consequence of the biprism location is that the biprism is placed before the projector lens array, leading to a fixed biprism size relative to the specimen for a given objective lens setting. Therefore, the biprism magnification relative to the specimen is dependent on the objective lens.

The objective lens itself requires explanation, as more than one physical lens can act as the objective lens in the TEM; in the microscope used in this work, two different physical lenses perform this operation in different imaging modes. The conventional objective lens is the one whose top and bottom polepieces encase the specimen. There is a small gap between them where the specimen is inserted – this distance influences the spherical aberration coëfficient C_s, and thus the resolution without an aberration-corrector. Decreasing the polepiece gap[5], before the advent of the spherical aberration-corrector, was the primary method of improving the resolution at a given accelerating voltage. However, increasing the distance between the specimen and the lens decreases the spatial resolution of the instrument – and the Lorentz lens is a prime example of this. As noted above, the Lorentz lens is designed to provide

[4] Because the monochromator is fixed to the end of the gun; therefore, gun alignments on a conventional microscope like gun tilt or gun lens settings are monochromator alignments. This does not impact this work scientifically, but is noted here for the practical reader unfamiliar with this specific instrument.

[5] With attendant decrease in available space in the gap – leading to more stringent tilt restrictions, etc.

magnetic-field-free imaging at the specimen plane, requiring the lens to be a greater distance away from the specimen than the conventional objective lens. To give an example of how this affects the spatial resolution of the TEM, at 300kV primary accelerating voltage (yielding a wavelength λ of 1.96pm), the spherical-aberration-limited point resolution *r*, calculated from the equation $r=0.64C_s^{1/4}\lambda^{3/4}$ [18] is *r*=0.20 nm for the conventional objective lens (C_s=1.2mm), and *r*=2.0 nm for the Lorentz lens (C_s=12m). It is illustrative to compare the latter to the *r*=1.6 nm limit cited by Gabor in his Nobel lecture as the practical TEM resolution limit in the late 1940s [19] to realize exactly how poor the Lorentz lens resolution is on a modern microscope. The interpretable spatial resolution in an image acquired using the Lorentz objective lens on an optimized microscope is considerably worse than an image acquired using the conventional objective lens on an optimized microscope. However, because the Lorentz lens provides a different objective lens magnification, it also provides a different magnification range for electron holography. Electron holography can be performed using either objective lens. Using the conventional objective lens yields "high-resolution" electron holography to characterize features on the Ångström length scale, with examples given in [20]; using the Lorentz lens provides "medium-resolution" holography on the nanometer scale, with examples given in [7]. The Lorentz lens is used for electron holography in this work in Chapter 4.

Two general operational modes of the transmission electron microscope are possible. One, conventionally referred to as TEM mode, uses a parallel beam, the other, conventionally referred to as scanning TEM or STEM mode, uses a convergent beam as a probe whose minimum-diameter "tip" is rastered across the sample. Both modes are used in this work, but in different ways. TEM mode is used for the holography images in Chapter 4 and the bright-field TEM images used for general characterization briefly in Chapter 3, and for nanowire dissolution in Chapter 6. STEM mode is used for the high-angle annular dark-field (HAADF) STEM images used for thickness determination in Chapter 3. TEM mode images in chapters 3, 4, and 6 are recorded using a standard CCD camera, in this case a Gatan Ultrascan 2048×2048 pixel array, and the entire image generated by the parallel beam is recorded simultaneously. In contrast, the HAADF STEM mode images used in Chapter 3 are recorded serially – with the probe scanning across the specimen plane, and the scattered electrons generated at each point recorded on a Fischione Instruments model 3000 detector, a ring-shaped scintillator-clad counting device of the same type characterized elsewhere [21] and used to detect electrons scattered to high angles (in this work, ~100 mrad), which is then reintegrated from the time each pixel was recorded, creating the HAADF STEM image. The TEM mode images used here are recorded using the entire CCD, which each pixel illuminated at the same time, so the time to acquire each TEM mode image is given by how long the detector shutter is open. STEM mode images are serial, so the time to acquire each STEM image is given by the probe dwell time at each point. The electron holograms used in Chapter 4, while recorded like conventional TEM images, are then reconstructed to yield the direct amplitude and phase of the electron beam. The reconstruction process is detailed in the next section.

The intensity recorded in STEM and TEM modes differ both in what is measured and how it is measured. STEM intensity is discussed in the HAADF STEM section later. TEM mode – including both holography and conventional imaging modes - measures electrons that are forward-scattered by the specimen; STEM mode measures electrons that have been deflected by some amount. Bright-field TEM mode measures these

forward-scattered electrons, even if they have lost energy when interacting with the specimen. For conventional bright-field electron holography, electrons that have undergone an inelastic (energy-dissipating) interaction event will not contribute to the holographic fringe contrast (and, thus, do not contribute to the reconstructed phase and amplitude) [22]. In the bright-field TEM image[6], the intensity is complicated to interpret, but, to a first approximation, the image intensity is a measurement of the number of electrons that have exited the specimen on the same trajectory as they entered it with. This includes *e.g.* double diffraction, referring to those that have been scattered twice inside the specimen – once away, and once back along the same vector, for no net change. Therefore, darker regions of the image generally indicate more strongly scattering material, but the image intensity undergoes orientation-dependent contrast reversals as thickness increases [1], so cannot be taken as directly interpretable. Intensity interpretation in HAADF STEM and holography will be detailed in the appropriate sections.

Electron Holography

Electron holography was first proposed by Denis Gabor in 1947 [2] as the first potential application of the principle of holography in an optical system – allowing for phase and amplitude retrieval from the original image using interference fringes – some of the consequences and historical concerns are described in Gabor's Nobel prize lecture [19]. However, it took some time for this to be used practically in the transmission electron microscope, because the TEM instrumentation required improvement. The approach taken in the original work was modified to perform off-axis electron holography, requiring the use of an electrostatic biprism invented by Möllenstedt and Düker [3] – although this does not preclude using Gabor's original idea; for a summary of multiple different approaches to electron holography, see [23]. There are other configurations than just the conventional bright-field approach used here for performing off-axis holography in the transmission electron microscope: the eponymous "Twenty forms of electron holography" [24], and some are also mentioned in [20]. It is also possible to perform electron holography in dark-field configurations, demonstrated in the 1980s as detailed in [25], but more recently the focus of practical and theoretical development starting in 2008 [5], with further development ongoing, as detailed in [26]. Bright-field medium-resolution off-axis electron holography is the method applied in this dissertation.[7] In this section, the technical requirements for bright-field medium-resolution off-axis electron holography are examined, then the role of phase and amplitude information is considered, and then diffraction contrast studies in the literature are explained. [7] is used throughout this section as a primary reference, and is recommended for further examination.

The conventional medium-resolution bright-field off-axis electron holography experiment requires two paths through the specimen region to intersect at one point in the detector plane. The hologram is formed with either one path passing through the specimen and one path through free space, creating an object hologram, or with both

[6] With the proper application of an objective aperture.

[7] There is a certain amount of ongoing discussion about off-axis electron holography using inelastically scattered electrons. In the conventional, bright-field holography case, where one path does not intersect the specimen, only electrons that have undergone energy losses on the order of $<10^{-15}$ eV should contribute, as noted in [22] and also mentioned in [27]. Holography using inelastically scattered electrons for specimen and reference paths has been previously demonstrated experimentally [28] and discussed in [29], but this topic is beyond the scope of this dissertation.

paths passing through only free space, creating a reference hologram. For each characterization image desired, an object hologram and a reference hologram are taken in quick succession. The reference hologram is taken immediately after the specimen hologram and without changing any lens parameters, meaning there will be the same distortions in the fringes of both holograms from the projector lens system. By comparing the phase and amplitude of these fringes in the object hologram against the reference hologram, these distortions can be removed, yielding the quantitative phase and amplitude of the electron beam as it passed through the specimen relative to vacuum.

For a successful off-axis electron holography experiment on a modern transmission electron microscope like the FEI Titan 80-300 used here, two features of the modern transmission electron microscope must be present. There must be an electrostatic biprism, and the electron source must be a FEG. The biprism allows for the path-overlap detailed in the previous paragraph to occur, and the FEG provides a coherent source to yield interference fringe contrast.

The electrostatic biprism is a critical device for off-axis electron holography [3]. As mentioned in the earlier paragraph on apertures, the biprism is a charged wire placed in the selected-area aperture plane. The voltage across the biprism is provided by an external power supply, which is under operator control. With the microscope image plane slightly offset relative to the aperture plane in "Mh" mode, the charged wire causes two parts of the beam with different trajectories through the specimen plane to overlap onto one another in the image plane. Rephrased, the charged biprism causes two paths through different parts of the specimen region to intersect at the same point in the image plane.

The field-emission gun (FEG) provides a benefit for electron holography from its "reasonably good stability, high brightness, and high coherence", to quote Missiroli *et al.* [23]. FEGs, because their electron emission is concentrated at the tip of a pointed needle, have a considerably higher brightness (defined as the current per unit solid angle [1]) than other electron sources, due to their small effective source size – tables of comparative order-of-magnitude values are available in several standard references (*e.g.* [18] or [1]). Coherence is directly proportional to brightness, as noted in [27]. The measured holographic fringe contrast is directly proportional to the coherence, also as noted in [27]. Therefore, the high brightness provided by the FEG allows for improved fringe contrast for electron holography. While a full explanation of coherence is outside the scope of this dissertation, other references such as [27] are recommended.

A brief practical note regarding the alignment of the TEM for electron holography might be of interest. The electron beam is carefully stigmated prior to an electron holography experiment, using the TEM condenser stigmators such that the incident beam is highly elliptical, with the major axis aligned perpendicular to the biprism wire, and the minor axis collinear with the biprism wire[8], which maximizes the spatial coherence in the major-axis direction [7]. This necessitated fine-tuning through empirical testing for practical operation on the DTU Cen ATEM. For best results, the third condenser lens was turned off, but the probe spherical aberration corrector

[8] In practical operation, this is routinely performed by maximizing one of the condenser stigmators to maximize the major axis of the elliptical beam and tuning the other to align the elliptical beam with the biprism.

created beam misalignments. To compensate for these beam misalignments, manual manipulation of the probe corrector lens excitations were performed to correct for coma, threefold astigmatism and beam-tilt in the pre-specimen condenser system – for an example of how this worked in practice, note the (much simpler) beam alignment before-and-after images in [7].

Quantitative Phase and Amplitude in Electron Holography

Electron holography allows for the quantitative preservation and retrieval of the phase and amplitude of the electron beam. The electron wavefunction ψ thus preserved is given by [31]

$$\psi(x,y) = A(x,y)e^{i\phi(x,y)}$$

where A is the amplitude and ϕ is the phase. This section details the influences on and utility of this phase and amplitude as measured with electron holography. These effects, neglecting diffraction contrast – which will be returned to later - are dependent on the specimen thickness. Neglecting diffraction contrast, the phase is linear with specimen thickness, and the amplitude exponentially decaying with specimen thickness. If diffraction contrast is not present, this should be orientation-dependent only in the sense that tilting the specimen will change the projected thickness of the specimen. Non-linear orientation-dependent diffraction effects in electron holograms are discussed in the following section.

The phase and amplitude are retrieved from the electron hologram using Fourier transforms and other software routines; this work takes the usual approach, detailed further in [7]. Holograms with interference fringes are loaded into the image-processing program (Gatan DigitalMicrograph), and a software plug-in designed for electron holography is applied (ASUHolography[9]), which works as follows. It applies a Fourier transform to the object and reference holograms, resulting in a centerband, generated by the non-interference-fringe sections of the image intensity, and two conjugate antisymmetric sidebands[10] generated by the interference fringes, with a streaking artifact radiating from the centerband towards the sidebands due to the Fresnel fringes resulting from the biprism not lying in a microscope image plane [30]. One sideband is masked manually in the Fourier transform of the object hologram; this mask is chosen to include the sideband and exclude the streaking. The selected sideband is centered and inverse-Fourier-transformed, yielding a complex image from which phase and amplitude can be extracted. Projector lens distortions are removed by taking the same process (using the same mask) on the reference hologram, and then renormalizing the reconstructed object hologram data values by the reconstructed reference hologram data such that the reference has a phase of 0 and amplitude of 1. If necessary, the phase can be unwrapped. Phase unwrapping might be necessary because the phase reconstructs to values on the semi-open interval $[0,2\pi)$, but the real maximum phase shift in the specimen may be $\geq 2\pi$. Several algorithms are included in the package to perform this, but they are not detailed here, because the phase images in Chapter 4 are unwrapped manually.

[9] My enduring thanks to Professor McCartney at Arizona State University for this.
[10] *i.e.* the only difference between the two sidebands is the sign of the phase – one is positive, the other is negative.

The experimentally measured phase shift depends on electric and magnetic potentials. Neglecting diffraction effects, which will be described in the following section, this dependence of the phase ϕ is given by [31] [32]:

$$\phi(x,y) = \frac{\pi}{\lambda E}\int_{-\infty}^{\infty} V(x,y,z)\,dz - \frac{2\pi e}{h}\int_{-\infty}^{\infty} A_z(x,y,z)\,dz$$

where λ is the relativistic electron wavelength, h is the Planck constant, e is the charge of the electron, E is the primary beam energy, the z-direction is collinear with the beam, A_z is the magnetic vector potential component in the beam direction, and V is the electric (scalar) potential. This dissertation does not concern itself with magnetic specimen properties, nor with electrical properties except for the mean inner potential (and incidental specimen charging), which itself will be treated near the end of this chapter. This phase shift is, therefore, quite characteristic of specimen electric and magnetic properties.

Because of this electric and magnetic potential dependence, the phase shift has many applications. It has been used for experimental exploration of the Aharonov-Bohm effect reported in [33] and additional work summarized in [34]. It has also been used for nanoscale magnetic and electric field mapping, with [4] providing a review of several examples: nanoscale magnetic elements' field profiles, transistor dopant profiles, *in situ* electrical biasing of *p-n* junctions, interface charge and polarization characterization in semiconductor heterostructures, and quantum dot charge accumulation. Another application is seen in [35], which characterizes dopants in focused-ion-beam-prepared GaAs *p-n* junctions. Characterization of *p-n* junctions is also reported in [36]. Long-range electric and magnetic fields in free space are characterized in several ways in [37], with a focus on semiconductor nanowires reported in [38]. A brief review of electron holography as applied to biological specimens is in [39]. Some other uses of phase images generated using electron holography can be found in the standard reference mentioned previously [7]. These are just a selection of the possible applications. For other further reading, two review articles covering the applications of analyzing phase information from electron holography to nanoscale characterization are [6] and [40].

The amplitude measured using electron holography also reflects material properties, but can be also analyzed as an amplitude-derived mean free path. Neglecting diffraction effects, the measured amplitude decreases exponentially with specimen thickness, and this rate of decrease yields a material-dependent absorption parameter, described in more detail in [7]. In addition, the measured amplitude can be used to calculate a parameter linearly dependent on thickness in the absence of diffraction contrast, t/λ_i, defined in [7] as:

$$t / \lambda_i = -2 \ln \frac{A_0}{A_r}$$

where A_0 is the amplitude in vacuum, A_r is the amplitude in material, λ_i is a mean free path, and t is the thickness of material (variable names from [7][11]). As described in [7], this mean free path is material-dependent, and is the mean free path between any interaction events that cause the electrons to no longer contribute to the recorded

[11] λ_i, the mean free path, is referred to elsewhere in this work as λ.

interference fringe contrast. As compared to the phase, the amplitude is generally used less frequently, but it has been considered. This consideration has been mostly but not entirely for thickness determination t through amplitude-derived t/λ_i measurements, like those shown in [41]. [42] applies this amplitude-derived thickness determination approach to remove thickness dependence in phase images, but [27] strongly contraindicates this approach, because the errors in the phase and amplitude are inversely proportional.

Because diffraction contrast is neglected, the amplitude and phase are considered above as both linear or predictable with specimen thickness, and not orientation-dependent. When diffraction contrast stemming from crystalline specimen orientation relative to the electron beam is considered, there is an additional effect that is nonlinear with thickness, explored and described in the next section.

Nonlinear Diffraction Contrast in Electron Holography

The previous section discussed the conditions where the phase shift and amplitude-derived t/λ are linearly dependent on thickness. In contrast, this section addresses an effect on the measured phase and amplitude-derived mean free path that is non-linearly dependent on the thickness: orientation-dependent diffraction contrast. This section will discuss first the source and importance of this nonlinear diffraction effect, and then the explorations of this effect reported in the literature will be summarized.

Diffraction contrast in electron holograms poses a specific challenge due to its inevitability. Diffraction contrast is mostly noted as an effect to avoid in [7]. However, if a crystalline specimen is used, then there will be additional amplitude and phase shift of the electron beam due to dynamical diffraction, and the magnitude of this shift can be sensitive to the specimen orientation, especially near a zone axis. This sensitivity is demonstrated and noted in [5] and [43] when crossing an extinction contour. Thus, it is necessary to also simulate these effects, with some discussion of simulations noted in [44], and simulations for corrections to the mean inner potential due to dynamical diffraction were performed for the mean inner potential measurements reported in [45].

In contrast to the effects in the previous section, there have been comparatively few reports exploring the role of diffraction contrast in electron holography. Early investigations focused on using bent specimens for measuring and determining the presence of diffraction contrast in unreconstructed holograms especially around extinction contours. The shift in hologram fringes near strongly diffracting conditions is investigated experimentally in the work of Hanszen *et al.* in [5], [46], and [43], while [47] presents experimental results and includes a theoretical discussion and modeling of strongly diffracting specimen orientation effects in electron holograms. These studies were limited by the lack of CCD cameras for reconstruction – this importance is discussed at length in [7]. More recent investigations have shown that diffraction contrast is, indeed, a potentially dramatic effect, but have only examined a few specimen orientations, and not usually compared the accuracy or precision of the simulations themselves to the measurements, meaning the simulations may be sensitive to different parameters[12]: [48] examines and models the phase shift at total of 8 specimen orientations, all strongly diffracting, while [45] simulates residual diffraction contrast at weakly diffracting conditions and also notes the effects of zone-

[12] Such as a different model for absorption parameter – see Chapter 4 for more details.

axis orientations; [49], [50], and [8] all use the output of simulations to compensate for residual diffraction contrast in phase images only at the measured orientation. [51] discusses dynamical diffraction effects on the mean inner potential in a theoretical context, and the measurements seen in [52] (included in this dissertation as **Paper II: Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials**) are also orientation-dependent, but diffraction contrast is not compensated for. However, the diffraction-contrast modeling work reported in the literature does not use a self-consistent dataset to perform systematic, large-scale investigations of a wide variety of strongly and weakly diffracting specimen orientations. Thus, the simulated amplitude and phase compensations for diffraction contrast reported in the literature may not be entirely accurate.

Simulations themselves are known to not be completely ideal. Notably, there is reported mismatch between high-resolution TEM image simulation and results, named the Stobbs factor after one of the co-authors on [53], and discussed more recently in [9]. [9] also points towards the use of electron holography for improving measurements in the search for the cause of the Stobbs factor, as high-resolution electron holography does not appear to show the same effect. For the use of simulations to compensate for diffraction contrast to be fully convincing, a wide range of specimen orientations should be systematically explored in a way that yields a self-consistent dataset, and the validity of simulations to model these effects can then be investigated. This is done in this work in Chapter 4.

The discussion moves now from electron holography to HAADF STEM, a different imaging technique. HAADF STEM can have intensity linearly dependent on specimen thickness, which makes it useful for thickness determination.

High-Angle Annular Dark Field STEM

STEM has multiple different possible imaging modes defined by their detector characteristics – three common ones are bright field, low-angle annular dark field (LAADF), and high-angle annular dark-field (HAADF). The objective in Chapter 3 is to obtain images where the intensity is monotonically dependent on specimen thickness, because then the image intensity can be easily related to the specimen thickness. Of these three, HAADF is the most useful technique for this dissertation for the following reasons.

Each of these three STEM imaging modes collects a set of electrons based on their angular deflection at the detector plane. The detector for bright-field imaging collects the electrons that are not deflected outside of the angle subtended by the detector – usually slightly larger than the original incident convergence angle. Both LAADF and HAADF collect electrons that have been deflected outside of the incident STEM probe convergence semi-angle, and the difference in collection techniques between LAADF and HAADF is the detector collection angle – HAADF usually indicates an inner detector angle[13] of at least 40-100 milliradians [54], and LAADF uses an inner detector angle less than that, but more than the convergence semi-angle. However, HAADF STEM is better for the following reason.

HAADF STEM image intensity has a signal dependent on the average atomic number Z in the sample as $Z^{1.7\text{-}2}$ - [54] has a brief review of the literature. While some debate continues about the exact atomic number dependence, the specimen in Chapter 3 has a

[13] The inner detector angle is the minimum deflected angle that contributes to the measured signal.

homogenous chemical composition and density in the region of interest. It is more important for this dissertation if the intensity in HAADF STEM imaging has a linear and monotonic response with thickness. In that case, the image intensity is then directly interpretable as material thickness. HAADF STEM image intensity is known to not be exactly linear with specimen thickness at high thicknesses, as noted for InP and InGaAs above approximately 150nm near a zone axis orientation [55]. It is also known to not be linear with thickness at some orientations - [56] considers HAADF and LAADF STEM, showing some of the complexities involved. However, the specimen used in chapters 3 and 4 does not appear to be thick enough for this to disrupt the monotonicity, and the angular dependence would necessitate only rescaling it for some specimen orientations to retain the same wire cross-sectional area[14]. The linearity of the HAADF STEM images themselves is considered in Chapter 4, through comparison between HAADF STEM images and holograms taken at weakly diffracting conditions. For further details, see chapters 3 and 4.

In Chapter 3, HAADF STEM data is taken and used as input to tomographic reconstruction algorithms – this is a standard approach, as pointed out in the review article [6]. Tomographic reconstruction is detailed briefly in the next section.

Electron Tomography

Electron tomography is an analytical technique that can reconstruct the specimen morphology from TEM data. This is done by solving an inverse problem: inverting the measurement process to determine the object that generated the measurements. One of multiple different algorithms can be chosen to perform this reconstruction. Further details are found in Chapter 3.

The purpose of electron tomography is to reconstruct the morphology of the original specimen from images acquired at multiple different specimen tilt angles, and this has been successfully applied to many different specimens – note the diversity listed in the review article [6]. However, the reconstructed specimen morphology is a product of both the experimental technique used for the original data acquisition, and the algorithm chosen to reconstruct the morphology. Thus, the assumptions that underlie these must be carefully considered before the product can be fully interpreted.

To perform an electron tomography reconstruction, an algorithm is used to process a stack of images acquired at different specimen tilt angles. The choice of algorithm must be compatible with the method that is used to acquire the images. As stated above, Chapter 3 uses a tilt series of HAADF STEM imaging data, and HAADF STEM imaging data is called “ideal” in [6]. There are different possible algorithms that may be used, but algebraic tomography (AT) [57] dominates the electron tomography literature (*e.g.* the review [6] primarily covers AT). While AT algorithms are used in this work, they are not exclusively used. Geometric tomography (GT) is applied, using the algorithm outlined in the paper [58], and this application can be seen in **Paper I: Reconstruction of an InAs nanowire using geometric and algebraic tomography**. Discrete tomography (DT) is also applied to the data in Chapter 3. This uses the DART algorithm first proposed in [59]. DART is applied to a variety of specimens in [60], and repeatedly mentioned in the review [6]. These are detailed and compared more thoroughly in Chapter 3.

[14] This is not considered in this work, but might be the subject of future work.

While the previous sections of this chapter have focused on TEM characterization techniques, the role and context of density functional theory for mean inner potential simulations in this dissertation is briefly explored in the next section.

Density Functional Theory (DFT)

Density functional theory (DFT) is a method for simulating the approximate electronic structure of materials, detailed briefly in this section. The background to DFT itself is not the primary focus of this research – this dissertation focuses on applying DFT to simulating mean inner potentials. While there are alternatives to DFT – such as quantum Monte Carlo [61] – DFT was chosen for several reasons: its availability (GPAW, the code used here, is developed in large part at DTU Fysik), because it can calculate mean inner potentials, and because previous mean inner potential simulations have used DFT codes and found mean inner potentials in line with experimental data [8], allowing for comparison between this work and reports in the literature. For further reading, and as a primary source for this section, see [61].

Very briefly, density functional theory determines the ground-state charge density of a system through an approximate theoretical framework. Solving for exchange and correlation interactions of a large number of electric charges is a non-local problem and computationally intensive to calculate exactly [61]. However, the Kohn-Sham *ansatz* [62] posits that this interacting system of electrons can be represented as a non-interacting system with a local interaction functional, making this problem computationally tractable [61]. The form of this functional is an approximation, calibrated against experimental results; several different functionals are usable, such as the local density approximation (LDA) [61] or one of several generalized gradient approximations (GGA), such as the PBE functional, named after the co-authors on [63], which proposed it. With this theoretical framework, the electronic structure of an arrangement of atoms can be modeled.

There are different implementations of DFT available for use. For an overview of many different approaches to DFT, see [61]. The only DFT implementation used here is the grid-based projector-augmented wavefunction method, implemented as the GPAW software package, further detailed in Chapter 5. Previous reports of mean inner potential simulations using DFT, and the mean inner potential itself, are covered in the next section.

The Mean Inner Potential V_0

This section explores one of the fundamental quantities for an electron holography measurement of material: the mean inner potential. In this section, the mean inner potential is defined. This definition is followed by discussions exploring the previous work on experimental measurement, theoretical frameworks, and values obtained using simulations.

There are other techniques that can measure the mean inner potential in the TEM, but these will not be examined in detail in this dissertation. One of these is Fresnel contrast analysis, which has been used for mean inner potential modeling at interfaces [64]. In the electron diffraction context, like that outlined in [49] and also in [65], the mean inner potential can be calculated from the electron scattering factor at zero scattering angle, as outlined in [66]. Tables of calculated isolated atom scattering factors (IASF) such as those in [66], or calculated more recently in [67], can be used for mean inner potential calculation. However, mean inner potentials calculated from

IASF have been shown to not be experimentally accurate: reference [8] compares experimental, DFT-simulated, and IASF-calculated mean inner potentials, finding the DFT-simulated results to be closer than the IASF.

The Mean Inner Potential From Electron Holography

The mean inner potential, as measured using off-axis electron holography, is the difference in electrostatic potential between material and vacuum, as noted in Chapter 1. In this section, the method of deriving the mean inner potential from the phase recorded using electron holography is defined. If there is no diffraction contrast, no electric fields (due to *e.g.* charging [38] or doping [35]), and no magnetic fields in the specimen, then the equation cited above for the phase becomes, for each point in the reconstructed phase image [7]:

$$\phi = C_E V_0 t$$

where $C_E = \pi/\lambda E$ is the accelerating-voltage-dependent prefactor, t is the thickness, and V_0 is the mean inner potential – which is zero for vacuum, and non-zero for material. An electron holography phase measurement measures the difference in phase between the vacuum path and specimen path. Thus, if there is a difference in electrostatic potential between material and vacuum due to the presence of electric charges (*e.g.* electrons and nuclear protons), then this difference in electrostatic potential will yield a phase shift when material is present.

Now that the method of measuring the mean inner potential using electron holography has been defined, the questions turn to what values have been measured for it before, and how the mean inner potential varies for different materials.

Previous Experimental Work

Previous experimental mean inner potential measurements using electron holography have evolved over the years. In the era before the CCD digital camera, optical benches were typically required for reconstruction, rendering quantitative mean inner potential measurements using electron holography very difficult to attempt: [68] details the first attempt to perform a mean inner potential measurement, which is an early effort on metals. Without good mean inner potential measurements, the experimental objective was usually to cancel out the effect of the mean inner potential on the experimental measurements. This led to the drive to combine and cancel mean inner potential and thickness effects in electron holograms when measuring electric potentials in semiconductor wedges [43] or small particles [31], or when examining diffraction effects in bent crystals [5]. However, in the same period, the mean inner potential was used to measure compositional variations [69]. The development of the CCD camera led to digital acquisition of electron micrographs. Soon after, quantitative measurements of the mean inner potential itself were proposed and taken at medium resolution on as-cleaved wedges with well-defined crystal facets, as reported in [70] and similar measurements were reported in [71], but that report focused on the camera characteristics. Follow-up work quickly branched out into mean inner potential measurements for different materials systems. The work in [70], a conference paper, was followed by a full article [45], which measured the mean inner potentials from cleaved wedges of GaAs, PbS, MgO, and Si, and also compensated for dynamical diffraction. Later, several different III-V semiconductor mean inner potentials (GaAs, InAs, GaP and InP) were measured from cleaved wedges, including dynamical diffraction effect compensation [50]. Devices and

device characteristics were also of new interest following this: the mean inner potential was measured for Si as part of a project to measure dopant profiles in transistors in [72], then a similar approach was also taken for Si doped structures in [73] to understand their basic properties. Changes in mean inner potential have also been interpreted as compositional variation and across interfaces. The work in [42] measured $CoSi_2$/Si interfaces, using amplitude images to remove thickness dependence. The mean inner potential as a marker of compositional change across a Mo/Si heterointerface and a Si_3N_4 grain-boundary were also studied in [74]. Grain boundaries in $SrTiO_3$ were studied in [75]. InGaN/GaN heterostructures were studied in [76], and AlGaAs/AlAs/GaAs heterostructures in [77]. However, reliability of thickness measurements is a priority for reliable quantification of mean inner potential measurements, leading to statistical considerations of regularity in thickness determination and amplitude-derived mean free paths, as reported in [78]. Most notably, several techniques were applied for the thickness determination used in the measurement of the mean inner potential of Ge reported in [79]. While the previous research listed above mostly looked at prepared wedge structures, some limited studies have looked at mean inner potential measurements from GaN nanowires [80], ZnO nanorods [81], or C nanotubes [82].

In addition, the surface of the specimen has attracted some experimental interest. The difference in mean inner potential between amorphous or crystalline Si nanosphere cores and their SiO_2 shells was measured in [83]. Anomalous mean inner potentials for Au nanoparticles were reported in [84] and followed up in [85]. Potential thickness-independent surface effects have been found in different studies: [86] reported a thickness-independent phase shift on a thin amorphous C film, and [87] examined thickness-independent phase shifts in C nanotubes.

In recent years, experimental studies have also been coupled with DFT simulations of mean inner potentials, with one example being [8]. This is in part because, as [8] outlines, when the same material was measured differently, the reported mean inner potentials were in disagreement. This is especially true for Si, which has been measured reasonably frequently, but which has a wide variation in mean inner potential values compared to the reported precision on each individual measurement. This inconsistency should not be present if the mean inner potential is material-dependent, and if the measurement is precise. Therefore, it makes sense to consider the previous reports on theoretical constructions of the mean inner potential and simulations, to see if there is a way to simulate mean inner potentials and evaluate the precision and accuracy of both simulations and measurements.

Previous Theoretical and Simulated Work

Theoretical frameworks and simulated values have long been part of the research interest surrounding the mean inner potential as measured using electron holography. Simulations of holograms themselves have been carried out previously to *e.g.* understand charging effects or small magnetic domains, both of which are considered in [88]. As noted above, early simulations of the mean inner potential relied on tabulated [67] isolated-atom scattering factors at zero scattering angle [66]; these are likely inaccurate due to the neglect of realistic electronic structure. More recently, density-functional theory calculations of the mean inner potential have yielded results similar to those measured experimentally, but there are still choices that need to be made when the simulations are carried out, and error analysis, when present, has not always been a priority.

The pioneering report of this, using DFT simulations of the mean inner potentials of Si, Ge, and MgO, is [89]. Zincblende III-V semiconductor mean inner potentials, Si, and Ge are treated in [8], with their wurtzite counterparts and Au in [90], using non-polar surfaces. Other mean inner potentials are calculated for II-VI semiconductors in [91], and amorphous C structures are treated in [92].

The theoretical framework for the mean inner potential as measured using electron holography is less extensively developed than the experimental measurements, but has not been completely unreported. The most important single paper in this respect, by Saldin and Spence [65], considers both low and high accelerating voltages, and relates the mean inner potential to bulk material properties such as the work function and Fermi energy; this paper is referred to extensively in Chapter 5. Some discussion of the mean inner potential and related quantities is found in [93] and [94], which, when considered together with [65], provide an overview of how the mean inner potential relates to specimen properties. Surface dependence of mean inner potential simulations was noted in passing in [89], which runs DFT simulations on several different bulk-terminated Si surfaces, but [89] does not fully describe the surface-dependent ramifications.

The discussion turns now to nanowires, which are the class of specimens used in this work. This discussion includes both the properties of the specimen, and also why they are useful for this work.

Nanowires

Nanowires are an important class of materials for a variety of applications. [95] provides a good review of the type of semiconductor nanowire used in this dissertation (and the metal-oxide vapor-phase epitaxy growth process used to generate the wires). Nanowires are whiskers of material, nanometers in diameter, but often microns in length, and can be grown with crystalline epitaxy to their substrate, all properties that are noted in [95]. However, in this work, their practical applications are less important than their morphology, and this section will give only a brief resume of the relevant characteristics of the types of nanowires used here. This work uses nanowires grown from III-V semiconductors, which can often have hexagonally-shaped cross-sections, as can be seen in similar wires grown in [96]. The nanowires are grown from heated Au catalyst particles, a routine method used since the days of micro-scale whiskers, where [97] was the pioneering work for this growth model, with the particle absorbing the nanowire source material (gas-phase in the case of this work), forming a eutectic between Au and the semiconductor material (noted in [95]), before the semiconductor material precipitates out and forms the wire. Specific properties of wurtzite InAs and InAs/InP wires can be found in [98], which covers similar wires to those considered here, and [99] provides an overview of the growth conditions used in general. InAs, GaAs, and InAs/InP nanowires are used at different points in this dissertation, and wurtzite crystal structures are allowed in these nanowires: [98] details and explores these wurtzite nanowires.

As nanowires have nanometer-scale diameters, they do not necessarily need extensive specimen preparation before TEM characterization; in this dissertation, the TEM specimen grid is simply scraped over the substrate with the wires. This preparation method allows for specimen preparation artifacts to be almost completely avoided.

The nanowires in this work are used as material references, because of this lack of specimen preparation artifacts. If they were grown with some variety of dopant included, then electron holography might show the difference between doped and undoped wires. This effect could, in principle, be measured, similar to work shown in [38].

Objective of the Dissertation

In this dissertation, the technical background outlined above led to specific experimental, analytical, and theoretical choices. The phase shift from the mean inner potential measured using electron holography in a TEM specimen is dependent on the thickness, requiring accurate and precise thickness measurements, which are outlined in Chapter 3. To obtain the mean inner potential using off-axis electron holography, phase changes must be measured accurately, and the non-linear effects on the phase and amplitude arising due to orientation-dependent diffraction contrast can be quantified, as shown in Chapter 4. Additionally, these experimental holography results are compared with the results of DFT electronic structure simulations of the mean inner potential in Chapter 5.

The quantitative measurement of the mean inner potential from experimental data requires a reference sample of known composition and structure, with minimal specimen preparation artifacts, and with a clear thickness profile. While prepared wedges have been used for much prior work (*e.g.* [78], [45], and [79]), semiconductor nanowires fulfill these requirements well [81]; unlike wedges, nanowires can be examined in the transmission electron microscope with minimal preparation, thereby avoiding specimen preparation artifacts as much as possible. Nanowires also allow for direct thickness determination using other techniques like electron tomography that are not feasible on crystal wedges due to wedges self-shadowing in the beam when tilted.

In Chapter 3, to quantitatively determine the thickness profile of a single InAs nanowire, a tilt series (140° span, sampled every 1°) of HAADF STEM images was used. To analyze these HAADF STEM images and determine a thickness profile, several different thickness determination methods were used, and two methods were shown to be effective. The first method is to apply the tomographic reconstruction methods detailed above to solve the inverse problem of specimen morphology by using this tilt series of images. Once the specimen morphology was determined, it was forward-projected to determine the thickness profile at the requisite tilt angles. Discrete tomography was found to be the most effective tomographic algorithm, but discrete tomography is not without its challenges, as discussed elsewhere [100]. The second method was to calibrate the intensity level in the HAADF images as a function of the width in the images taken at orthogonal angles, allowing for the intensity levels in the image to be directly translated into a thickness measurement. Both of these methods are employed in Chapter 3, and the advantages and drawbacks with each of them are explored quantitatively.

Once the nanowire thickness profile is determined, two different methods of determining the mean inner potential from phase data recorded using electron holography are considered in Chapter 4. The first is to average over the constant maximum thickness region of the nanowire, providing statistics, but including the phase contribution from the surface layer. The second is to use the gradient along the sides of the nanowire, determining the change in mean inner potential with the change

in thickness. This approach allows for the detection of the surface layer, which is then compensated for, and the mean inner potential of the InAs nanowire core is measured.

A tilt series of electron holograms (11°, sampled every 0.1°) is used to account for nonlinear diffraction effects in this nanowire specimen [48]. By systematically tilting a single nanowire through both strongly and weakly diffracting conditions, the effect of diffraction contrast on the phase and amplitude images is measured and compared with quantitative simulations. Weakly diffracting conditions where the phase is slowly varying with tilt angle are averaged over, allowing for an improved signal-to-noise ratio. Strongly diffracting conditions are also examined; by combining analysis of both strongly and weakly diffracting conditions, the accuracy and precision of simulations of dynamical diffraction effects on the amplitude and the phase are explored, with particular note taken of the absorption parameter.

These effects all influence the measurement of mean inner potentials; what influences the mean inner potential itself? To explore these dependencies, DFT is used to simulate the electronic structure of materials as a function of geometry, chemical composition, and surface state. Mean inner potentials are derived from these calculations.

The first task is to determine the precision of mean inner potentials calculated using DFT. This allows for the creation of acceptance tests for determining whether a specific parameter modification actually changes the mean inner potential, outside of the margin of error. These are compared to the margins of error for the experimental work performed in the previous sections.

Once acceptance thresholds are determined, specific investigations can be carried out *in silico*. The effects of surface modification are primarily examined in this work. While the mean inner potential is usually assumed to depend only on chemical composition and lattice parameter; it is shown that, in fact, there is an effect from the surface in some configurations, through the choice of surface facet, adsorbate layer, or surface reconstruction. With DFT results as a guide, experimentally testable situations for this surface dependence are identified.

With this potential surface dependence in mind, an attempt at nanowire surface modification is then described in Chapter 6. In this attempt, there was no sign of surface compositional change; however, the nanowires themselves underwent a dissolution process, decomposing along the nanowire/catalyst-particle growth interface. This dissolution process is then quantified, to a limited extent, and its temperature dependence and the role of beam effects are both examined.

After setting out this framework and technical background in this chapter, the following four chapters are experimental results and analysis, as mentioned above. The first of these chapters, and the next part of this dissertation, is the thickness determination chapter, Chapter 3.

Chapter 3: Thickness Determination

Outline

In this chapter, the thickness profile of an InAs nanowire cross-section used for electron holography in Chapter 4 is determined. This is done using both tomographic and non-tomographic techniques, and the efficacy of these techniques at thickness determination is evaluated, including both random and systematic errors.

This chapter starts with the data input and analysis used for each of the subsequent steps, and then briefly discusses the full relevance of calibration in this work. After that cross-section reconstruction results from the three algebraic tomography (AT) techniques are briefly discussed (with their full results in Appendix B). Then, geometric tomography (GT), and discrete tomography (DT) are applied to determine the nanowire cross-section. The DT results are then compared to the results from the renormalized orthogonal-images routine (RENOIR), which is a different, non-tomographic method for thickness determination applied to the same data. RENOIR and DT are shown to yield the best cross-sectional information, and are used in Chapter 4.

The factors affecting accuracy and precision of the DT and RENOIR data are then examined, and discrepancies between the two datasets are investigated.

It should and shall be noted that the work of other people contributed materially to the results and data in this chapter, and are thanked for their contribution. L. Houben provided the nanowire image alignment used for several of the reconstructions (details in the section on data processing), A. Alpers performed the geometric tomography reconstruction, and K. J. Batenburg performed the discrete tomography reconstruction. In 2006, L. Fröberg grew the InAs wire specimen used here.

Experiment, Data Input and Treatment

Experimental Conditions

After growth using metal-organic vapor-phase epitaxy [95], InAs nanowires were mechanically harvested from the growth substrate onto a holey carbon film, mounted in a Fischione Instruments model 2020 single-axis tomography holder with a possible axial tilt range of ±80° from horizontal, and examined in a FEI Titan 80-300 transmission electron microscope (TEM) installed at the Center for Electron Nanoscopy at the Technical University of Denmark. This instrument is detailed in Chapter 2 as the ATEM. It was operated at 120kV for bright-field TEM (BFTEM) imaging of a single nanowire and, during a different session, the same instrument was operated at 300kV for high-angle annular dark-field scanning TEM (HAADF STEM) for thickness determination of the same wire.[15] 2048×2048 pixel images were acquired for both techniques.

HAADF STEM imaging was carried out with a probe semi-convergence angle of 17 mrad, and an inner detector semi-angle of 99 mrad. The probe size was not measured. The inner detector semi-angle makes it clearly in the high-angle regime of ADF

[15] The BFTEM images were recorded at 120kV because they were acquired along with the holography data detailed in Chapter 4.

STEM imaging – as noted in Chapter 2, [54] states that HAADF imaging uses an inner detector semi-angle of at least 40-100 mrad.

The nanowire for study was selected from the wires on the specimen grid based on its orientation and support. The long nanowire primary axis was approximately collinear with the alpha-tilt axis of the holder, making the cross-section of the wire easier to determine. The extensive contact with the carbon film kept the wire from moving during data acquisition.

The acquired HAADF image series spanned the tilt range from -65° to +74°, measured relative to the horizontal on the stage goniometer, with 1° increments between images, but some of the most extreme angles showed signs of shadowing in the image, limiting the tilt range that could be used. This shadowing is from the copper grid holding the carbon film, not the holder, but was inevitable in this experiment. It would be possible, in principle, to improve this through mounting the nanowire on the tip of a needle-like specimen holder, but this was not done in this dissertation due to the specimen preparation required.

A representative HAADF STEM image from the tilt series is seen in Figure 3.1, including the as-acquired spatial calibration. The image in Figure 3.1 also highlights the three cross-sections (in blue, green, and red) that are analyzed using electron holography in Chapter 4. Figure 3.1 also includes intensity line profiles from the HAADF image, with and without spatial averaging, to demonstrate the noise levels. Note the original spatial calibration on the microscope, discussed more in Figure 3.2.

Calibrations were carried out under the same microscope alignment settings as the experimental data, including objective lens defocus. These alignments were restored between sessions from the saved, maintained microscope alignment files, with no software upgrades between sessions, and the same alignment procedures were carried out.

Data Processing

This section explains how the HAADF STEM images used in this chapter were processed after acquisition using a combination of programs. This processing includes correcting for the background intensity from the original images, image alignment, and some spatial averaging was applied for some techniques. The magnification calibration data was fast-Fourier-transformed for lattice-fringe analysis where required.

The first step in analyzing the HAADF data was to subtract off the background intensity (see [21] for further discussion of this). The intensity in vacuum should be zero, as the vacuum should not scatter electrons into the HAADF detector. The background intensity is controlled by the “brightness” detector setting. This setting is an additive value; thus, by subtracting off the average intensity in vacuum, the HAADF images were background-corrected.

Tomography data were aligned systematically using ImageJ [101] as follows. The stack of images *S* was copied, making an identical stack *R*. A Canny-Deriche edge-detection filter [102] was applied to *R* to highlight the nanowire edges. Then, to bring each image in the stack into registration, image shifts were calculated sequentially from *R* using cross-correlation on adjacent images, then these image shifts were

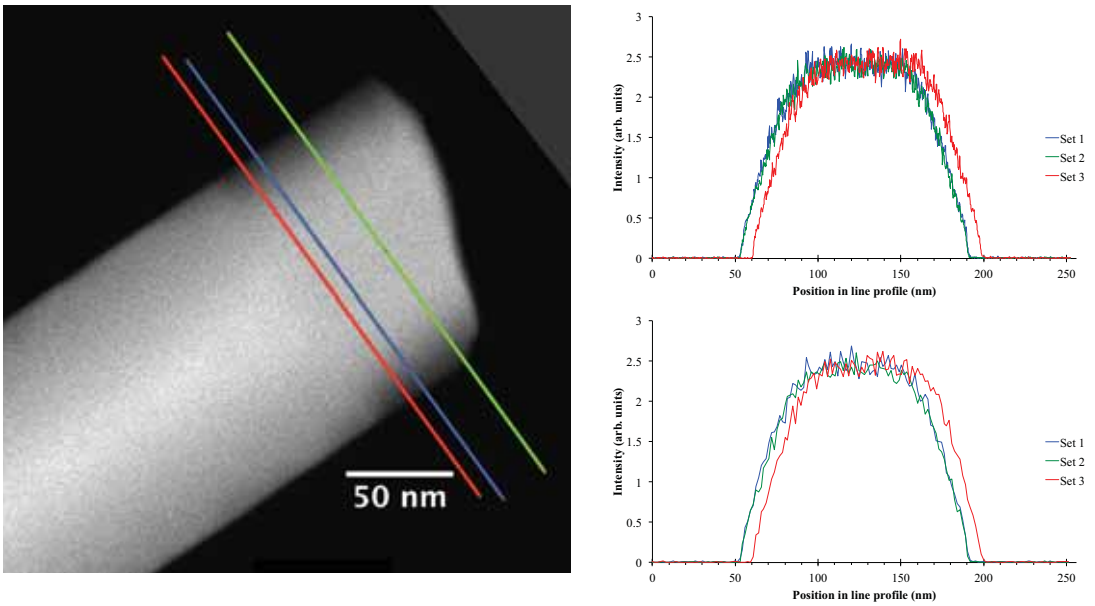


Figure 3.1: HAADF STEM image and intensity line profiles of InAs nanowire. The image and the line profiles are displayed with the microscope's as-acquired spatial calibration, which is ~2% too large. The wire diameter is ~133 nm. This is recalibrated according to Figure 3.2. The blue, green, and red lines indicate where electron holography line profiles are taken in chapter 4, and where reconstructions are taken in chapter 3. The grey triangle in the upper right of the image is due to image rotation during processing. The lower graph is the upper graph averaged to approximately the spatial resolution used for the phase and amplitude in chapter 4.

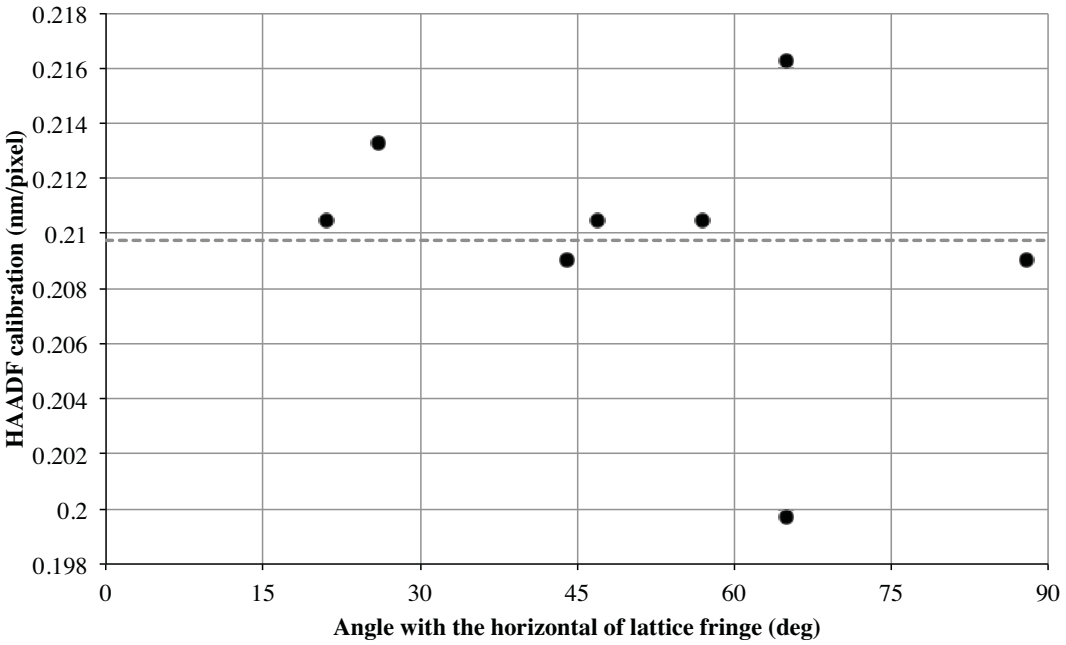


Figure 3.2: Test for angular dependence of HAADF spatial calibration. In the plot of the angular dependence of the HAADF STEM spatial calibration, performed on lattice-fringe images on a Si bulk-like calibration specimen, no systematic angular trend is present, although there is a noisy pair of points around 65°. The dotted line at the average of these calibrations (0.2097 nm/pixel) is to be compared with the nominal calibration of the microscope, 0.2129 nm/pixel.

applied to *S*. This provides a stack of images where the nanowire is in the same position throughout.

There were additional alignment and averaging steps applied to the input data for some of the algorithms used here: RENOIR uses carefully averaged line profiles, and geometric tomography uses an additional alignment algorithm, and discrete tomography uses both of these. There does not appear to be significant variation in the cross-sectional profile along the region of the wire studied here (the red, blue, and green lines in Figure 3.1), so the nanowire cross-section in this region can be taken to be the same throughout the region. Therefore, for the input data to the RENOIR thickness-determination technique, the intensity in the HAADF images of the wire is averaged over the homogeneous region of interest to increase the signal-to-noise ratio. Given the high inner detector semi-angle, the original data are somewhat noisy, as it captures relatively few scattered electrons, so improving the signal-to-noise ratio may help. These data in the region of interest were re-aligned in each slice before averaging, in order to retain both the features of the wire and the original tilt axis. This averaging technique should retain the same cross-section, only with less noise. This procedure reduces each image in the HAADF tilt series to a one-dimensional intensity line profile. For the geometric tomography reconstruction detailed here, the input image stack was additionally run through a self-consistent iterative alignment algorithm developed by L. Houben [103], which yields a stack images as output. The discrete tomography data was run through both of these routines: first, the self-consistent iterative alignment algorithm on the images, then the profile averaging to reduce the noise.

This nanowire HAADF data are then used as input into the reconstruction algorithms. However, to obtain quantitative information measured in nanometers from these images measured in pixels, there must be a spatial calibration. The spatial calibration is examined in the next section.

The Importance of Calibration

The spatial calibration is the most vital single component of the quantitative thickness determination. The spatial calibration to turn pixels into nanometers allows the tomographic reconstructions to be directly convertible into thickness profiles, the widths for RENOIR to be expressed in nanometers, and for the data from BFTEM and HAADF STEM to be directly compared.

A good spatial calibration is vital for interpreting the tomography reconstruction as quantitative cross-sections. The acquired images are projected in the z-direction, so the images are all in the x-y plane, and the reconstructions can be in the x-z or y-z planes. The spatial calibration for the x and y directions in the original images would be the same as that in the z-direction in the reconstructed cross-sections[16]. Therefore, with a spatial calibration, the cross-sections generated by tomographic reconstruction can be used for a direct measurement of the thickness.

Additionally, the precision of the spatial calibration for thickness determination in this chapter will directly affect the precision of the holography results in Chapter 4. As the phase measured using electron holography is directly proportional to both specimen

[16] Provided that the angular tilt calibration is accurate. This is not a routine user-serviceable alignment, and is thus taken to be accurate, because an inaccuracy in the tilt step may cause the specimen holder to collide with the objective lens. Chapter 4 briefly discusses the accuracy of fine tilt steps (0.1°).

thickness and mean inner potential, any inaccuracy or imprecision in the thickness determination will directly affect the mean inner potential measurement. This is discussed more in Chapter 4.

Both BFTEM and HAADF STEM images are used for calibration in this chapter, and, by combining them, more information is available about the specimen. The BFTEM data, in Figure 3.3, shows both the nanowire core and an amorphous shell encasing it, allowing for the widths of both to be measured and quantified. The amorphous shell is expected due to the long exposure of the nanowires to air between growth and characterization. On the other hand, whether the shell is or is not contributing to the HAADF images is unknown without additional, non-HAADF information. While the shell should give some HAADF signal, the density, composition, and lack of crystalline structure render the HAADF intensity not necessarily predictable. An elementary calculation would assume that the core is InAs (atomic numbers Z_{In}=49 and Z_{As}=33) and the shell is *e.g.* C (Z_C=6) of the same density. Thus, if the intensity were dependent on $Z^{1.7\text{-}2}$, according to [54] as noted in Chapter 2, then, even using a conservative estimate, the core would yield well over an order of magnitude higher HAADF image intensity than the shell. Therefore, it is reasonable that the shell may not be detectable in HAADF images. However, by measuring the width of the InAs nanowire from both BFTEM and HAADF images, it is possible to determine whether the HAADF intensity includes the shell or not.

Calibration of BFTEM Images

The BFTEM images are calibrated by reference to the known lattice parameter of gold. A commercially available polycrystalline gold nanoparticle sample was used, with particles suspended on a carbon film. Bright-field TEM images were taken at the same optical conditions, under which <111> lattice fringes were resolvable. A fast Fourier transform (FFT) of the BFTEM image, then a radial average of the FFT, allowed for calibration. The radial average of the FFT has peaks corresponding to the {111} planar spacing, and this spacing is 0.2355 nm for Au [104]. From the measured spacing and the known distance of that spacing, the BFTEM image is calibrated. This polycrystalline Au method may be less accurate than the single-crystalline Si specimen used in the next section if the Au nanoparticles have a size-dependent change in their lattice parameter; the Au specimen was used instead of the Si one because the Au specimen was available at the time.

Calibration of HAADF Images

The HAADF images are calibrated by reference to the known lattice parameter of silicon. A commercially available reference sample is, like that detailed in [105]. This reference specimen has a manufacturer's precision of <1%, and consists of a prepared section of silicon cut so that the low-order zone axis [110] is easily accessible. This leads to {111} lattice fringes in the images, which are spaced 0.3136 nm apart for Si [104]. A FFT of this lattice-fringe image yields discrete spots corresponding to these planes.

Angular Dependence of HAADF Calibration

Each Si {111} spot should be equidistant from the center of the FFT. However, this may not necessarily be the case for the STEM images in this work, as the STEM probe must raster over the specimen. The distance that it rasters in the *x*- and *y*-directions (*x*-direction being the one between individual points, *y*-direction being the one between lines) is a part of the TEM alignment file. Therefore, the HAADF images

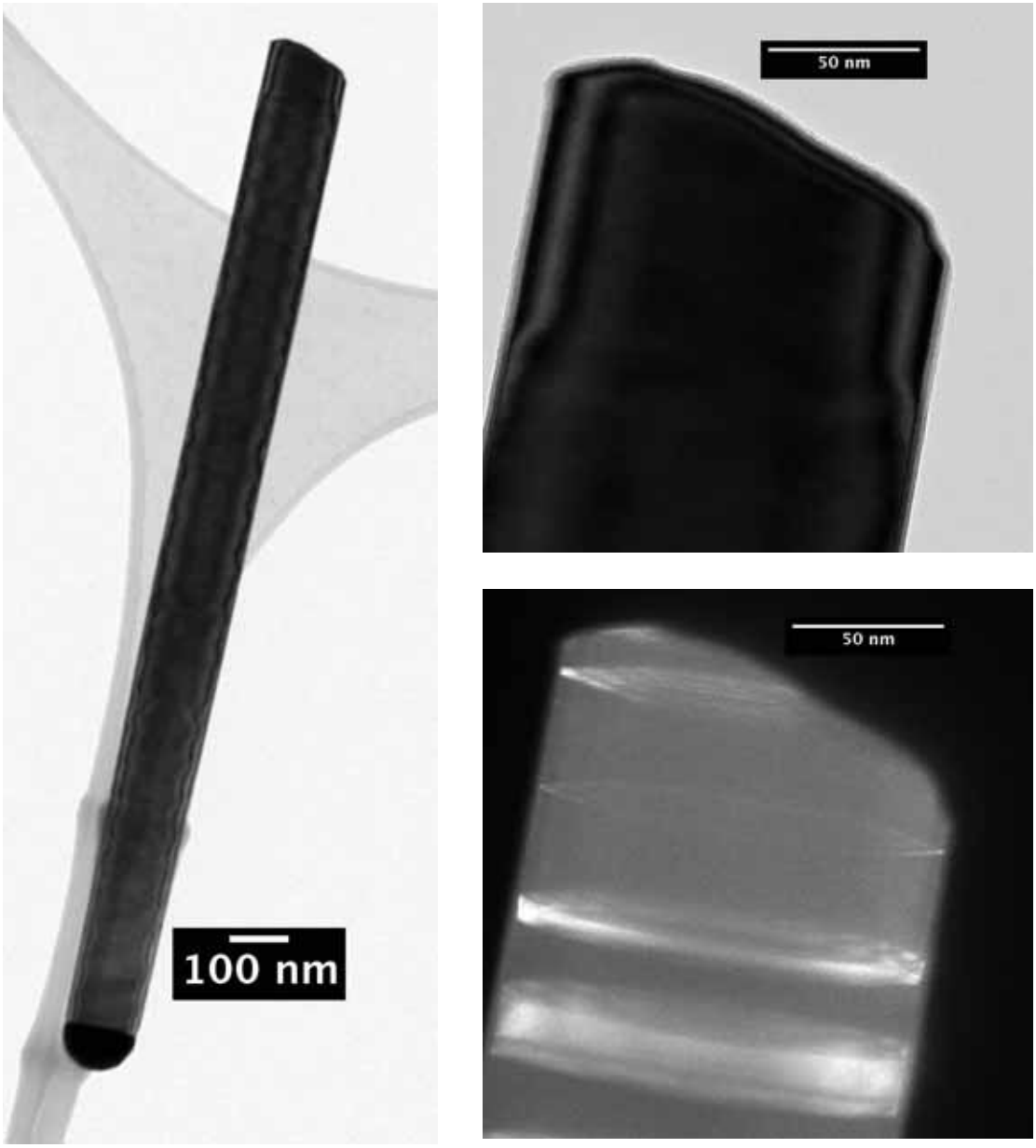


Figure 3.3: Calibrated bright- and dark-field TEM images of the whole InAs nanowire and the region of interest. Bright-field images at left and upper-right; dark-field image in lower-right. This nanowire is used in both Chapters 3 and 4. From the bright-field TEM (BFTEM) image in the upper-right, taken in a plateau orientation used for holography (see Figure 3.3) the width of the core of the nanowire is measured at 133.6 ± 1.3 nm, and the width of the core and shell together is 142.0 ± 1.4 nm. Note extinction contours from diffraction contrast in the bright-field images above. The nanowire has several crystallographic defects, seen in the dark-field TEM (DFTEM) image at lower right from a dark-field orientation chosen to remove extinction contours and highlight the defects, but not recorded. Note that the left image also appears in **Paper I**.

have a possible reason to have a difference in calibration between the *x*-direction and the *y*-direction. If that were the case, the disparity in the calibrations from different spots in the FFT might be determinable by the calibration of individual points at different angles. A systematic difference in calibration would make the unit circle of magnification into a unit ellipse. The direction of the nanowire in the HAADF images must be stated for the exact calibration to be verifiable. Happily, the nanowire was pointed in almost exactly the *y*-direction when the HAADF images were acquired, meaning each nanowire cross-section is in an *x*-*z* plane.

In the event of an orientation-dependent spatial calibration, the *x*-direction calibration can be used to calibrate the images, including the tomographic reconstructions. This is because each nanowire cross-sectional slice through the reconstruction volume is an *x*-*z* plane. Since the reconstruction process acts on one *x*-*z* slice through the reconstruction volume at a time, and an *x*-*z* slice only includes the calibration in the *x*-direction, that is taken to be the calibration of each slice of the tomographic data. If *y*-*z* slices were of interest for this work, their calibration would also need to be determined.

In Figure 3.2, the orientation dependence of the HAADF image calibration is plotted. This compiles the data from multiple images, taken with the specimen lattice rotated to different angles. This lattice rotation is performed by removing the specimen from the TEM and seating it at a different orientation in the holder. Figure 3.2 shows that the HAADF image calibration varies randomly, not systematically, with tilt angle. The dashed line on the graph is the average spatial calibration: 0.2097 nm/pixel, rounded here to 0.210 nm/pixel. This is ~2% different from the as-acquired calibration of 0.2129 nm/pixel. The maximum and minimum values occur at almost the same orientation, which indicates that the spread in values is likely due to random error, given that they are measured on the same bulk-like Si specimen, and the lattice parameter of the specimen is taken to be that of bulk Si. These two extreme points are +3.1% and -5.8% from the mean. If the outlier points were the result of random errors from those images that also appeared during HAADF tilt series acquisition, then there would be problems with alignment of the tilt series. However, since no such errors were encountered, the outliers are likely a random occurrence. After these two outlier points, the maximum measured deviation is +1.7%, and the rest of the points are close to the average. Based on Figure 3.2, it is assumed in this work that the calibration imprecision in the HAADF images is ±1.5%.

Nanowire Width After Calibration

The acquired bright-field TEM image[17] was calibrated as detailed above, yielding 0.0877 nm/pixel and is depicted in Figure 3.3. In the bright-field TEM image in Figure 3.3, a clear distinction is visible between the core and the shell of the nanowire. Measuring the core and the shell of the nanowire yielded a nanowire diameter of 133.6 nm for the core, and a total additional diameter of 8.4 nm for the shell (4.2 nm radial thickness). The previous BFTEM spatial calibration carried out separately by Chris Boothroyd, when applied to this image, would yield a nanowire width of 133.3 nm; a difference of 0.2%. Polycrystalline Au, however, may have inaccuracies in lattice parameter for small particles, as previously measured in [106]. Because the specimen in this work is made from Au in an agglomerated film of

[17] This image was taken at a stage goniometer tilt of 7° relative to the horizontal, which is chosen as the 0° orientation in the following holography chapter, due to its crystallographic orientation.

nanoparticles approximately ~10-20 nm in diameter, according to [106], particles of this size correspond to a lattice parameter change of ~1%. This work then assumes a 1% margin of error on the BFTEM calibration, giving a bright-field width at this orientation of 133.6 ± 1.3 nm for the InAs core, and 142.0 ± 1.4 nm for the core and shell combined.

The HAADF STEM images, when calibrated, yielded the results seen in Figure 3.4. However, determining the exact width of the nanowire from images like those in Figure 3.4 is not straightforward, because the STEM probe has broadened the abrupt interface here between vacuum and material. This beam broadening leaves low levels of residual intensity in vacuum around the wire. This is detailed more in the RENOIR section later in this chapter, as RENOIR relies upon a maximally accurate width determination. From that section, the optimal measurement of the width of the wire from the HAADF images is 134.0 nm. As determined in the previous section, this is subject to a ±1.5% margin of error, yielding a width of 134.0 ± 2.0 nm.

The measurement from the calibrated HAADF and BFTEM images indicates that the HAADF images are only capturing the core, without the shell. The width measurement of 133.6 ± 1.3 nm from the BFTEM image of the core alone agrees with the width measurement of 134.0 ± 2.0 nm from the HAADF image. While the HAADF width measurement involves some uncertainty, there is enough of a margin – approximately 6% - between the core-only and core-and-shell widths that the two cases can be distinguished.

As detailed above, this is an understandable result. The HAADF images are capturing only the core likely because of the significant difference between the known density of the core and the assumed density of the shell. However, while this does not greatly impact this chapter, it has significant ramifications for the holography analysis in Chapter 4, and is discussed further there.

Eliminated Techniques

Several TEM acquisition techniques were considered for thickness determination; however, the one that was deemed most useful was HAADF STEM imaging. This section discusses techniques that were not used for thickness determination, and why they were not used. While HAADF STEM is not the only possible choice (see [79] for a potential alternative: weak-beam dark-field TEM[18]), it was the most convenient and straightforward method in this work.

Why Not Conventional Dark-Field TEM?

Conventional dark-field TEM might seem, at first glance, to be a good candidate. The stacking fault in the middle of the region of interest, visible in Figure 3.3, looks to yield a faithful cross-sectional profile, which might be convertible into a cross-sectional thickness profile if the tilt angle of the nanowire relative to the specimen plane is known.

[18] Weak-beam dark-field (WBDF) (the fundamentals of which are in [1]) was not used in this work (like it was in [79]) for two reasons: First, the nanowire in question was mounted in a single-tilt specimen holder, greatly limiting the tilt range and meaning the careful diffraction condition required for WBDF imaging was not achievable. When the specimen grid was carefully re-mounted in a double-tilt holder, the specific nanowire was no longer present, evidently having been detached during handling. Second, WBDF does not yield a thickness profile in the same way as HAADF STEM – it yields thickness contours rather than fully resolved monotonic intensity-thickness. This would be useful complementary information – if it had been possible.

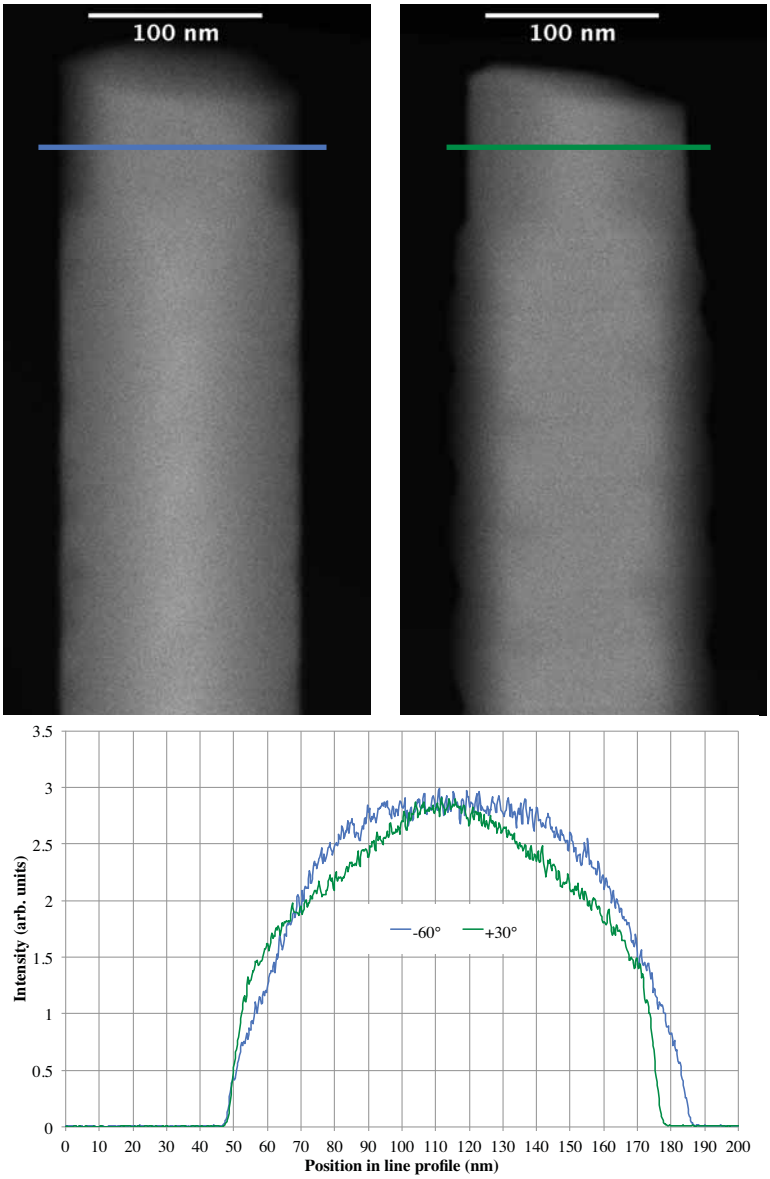


Figure 3.4: Orthogonal HAADF images of InAs nanowire. The green HAADF line profile shows the "house" orientation, and, orthogonally, the blue line profile at the "plateau" orientation. This work focuses on the "plateau" orientation. Intensity not in counts, but in arbitrary units, and with vacuum intensity set to 0 (see text From same dataset as used in **Paper I**.

As an illustrative exercise, it is possible to measure the nanowire thickness from this image. One simple approach to determining the total thickness would be to take the dark-field image seen in Figure 3.3 and measure the projected width of the tilted stacking fault. This yields a tilted projected thickness of 12.2 nm, which, when the 7.66° angle the nanowire makes with the horizontal is taken into account, yields a total nanowire thickness of 91 nm. This is assuming the crystal tilt measured from the Kikuchi pattern is correct and the dark-field beam tilt is irrelevant.

This nanowire thickness is much smaller than that yielded from other methods, which are all about 40% larger. This is likely because the incident tilt angle of the beam relative to the sample is not known in this particular experiment. In the dark-field TEM imaging mode, the incident beam is tilted off the optical axis in order to provide a beam corresponding to diffracted intensity down the optical axis, which is then selected with an aperture. This means the beam is tilted relative to the sample. This incident beam tilt is not known in this experiment, because the beam was tilted until the "cross-sectional" stacking fault was maximally visible. This tilt would affect the projected width, as the tangent used in this case is very sensitive to small changes in tilt angle. As a result, this thickness number is not used or suitable for the quantitative analysis later.

Ideally, the thickness information would come from a technique that does not have this misorientation. The next section considers electron holography.

Why Not Amplitude Images from Electron Holography?

Intuitively, it might be ideal to use thickness information that was acquired using electron holography at the same time as the phase information from electron holography. No image alignment would be necessary. However, diffraction effects would possibly be a problem, and using the amplitude and phase together is deeply suboptimal because their errors are inversely correlated with each other.

As detailed in Chapter 2, amplitude information from the nanowire can be turned into thickness information by taking the logarithm, which yields a t/λ plot. If the maximum thickness of the nanowire is known, the t/λ plot can be calibrated. However, dynamical diffraction, as detailed in Chapter 2, would affect both the amplitude and the phase of holograms. Therefore, the t/λ plot may no longer be a monotonic representation of the thickness, and using the diffraction-affected amplitude and diffraction-affected phase together may yield unpredictable results.

Even if there is no diffraction contrast, amplitude-derived thickness information is not practical to use for phase images. While it is theoretically possible to apply this method for thickness determination for a phase image [42], it is undesirable because the errors in amplitude and phase are inversely correlated. A region with heavy amplitude attenuation (and, therefore, low amplitude noise) will have a marked decrease in fringe contrast, yielding a noisy phase signal. However, a region with good fringe contrast and, therefore, low-noise phase information will have minimal amplitude attenuation and, therefore, noisy amplitude information [27]. This compromise is mathematically unavoidable if both phase and amplitude images are used together. Therefore, a different thickness-determination technique should be used.

Verifiable Features Visible In Images

Some features are directly visible in BFTEM or HAADF STEM images of the nanowire. These features must be reproduced in any tomographic reconstruction or intensity-normalization technique for the result to be trustable and verifiable as accurate. Failing this test does not mean that the algorithm is fundamentally flawed, and passing it does not mean that it is of general applicability. These observations and requirements are not intended to be universal, but might be taken as generally useful for this class of specimen and this application.

From the images (*e.g.* Figure 3.3), the nanowire region of interest is suspended over vacuum. This is visible in the BFTEM images (Figure 3.3) as a bright region around the darker nanowire, and in the HAADF images (Figure 3.1 and Figure 3.4) as a dark region around the brighter nanowire. This means the only material in the region of interest belongs to the nanowire. While there is lacey carbon film visible in Figure 3.3, it is not at the region of interest. This is the result of suspending the nanowire on the lacey carbon film during specimen preparation, such that some parts of the wire are over the lacey carbon film, and some are over vacuum. Any reconstruction needs to reconstruct both the nanowire and the empty space surrounding it.

The clearest quantifiable feature of the nanowire is the width of the InAs core, seen in Figures 3.3 and 3.4. This is measurable directly from the images, and provides critical calibration information across techniques, including for electron holography in the next chapter. A more comprehensive look at the InAs width measurement is in the RENOIR section in this chapter, and this is also briefly considered in Chapter 4.

Another clear feature is the truncated-corner hexagonal cross-section. This is visible from the HAADF line profiles in Figure 3.4. This means that the nanowire cannot be taken to be a regular shape, and the morphology of these corners must be determined. This truncated-corner hexagonal cross-section should be reproduced in a tomographic reconstruction of the nanowire. The form of the cross-section is not a problem for RENOIR, as RENOIR simply recalibrates this line profile, so it is automatically consistent.

There are several nanowire features visible in the original image data: truncated-corner cross-sections, a nanowire surrounded by vacuum, and the core width. These clear features must be present for each thickness determination technique to be trusted – hence the calibration. If these features are not present in the cross-section determined by a given technique, then that technique's cross-section is inaccurate. This provides for quantitative calibration of one technique using another. However, the properties of the surface layer are more difficult to determine.

The Surface Layer

There is a visible surface layer around the nanowire of approximately 3nm in thickness, forming a shell. This surface layer is of unknown composition, but perhaps is comprised of a native oxide or atmospheric hydrocarbon, and is referred to throughout this work occasionally as an oxide layer for this reason. The surface layer adds approximately 6nm to the total width. The surface layer is visible in the bright-field TEM images as an amorphous shell around the central core of the wire – see Figure 4.9 for one hypothesis as to what the surface layer morphology might be. This surface layer might be due to the exposure to air of the nanowires after growth, and attempts to remove the surface layer are discussed in Chapter 6.

The surface layer will remain the same throughout the characterization, but different techniques will capture it differently. This shell should be visible in both holography (phase and amplitude included) and bright-field TEM. The surface layer poses a larger practical problem in HAADF image interpretation. The above sections indicate that the nanowire width as measured using HAADF does not include the shell. However, there might be beam-broadening effects due to the probe size, and the unknown density and unknown composition of the surface layer introduces some ambiguity. What is clear about the oxide layer, despite the above unknowns, is that it is present in the TEM image modes and is unlikely to be present in the STEM image modes.

Electron Tomography

From noting direct features of the images, the next step is to perform and analyze tomographic object reconstructions generated from the experimental tilt series of HAADF images. Electron tomography allows for reconstruction of an object from individual projection images acquired at different tilt angles, as noted in Chapter 2, see [57] for further details. Electron tomography reconstruction is discussed in Chapter 2, and contrasts with the direct intensity-calibration method of RENOIR used later in this chapter.

Tomographic reconstructions can be performed using different algorithms. Here, five different reconstruction algorithms are each independently applied to the HAADF STEM tilt series. These five algorithms fall into three different categories: algebraic tomography (AT), geometric tomography (GT), and discrete tomography (DT). First, three conventional AT algorithms for reconstructing specimens comprised of arbitrary density distributions (detailed in [57]) are used independently of each other – weighted back-projection (WBP), algebraic reconstruction technique (ART) and the simultaneous iterative reconstruction technique (SIRT). These AT results do not yield quantitatively interpretable cross-sectional profiles of the nanowire, and their discussion is in Appendix B, with a brief summary in this chapter. Then, in the following section, one GT algorithm is applied that reconstructs the shape of the specimen instead of the density distribution. This algorithm is also detailed in [58] and applied in **Paper I: Reconstruction of an InAs nanowire using geometric and algebraic tomography**, included in this dissertation. However, the GT algorithm used here is not completely suitable for the data, as there is a significant missing wedge. Therefore, in the section after, a DT algorithm is applied: DART, which is detailed in [60], with several additional examples in [6]. DART assumes discretized intensity levels in the object to be reconstructed, which yields a more complete picture of the cross-section. For all three types of algorithms (AT, GT, and DT), the reconstruction artifacts and assumptions are examined. However, only DT yields a fully interpretable nanowire cross-section.

Determining Thickness From Tomography

To be useful in this dissertation, once the tomographic reconstruction is performed, the reconstruction must be translated into a nanowire cross-section, which must be translated into a one-dimensional cross-sectional thickness profile. This is true for any algorithm.

While tomography solves the inverse problem of determining the object itself, this full solution is of limited use for this experiment. The primary quantity of interest is the projected cross-section, such as might be obtained in a single HAADF image or

phase image from electron holograms. Therefore, no matter which reconstruction algorithm is applied to the HAADF STEM data, the reconstructed object must be forward-projected as if it were a real object, rendering a one-dimensional line profile from a two-dimensional reconstructed object cross-section. In principle, this allows for both verification of the tomographic reconstruction and sensible object precision. Verification, as discussed above, is through comparing known nanowire features with features in the reconstruction. The nominal real projected cross-sections are known, including some noise, but the tomographic forward-projections, by using more data, might have less noise.

Prior knowledge plays an important role in determining the faithfulness of the reconstruction. The prior knowledge can be applied that, neglecting the surface layer, the nanowire is a binary object, thus the density inside the wire does not vary. The surface layer can be neglected because it does not appear in the HAADF images, according to measurements of the width of the wire. Therefore, the reconstructed greyscale object could potentially be binarized before re-projection into a line profile; however, if the reconstructed object itself is used to determine this threshold, then misinterpretable artifacts may become more likely. A good threshold is to use the width measured from the experimental HAADF images. This width is internally consistent, because the reconstruction procedure uses the same images. If a different threshold were used, then the reconstruction would be clearly unfaithful to the original object. This width is then used to determine the threshold to be applied to the reconstruction such that the width of the thresholded reconstruction and the width of the nanowire as measured are the same. This is appealing because it gives a threshold determined from a specific, quantitative experimental quantity.

It is perhaps worth reiterating that three selected cross-sections are used for both holography profiling and tomographic profiling; where colors are used, they are colored here blue, green, and red. See Figure 3.1 for the locations of these cross-sections.

Algebraic Tomography

Algebraic tomography in this work is more fully explained in Appendix B, because it was found to be unsuitable for quantitative thickness analysis. A brief summary of the AT results is provided in this section.

AT applied to TEM data solves the inverse problems of object morphology determination by taking the intensity values from a series of images acquired at different tilt angles and determining a greyscale object from it. It does not put constraints on the density or intensity distribution of the object, but solves it generally.

Appendix B shows the results of applying three different algebraic tomography algorithms to the data – the weighted back-projection (WBP), the algebraic reconstruction technique (ART), and the simultaneous iterative reconstruction technique (SIRT), implemented in the TomoJ software package [107]. However, the tomographically determined cross-sections from these algorithms require thresholding to interpret their cross-sections as thickness information, and this processing renders the output of WBP, ART, and SIRT not quantitatively interpretable.

This challenge arises because there are only two potential densities to be reconstructed from the HAADF STEM information: the InAs nanowire core, and

vacuum. Therefore, to interpret the greyscale reconstruction as a thickness profile, the reconstructions for all three AT techniques must be thresholded. The post-thresholding results do not correspond to the wire, and are described in Appendix B. This fundamental compromise leads to the desire for a different technique that does include this information, and, therefore, to either geometric tomography or discrete tomography, both of which are treated in the coming sections.

Geometric Tomography

While algebraic tomography involves determining a greyscale image of the intensity distribution of the object under study, geometric tomography uses a fundamentally different set of assumptions that are more valid for this experiment. It assumes that what is to be reconstructed is the outline of a convex object, which is a good assumption for this nanowire. Therefore, before performing any reconstruction steps, the tomography input data is thresholded to determine the object outline, and a polygon is fitted to the outline at each tilt angle [58]. More details can be found in the attached **Paper I: Reconstruction of an InAs nanowire using geometric and algebraic tomography**. One critical drawback of GT is that analysis must completely neglect the missing wedge, as there are no orthogonal projections to take an outline measurement from. However, it might be useful in the non-missing-wedge region.

Cross-sections reconstructed using GT are seen in Figure 3.5, from the same dataset as in **Paper I**. The pointed tops and bottoms of the nanowire are clearly missing-wedge artifacts, but the left and right sides are potentially usable[19]. The ends of the missing-wedge region are clearly defined, and the artifact-free region that is reconstructed can be selected.

Forward-projecting the reconstructed cross-sections seen in Figure 3.5 results in the line profiles seen in Figure 3.6 on the left; the difference in vacuum level is simply to be able to show all three forward-projected line profiles on the same graph without overlap. Removing any region that has missing wedge artifacts yields the graph seen in Figure 3.6 on the right. This has well-defined sloping regions, and would allow for the thickness in the wedge regions to be determined from the calibration of the tomography images. Figure 3.7 shows forward-projections after rotating the nanowire to the orientation used for electron holography (7° off horizontal), with the left graph in Figure 3.7 including the missing-wedge artifact, and the right graph in Figure 3.7 with the missing wedge excluded. Figure 3.8 shows the results of re-sampling the right graph in Figures 3.6 and 3.7 to the holography calibration for direct comparison.

Why Not Geometric Tomography?

GT does determine the object cross-section in a manner consistent with the object itself, because it reconstructs a binary object. AT assumes that what is to be reconstructed is an arbitrary density distribution, but GT, as implemented here, attempts to reconstruct the outline of a convex object, which, for this particular nanowire, is a good assumption, given that the nanowire is a binary object.

However, the missing wedge in geometric tomography, clearly visible in Figure 3.5 poses a significant and obvious problem. If the maximum thickness of the wire is not known, then the analysis is limited to only the thickness wedge regions, and not even

[19] Because the missing wedge is slightly offset, the reconstructions in Figure 3.5 appear asymmetric.

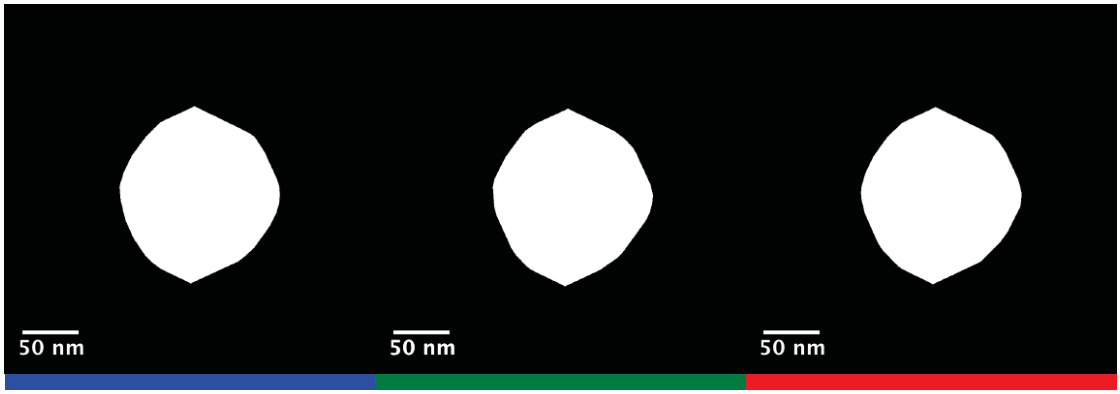


Figure 3.5: Geometric tomography reconstructions, selected cross-sections. The missing wedge at the top and bottom of each reconstruction leads to a pointed, house-like reconstruction, where there should be plateaus on the top and bottom. The colors correspond to the colored line profiles in Figure 3.1. From same HAADF STEM dataset as **Paper I**, as are subsequent geometric tomography figures.

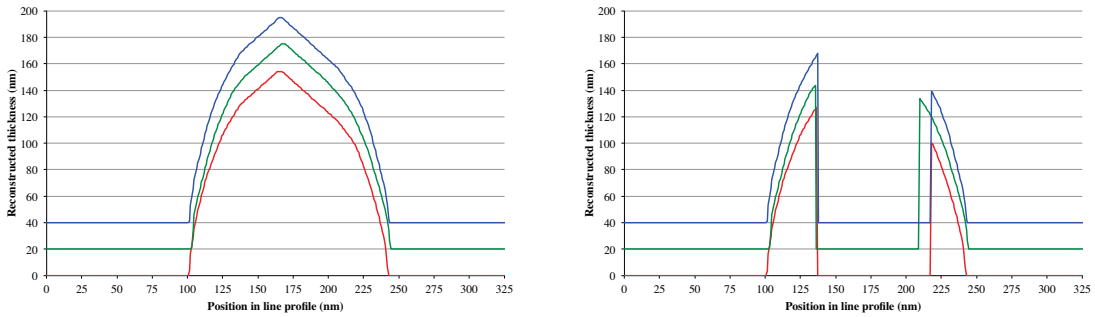


Figure 3.6: Forward-projected geometric tomography reconstructions. The line profiles generated from the forward projection of the geometric tomography reconstructions in Figure 3.5: (left) with and (right) with the missing wedge regions set to 0. Profiles are offset in thickness solely for clarity.

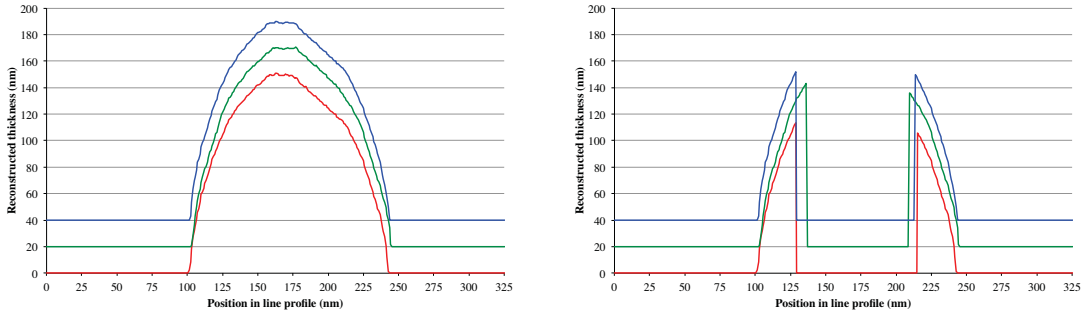


Figure 3.7: Line profiles from forward-projected geometric tomography reconstructions, rotated to match holography orientation. The same missing-wedge challenges are present before and after rotation. Compare with Figure 3.6.

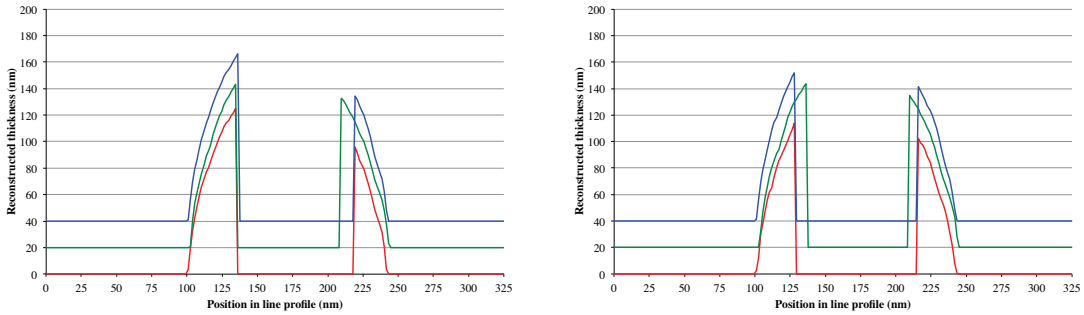


Figure 3.8: Geometric tomography reconstruction interpolated to the holography data sampling. Before rotation (left) and after rotation (right) to holography orientation. Compare to Figures 3.5, 3.6, and 3.7. The missing-wedge region makes this impossible to use for the maximum-thickness region.

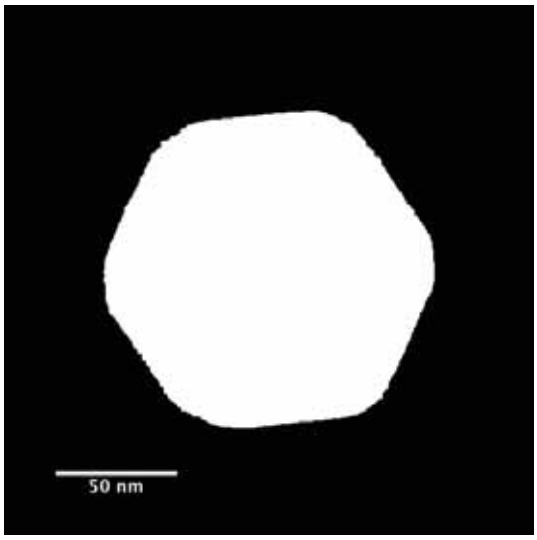


Figure 3.9: Discrete tomography (DT) reconstruction of nanowire. This reconstruction from HAADF tomography data made using the DART algorithm (see text). This method reconstructs a hexagonal cross-section of the nanowire with truncated corners. Magnification calibration from HAADF images.

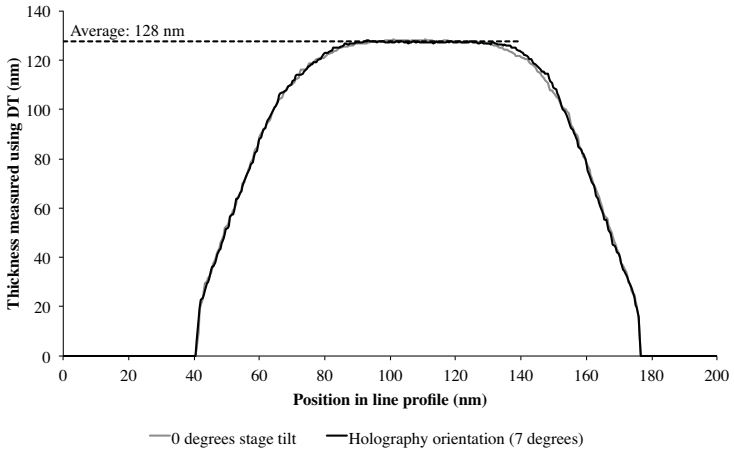


Figure 3.10: Forward-projection of DT reconstruction of nanowire. The reconstructed cross-section is shown in Figure 3.9. For the tilt range examined using electron holography in chapter 4 (+2° to +13° stage alpha tilt), there is almost no cross-sectional change. Compare to HAADF profiles in Figures 3.1 and 3.3.

their full span, but a very limited portion. This throws away much of the useful holography data.

More subtly, the determination of the nanowire pre-reconstruction intensity threshold is subject to a certain degree of interpretation. A different threshold intensity would yield a different reconstruction. In this work, a single threshold intensity was taken to determine the width of the nanowire, and was chosen to make the reconstruction optimally stable. As will be detailed on the section on RENOIR in this chapter, nanowire width determination is nontrivial, and should be considered very carefully.

It is worth noting here that the GT algorithm used is suboptimal for treating TEM data, as it is very susceptible to missing wedge information. The information used in this reconstruction is the boundary of the nanowire, measured in the plane of each image. However, other GT algorithms in the general literature (*e.g.* the one detailed in [108]) present more promise for electron tomography applications, because they can use the intensity, instead of requiring pre-reconstruction thresholding.

Therefore, the discussion turns now to a method that allows for the reconstruction of the binary object manner from non-binary input data: discrete tomography.

Discrete Tomography

Discrete tomography (DT) allows for determination of a binary object from non-binary input data. This is an accurate description of the nanowire in this work, as the specimen region in the HAADF images contains either nanowire or vacuum. Therefore, in this section, a DT algorithm is applied to reconstruct the nanowire. The DART (discrete algebraic reconstruction technique) algorithm was first introduced in [59] and applied more extensively in [60]. DART is applied to attempt to answer the question of the thickness profile of the nanowire. DART, as detailed in [59], reconstructs the specimen by determining the best-fit boundary to the object that would yield the input set of forward projections, and uses as prior information a certain intensity-thickness relationship in the specimen. This intensity-thickness relationship is considered prior information in [59], but is determined from a SIRT reconstruction[20] in [60].

The DT reconstruction of the cross-section of the nanowire is seen in Figure 3.9. Several of the verifiable features listed in the previous sections are immediately visible: binary object reconstruction, truncated-corner hexagonal cross-section, and no obvious missing-wedge effects (compare the GT reconstructions in Figure 3.5).

First, the hexagonal corners on the reconstruction in Figure 3.9 appear truncated, with small but distinct corner sub-faceting. This is consistent with the original intensity profiles from the HAADF images, which show rounded corners at the vertices in the middle of the thickness profile, and a truncated point in the “house” orientation (see Figure 3.4. This is also important, as, in the reconstruction in Figure 3.9, all corners are truncated. Therefore, the corner truncation is likely, though not certainly, due to a physical change in nanowire morphology from an ideal hexagon, instead of nonlinearity in HAADF intensity-thickness response.

In the DT reconstruction in Figure 3.9, there are no obvious missing wedge effects, in contrast to the AT and GT reconstructions; the reconstructed nanowire cross-section

[20] This DT reconstruction does not use the SIRT reconstructions seen in Appendix B.

appears reasonable along all angles. This increases confidence in the nanowire cross-section determination.

In sum, the DT cross-section appears qualitatively reasonable. Because there is a spatial calibration and a binary object, the reconstructed cross-section can be directly forward-projected into a cross-sectional thickness profile. To do this, the cross-section in Figure 3.9 is rotated into the "plateau"-like holography orientation (7° off horizontal) and forward-projected, as shown in Figure 3.10. Qualitatively, the DT thickness profile compares favorably to the HAADF line profiles in Figure 3.1. From this orientation, the DT cross-sectional thickness profile shows low variation between the 0° and 7° line profiles, indicating that the same thickness profile is usable across the short angular range sampled for holography[21].

Therefore, the quantitative thickness profile from the DT reconstruction is useful, and will be applied in the holography chapter to determining the thickness. Now, the next section examines the potential downsides of DT – most importantly, that the intensity-thickness relationship DT depends on is not verifiable.

Why Not Only Discrete Tomography?

For the thickness determination in this chapter, discrete tomography is the most useful of the tomographic techniques described here, because it most accurately reflects the assumptions known to be true about the nanowire: the binary nature of the nanowire as imaged using HAADF STEM (excluding the surface layer), the rounded corners, and the general nanowire morphology. This reconstruction appears to be very faithful to what is known the original object.

Even though there is good agreement between prior knowledge and the reconstruction in this case, there still may be reconstruction artifacts or distortions. The thresholding process for DT occurs as part of the reconstruction algorithm, and is dependent on determining the correct intensity-thickness relationship, something that is not straightforward and the focus of ongoing research, as noted in [60]. On the other hand, an approach like the one in [100] may improve DT reconstruction of specimens like nanowires. This makes the DT reconstruction potentially easier to trust for a general specimen, as it does not rely on arbitrary, human-driven thresholding, but on a reproducible, rules-based metric. On the other hand, in this case, human-driven thresholding can make the reconstruction consistent with prior information like the known nanowire width, while the DT width is not chosen specifically to be consistent with the BFTEM width, and may incorporate beam tail effects from the STEM probe. Therefore, the DT reconstruction is not guaranteed to yield the correct values for known parameters.

The end result of DT is to yield a thickness profile that is likely trustable, but the intensity-thickness relationship is not guaranteed to be accurate. This is because, if the reconstruction algorithm determines the wrong intensity-thickness relationship, then the DT reconstruction will not be accurate, but may nevertheless appear accurate. This intensity-thickness relationship is determined from a SIRT reconstruction. If the SIRT reconstruction is not an accurate representation of the original object – and SIRT reconstructions can be tricky, as Appendix B shows – then the DT reconstruction will be flawed. Now, the dissertation turns to a technique to calibrate the HAADF profile

[21] The holography section uses -5° to +6° from the 7° stage-tilt holography plateau-like orientation, corresponding to +2° to +13° in stage-tilt.

directly, without tomographic reconstruction, through determining the best intensity-thickness relation for the HAADF tilt-series data.

RENOIR

In this section, the RENormalized Orthogonal-Images Routine (RENOIR) is invented and used to compare intensity and width along mutually orthogonal directions in the tilt series. The objective of RENOIR is to generate an intensity-thickness ratio to be used to calibrate the acquired HAADF intensity profile.

RENOIR relies on measuring the average maximum intensity in the nanowire in one cross-section with the nanowire width measured in the cross-section acquired after 90° of specimen tilt. This yields two measurements that measure the same quantity from two orthogonal images, where one image measures it in *e.g.* the *x*-direction (width of the nanowire, from the image), and the other image measures it in the beam direction (intensity at maximum thickness). This approach allows for non-tomographic quantitative interpretation of projection information. A sample pair of orthogonal images is seen in Figure 3.4. From Figure 3.4, the blue HAADF intensity line profile's width will correspond to the green HAADF intensity profile's average maximum intensity, and the green HAADF intensity line profile's width will correspond to the blue HAADF intensity profile's average maximum intensity.

The objective of RENOIR is to use the spatial calibration of the HAADF images in the tilt series to determine their intensity calibration. The spatial calibration of the HAADF images is known, but the calibration between intensity and thickness is not known. Therefore, if the width of the nanowire in one image is measured, and coupled with the maximum HAADF intensity measured in an image taken along an orthogonal direction, then this gives the intensity-thickness ratio. This intensity-thickness ratio is then used to translate the HAADF intensity cross-sectional line profile of the nanowire into a thickness cross-section, without the reconstruction artifacts introduced by the tomography algorithms detailed in the previous sections.

In this manner, the tilt series used for tomographic reconstruction in the previous sections is used here for a direct, non-reconstructed determination of the nanowire cross-section. Due to the limited tilt range of the acquired series, the direct width of the nanowire along the projections used for the holography measurement is not possible. However, as the same microscope settings were used throughout the HAADF tilt series (*e.g.* detector brightness/contrast and beam settings), the intensity-thickness data is internally consistent throughout the tilt series and can be compared between different images in the series.

To perform this intensity-thickness calibration, both the region of maximum intensity and the width of the nanowire in the cross-section of each projected image were identified. In this case, the total tilt series continuously spanned 140° with an image every 1°, but part of the end of the series was not usable due to intensity falloff from grid shadowing, yielding only 132° of continuous usable angular span. In this 132° span, there are 43 pairs of images (86 individual images) acquired from orthogonal directions. From each image, both intensity and width are measured, yielding 86 intensity-width pairs from orthogonal images. The full intensity-width series can be fitted linearly, with a forced 0 intensity = 0 width intercept (because the measured HAADF intensity should be 0 at 0 thickness). This yields the intensity-thickness calibration. Each intensity-width data points can also yield an intensity-thickness

calibration, through a linear fit to two points: the data point and a forced 0 intensity = 0 width intercept point; the ensemble of these individual fits is used for error analysis. This procedure is now carried out for the data with and without renormalization.

Simple Results Without Intensity Renormalization

Using a set of simple, incorrect assumptions – assuming that the recorded intensity is exactly reflective of thickness, neglecting beam-spread effects or renormalization – and the approach detailed above yields the results in Figure 3.11, showing intensities and widths. A threshold intensity of 100 counts (to overcome shot noise) is set for the width, and the intensity is used as-is. The results in Figure 3.11 show the results of measuring the maximum HAADF intensity of the nanowire (top graph), the width of the nanowire (middle graph), and co-plotting the width from each image and the intensity from the orthogonal images (bottom graph). This yields 86 pairs, as detailed above.

Performing this simple procedure yields the series seen in Figure 3.12. However, this spans a relatively short range of widths, as can be seen in the top part of Figure 3.12, and, while the linear fit appears locally to be reasonable, the display of total widths in the middle part of Figure 3.12 shows a broad distribution.

Since there are two known points on an intensity-thickness calibration graph for each intensity-thickness pair – the 0 intensity = 0 width point, and the intensity-thickness measurement – dividing the intensity by the thickness for each of the 86 pairs yields the intensity-thickness ratio for that pair. This is plotted as a histogram in the last part of Figure 3.12, with the 2-sigma error estimate shown calculated from the standard deviation of the full ensemble of 86 pairs. This shows a large imprecision in the determination of the intensity-thickness relation using this method.

Accounting for this imprecision, the HAADF line profile is calibrated, yielding the thickness profile seen in Figure 3.13. Figure 3.13 also displays the error estimate; as the imprecision is in a multiplicative quantity, there is a difference in the positive and negative error. While this simple method gives an intensity-thickness calibration, this calibration is not very precise.

To improve the precision of this method, both the intensity determination and the width determination can be improved. First, the intensity is renormalized to account for random differences in individual images, tightening the distribution. Then, the width determination is improved by identifying and optimizing an objective figure of merit.

Improving the Intensity Determination

This section discusses how to improve the HAADF maximum-intensity determination through renormalizing the HAADF intensity. The total area of the nanowire cross section in each projection should be the same, even if the distribution of that intensity changes at different tilt angles. This might occur due to diffraction contrast from a strongly diffracting orientation, or from fluctuations in the electron source. This can be compensated for by dividing the maximum intensity by the total integrated intensity, yielding the same total intensity in the sum of the cross-sectional line profile. This renormalization process is what gives RENOIR the first part of its name.

The effect of HAADF maximum-intensity renormalization is seen in Figure 3.14. The first part of Figure 3.14 plots the original as-acquired HAADF intensity against the re-

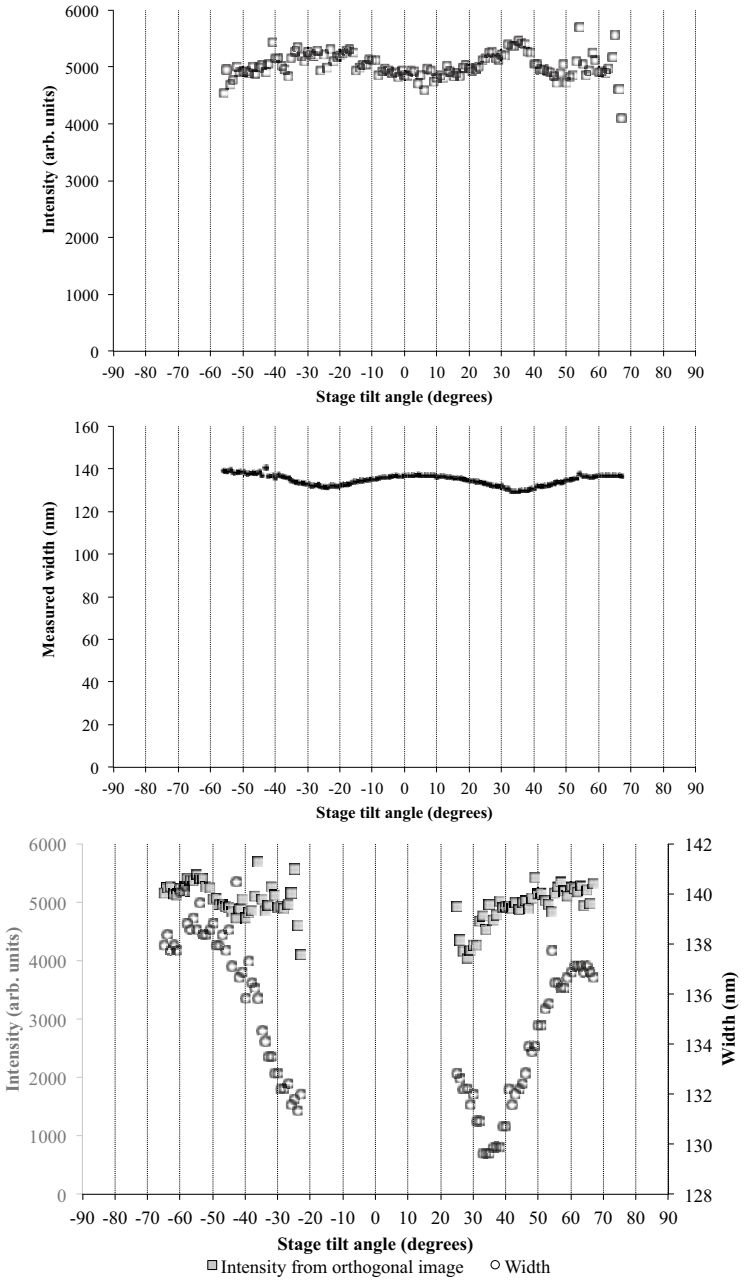


Figure 3.11: Intensities and widths measured from HAADF images in tilt series. Top: Average intensity at maximum-intensity region in line profile. Middle: Width from HAADF line profile at same tilt angles. Bottom: Widths (circles) from middle graph, and intensities (squares) from an orthogonal orientation in top graph, blank where pair does not exist, plotted for the orientations that have orthogonal images.

Images taken from across the tilt series described in chapter 3. Widths determined by original HAADF image threshold set to T=100 counts, as described in the text.

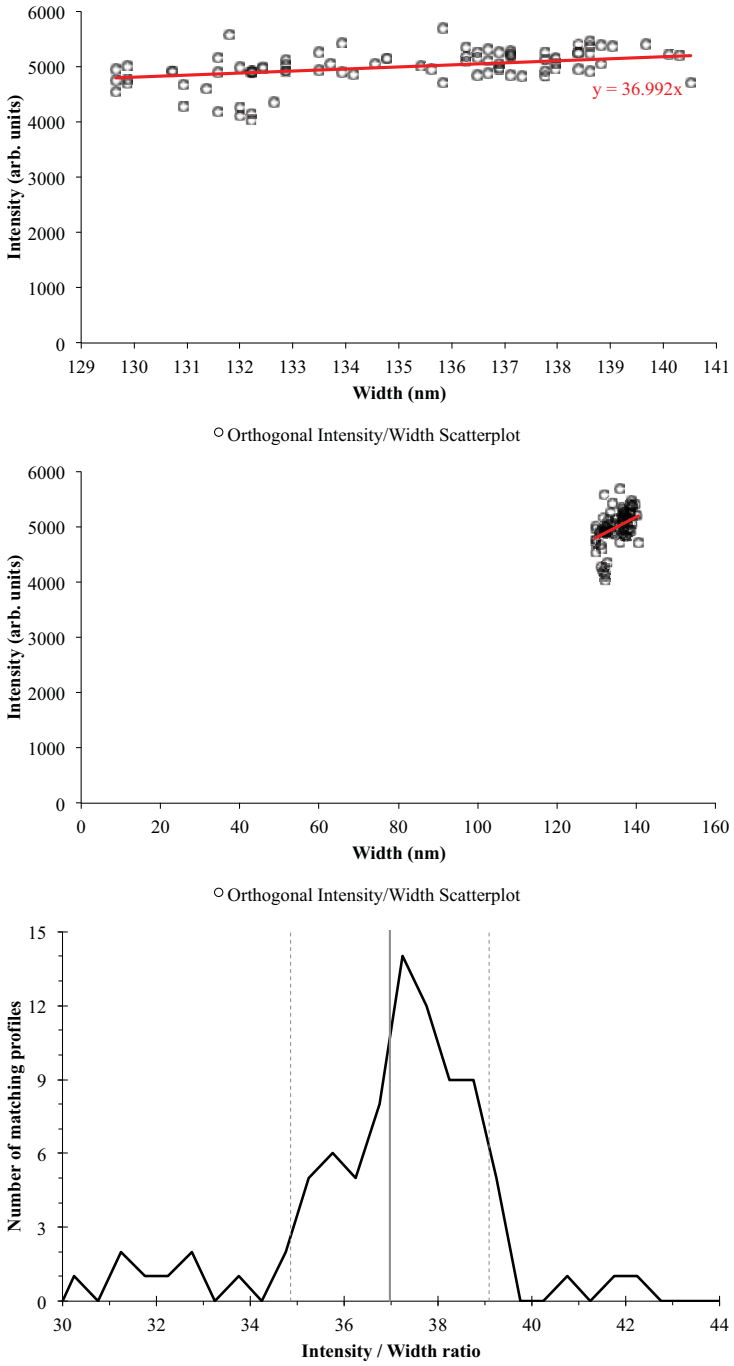


Figure 3.12: Intensity-thickness ratios. Widths are used to determine intensity-thickness relations. Top: Scatterplot and linear fit of the individual intensity-width data from Figure 3.11, bottom graph, with fit line forced to have a 0-intercept (zero intensity = zero width). Middle: Same as top graph, but scaled fully down to 0 intensity and 0 width, showing the limited range and high standard deviation. Bottom: histogram of the intensity-width ratios, with the average (solid grey) and the first standard deviation (dashed grey) indicated.

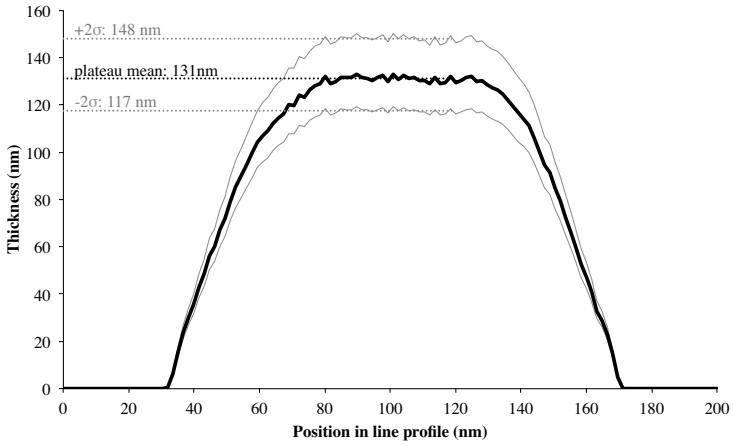


Figure 3.13: Thickness-calibrated HAADF intensity line profile, using Figure 3.12, with 2σ error margins indicated. Calibration and error estimates from values in Figure 3.12. The error is too high to be usable. These are very wide margins, and lead to renormalizing the intensity in the subsequent graphs.

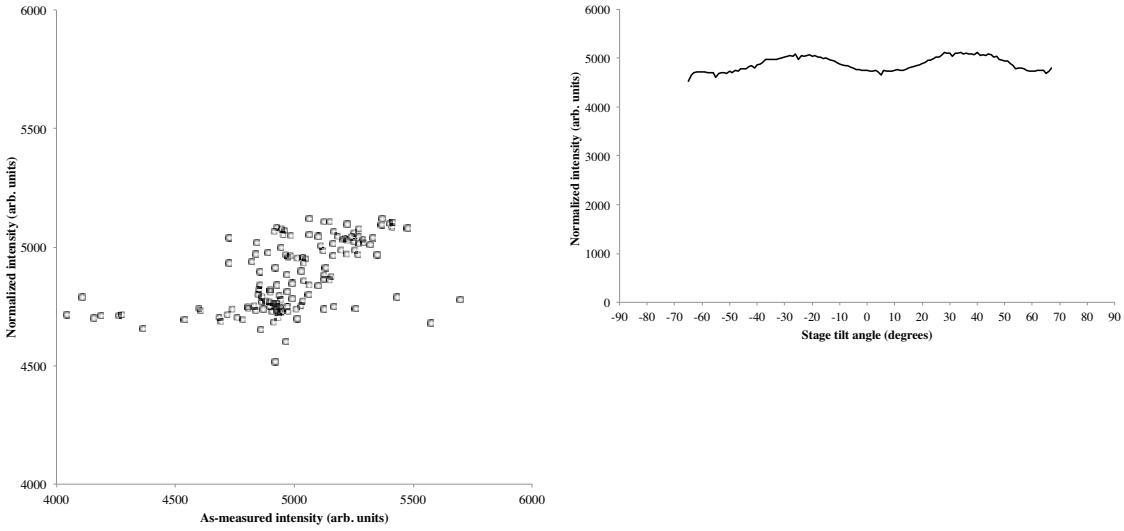


Figure 3.14: Intensity renormalization on the HAADF images. By renormalizing the intensity such that the total cross-sectional area remains constant across the series, the consistency of the data is improved. Left: The normalized maximum intensity plotted as a function of the original, as-measured maximum intensity. The wide original intensity (horizontal) scatter becomes narrow with normalized (vertical) intensity around 4700 and 5000 (normalized intensity). As the wire cross-section has the same total area no matter how the wire is tilted, the variations here are artifacts and can be removed. Right: The results of renormalizing the intensity, as a function of tilt angle, showing smoother results than in Figure 3.11, top graph.

normalized HAADF intensity; the compressed vertical spread in this plot indicates that some points appear to be closer in renormalized intensity than their original. The second part of Figure 3.14 shows renormalized intensity spread as a function of tilt-angle, showing a smoother distribution than the original HAADF intensity spread, shown in the first part of Figure 3.11.

Figure 3.15 shows the quantitative effects of renormalization on intensity-thickness calibration, by comparison with Figure 3.12. The linear fit is much improved on both the compressed (Figure 3.15, top) and complete (Figure 3.15, middle) length scales than before renormalization (Figures 3.12, top and 3.12, middle, respectively), and the distribution of intensity in Figure 3.15, bottom, has much more precision than that in Figure 3.12, bottom. The precision of the intensity-thickness calibration is improved by about a factor of 4 by using renormalized intensity instead of the original HAADF intensity.

When this RENOIR intensity-thickness ratio is applied to the (renormalized) HAADF profile, the result is seen in Figure 3.16. The average maximum thickness in Figure 3.16 is 137 nm, and the 2-sigma error bars on the intensity-width determination from Figure 3.15 give a thickness range of 134-141 nm, which is vastly improved over the 117-148 nm thickness determination in Figure 3.13. This RENOIR thickness value is within the original thickness error range. In this way, RENOIR represents perhaps the easiest and most direct method for intensity-thickness calibration. However, the width determination can be further improved and verified by making width determination a systematic process.

Improving the Width Determination

Now, by improving the width determination, it is possible to improve and verify the other pillar of the RENOIR thickness determination. This is done by systematizing how the HAADF intensity threshold for the nanowire width is determined (from intensity-renormalized images).

Due to the "beam tails" on the STEM probe used to form the images, when the center of the probe is placed immediately outside the edge of the nanowire, a non-zero amount of the electron probe will pass through the near-edge region of the wire, which generates detectable HAADF intensity. This is visible on *e.g.* Figure 3.16, where the regions on the very edges of the nanowire have a slight tapering to 0 intensity.

Changing the intensity threshold T for the wire width determination can have a noticeable effect. The effect at a single orientation – the one examined previously, including in the bright-field TEM image – of choosing different threshold values for the width can be seen in Figure 3.17. While the choice of T=100 appears nominally reasonable if there is no intensity in vacuum, or no beam tails, using T=100 yields a nanowire width that is higher than the nanowire width measured using other techniques detailed in this chapter (DT or BFTEM). This may be a sign of detected intensity from the shell surrounding the wire, but the width of the shell is too low for a conclusive determination. Figure 3.17 shows that, an intensity threshold value for width determination of T=200 or T=300 would be consistent with the experimental calibrations. However, simply choosing an arbitrary value of T would not necessarily be optimal for the entire tilt series.

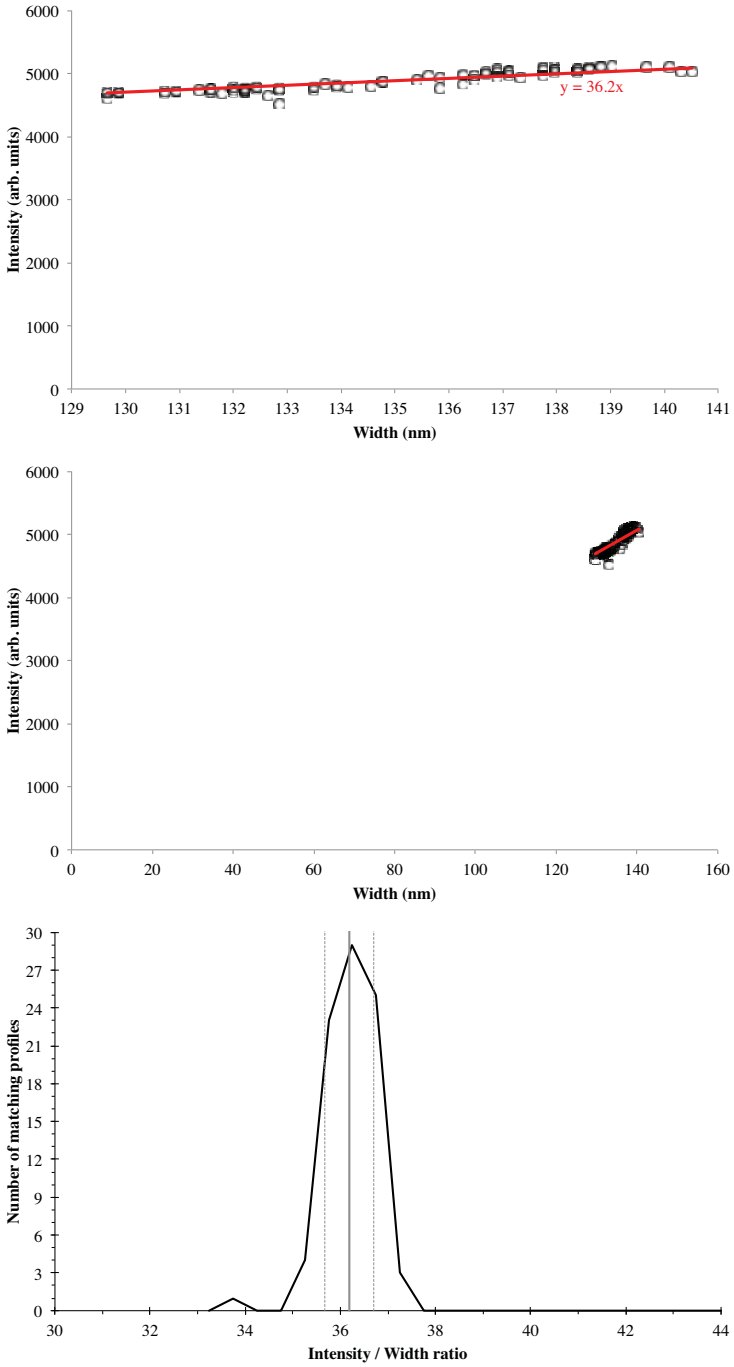


Figure 3.15: RENOIR graphs at intensity threshold *T*=100 counts. This uses the same graph legends as in Figure 3.12; see that figure for technical details. This is carried out on renormalized intensity data – making this RENOIR (RENormalized Orthogonal-Images Routine) data. This can be compared to the results in Figure 3.12, with RENOIR showing a better linear fit (top and middle), even when forced to have a zero intensity = zero thickness intercept, and much more precise intensity-width ratios (bottom). The intensity threshold used here to determine the width – *T*=100 counts – is relatively arbitrary in the presence of STEM probe tails, and improved in the following figures.

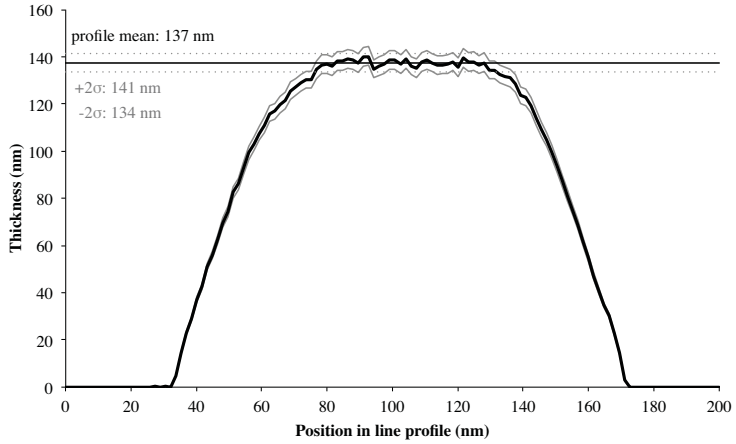


Figure 3.16: RENOIR thickness-calibrated HAADF intensity profile, with 2σ error margins indicated. This thickness profile, calibrated from the RENOIR data in Figure 3.15, is much more precise than the one in Figure 3.13, showing the value of renormalization. This, like Figure 3.15, is using the *T*=100 counts threshold. Noise is slightly different than Figure 3.13 due to the use of an intensity line profile that had been averaged over more points.

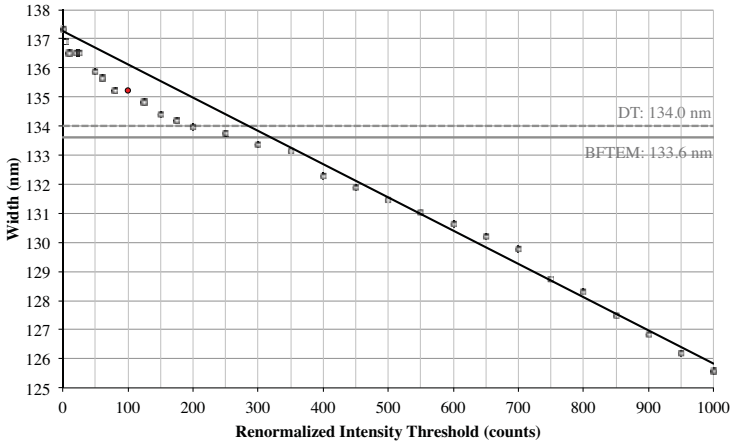


Figure 3.17: Width as a function of RENOIR intensity threshold. By changing the intensity threshold value that defines the presence/nonpresence of the nanowire, the measured width of the nanowire changes. The *T*=100 count threshold used in e.g. Figure 3.15 is noted in red; the core widths from discrete tomography (DT) and bright-field TEM (BFTEM) are plotted. A fit line between 400-1000 counts provides a guide for the linear behavior. However, below *T*=400, because of the abrupt nanowire edge, it is not intuitively obvious from the RENOIR data alone where the threshold should be chosen. However, if this threshold is *T*=200-250, then all the widths will agree.

This plot is derived from the intensity line profile across the nanowire at the holography tilt angle (+7° stage tilt). While the holograms were taken over a range of tilt angles, this was the plateau orientation near the center of the tilt series.

The optimal intensity threshold value for width determination is determined by statistical matching to the entire data set of orthogonal points. First, the threshold value is adjusted to yield a range of different RENOIR width-intensity plots. By choosing the value of T to maximize the R^2 value of the zero-intercept linear fit (and therefore optimizing the linear fit), it is possible to determine a best-fit value of T from the entire RENOIR data set without recourse to arbitrary methods or additional data. In Figure 3.18, the results of this are shown, and $T = 200$ counts is determined to be optimal. At this threshold value, the core width w is 134.0 nm, which is identical to the core width determined from discrete tomography in the previous section. Figure 3.18 also shows one line at the core width measured using the BFTEM spatial width calibration determined in this work (w = 133.6 nm), and the lower from Dr. Boothroyd's BFTEM calibration (w = 133.3 nm), showing the close agreement between different techniques. The total effects on the width of choosing a single intensity threshold can be seen in Figure 3.19, which will be returned to in the discussion on DT vs. RENOIR.

The plots of the individual RENOIR points can be seen in Figure 3.20, laid out in the same fashion as Figures 3.12 and 3.15. The intensity distribution appears to be bimodal, in comparison with Figure 3.15, but it retains the tight precision of Figure 3.15, and is visually comparable. Therefore, this appears to be perhaps the least suboptimal width determination for RENOIR.

Results of Full RENOIR Approach

With this information, and following the full RENOIR approach, the measured HAADF profile can be calibrated from intensity into thickness. This is seen in Figure 3.21, with the RENOIR and DT thicknesses co-plotted. This RENOIR profile is used as the RENOIR thickness profile of the nanowire in the subsequent chapter. While the DT and RENOIR thickness profiles are similar, they are different in measurable ways.

Why Not Only RENOIR?

RENOIR is a useful, non-tomographic technique for calibrating HAADF images, but RENOIR relies on several assumptions. It will not have tomographic reconstruction artifacts, and some of the information in it is automatically directly verifiable by reference to the original images (*e.g.* the nanowire width will be the same in the HAADF profile as in the RENOIR thickness determination). However, there are several reasons why RENOIR is not the only technique employed in this work for thickness determination.

One potential weakness of RENOIR is in determining the exact width, as mentioned above. Despite the systematic and optimized width-determination method used here, the width is still sensitive to the exact threshold chosen, and the R^2 value of the linear fit with a zero intercept (seen in Figure 3.18) is still rather less than 1, indicating that the linear fit is not perfect. This is possibly due to the intensity of the beam tails depending on the thickness of the nanowire at the edge – the "house" and "plateau" orientations have very different edge thicknesses. Therefore, while the width determination threshold used here is a globally optimized quantity, a better width determination would be useful. This improved width determination should be general enough to be one ruleset, applied across the entire series – perhaps including deconvolution of a probe shape.

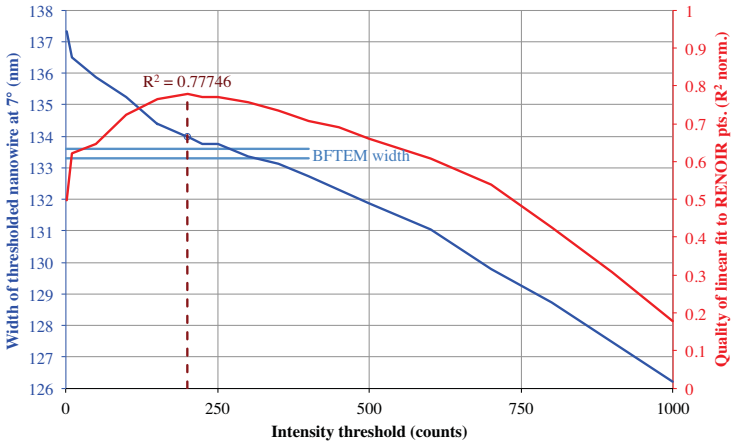


Figure 3.18: RENOIR intensity-threshold determination and quality of RENOIR intensity-threshold determination. In this figure, the optimal threshold is determined by testing the quality of the linear fit (as in the red fit in Figure 3.15) as a function of the threshold. The R^2 value of the linear fit to the ensemble of tilt angles is used to evaluate the quality of the fit, and is plotted in red. The corresponding width at that threshold is plotted in blue. The threshold with the maximal R^2 value is the threshold at which a single RENOIR intensity-width relation is most applicable to the entire tilt series. This is determined to be T=200 counts, agreeing with the width in the DT reconstruction (see Figure 3.17).

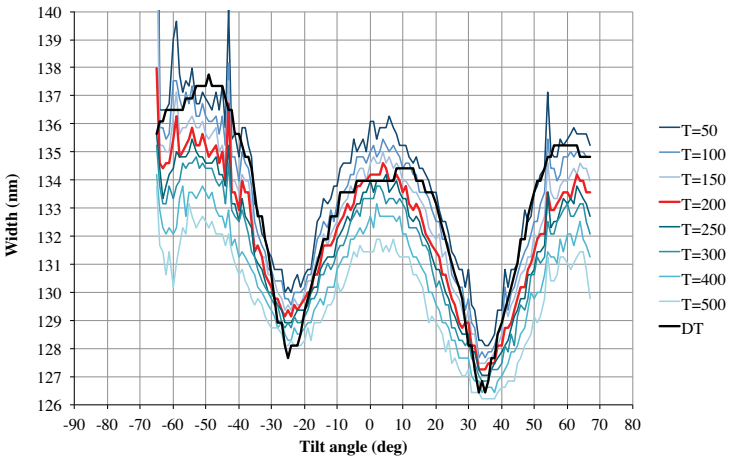


Figure 3.19: RENOIR width as a function of tilt angle at different intensity thresholds. The widths determined from discrete tomography (DT) are plotted in black, and the T=200 optimal RENOIR threshold, determined from the R^2 fit shown in Figure 3.18 is plotted in red. This shows disagreement between the DT and RENOIR data, especially at high tilt angles, but general reproducibility of similar features – note how the T=200 threshold matches the DT data near the local maximum at 0° tilt, but not near the local minima near ±30°, or the other local maxima around ±60°.

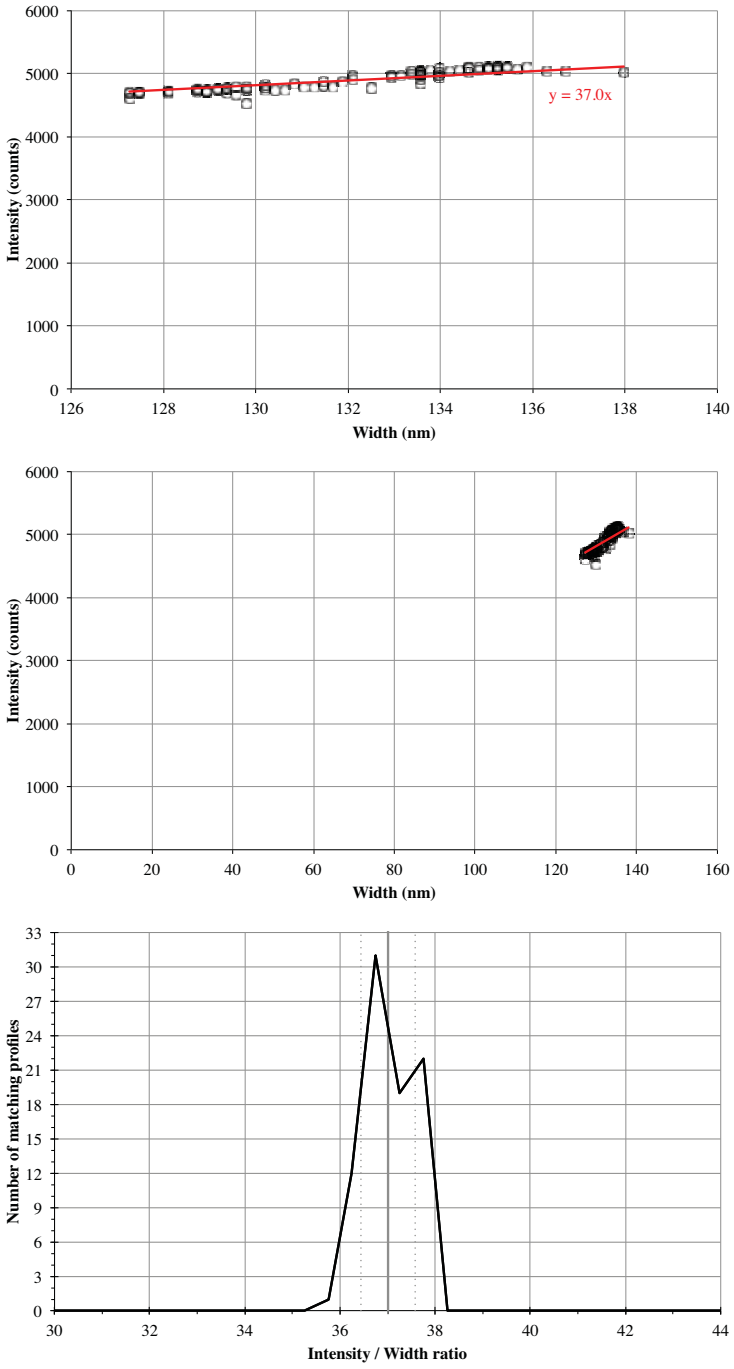


Figure 3.20: RENOIR intensity/width ratio at *T*=200 count threshold. This is a counterpart to Figure 3.15, but with a different intensity threshold for width determination. Here, the threshold determined in Figure 3.18 is applied to the data. This shows good linearity and, in the bottom graph, the average width and 2σ margins of error are indicated, comparing favorably with those in e.g. Figure 3.15.

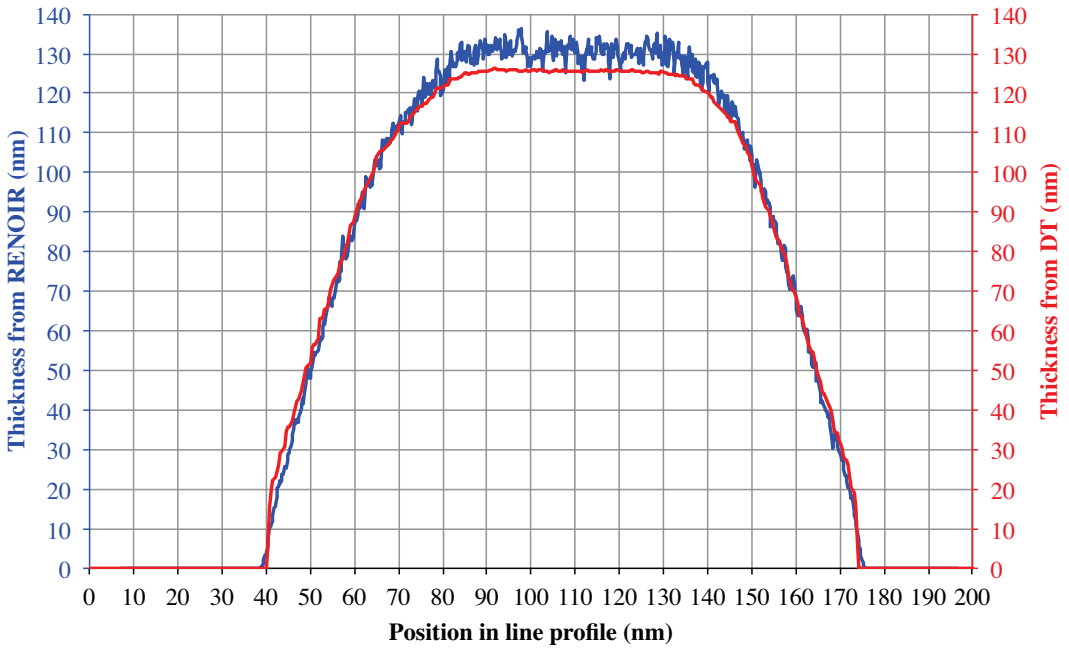


Figure 3.21: RENOIR vs. DT thickness profiles. This graph shows that, while the thickness profiles determined using RENOIR (blue) and DT (red) are similar, they are not identical. They appear to agree around 50-100 nm; DT yields higher thicknesses than RENOIR at approximately $t<\sim50$ nm, and RENOIR yields higher thicknesses than DT at approximately $t>100$ nm. The input data has a precision of 1.5%, but the mismatch at maximum thickness is ~4%, and, at lower thicknesses, the DT thicknesses are slightly overestimated relative to the RENOIR thicknesses. The "true" thickness of the specimen is not known *a priori*, but must be measured, therefore, the ultimate accuracy of these two techniques is not known – their precision is, though, and on the order of 1.5% due to magnification calibrations. The ramifications of this are examined in chapter 4.

The RENOIR approach does still rely on several assumptions. The intensity-thickness relationship in the HAADF imaging condition is assumed to be linear for the intensity in the HAADF profile to be an accurate reflection of the nanowire thickness. This assumption is examined in Chapter 4. As with all of the data used here, imprecision in the magnification calibration would lead to additional imprecision in the thickness profile determined with this method. While there is random error due to the noise on the intensity profiles, this noise is not critically significant; the "synthetic wedge" approach used in Chapter 4 reorders the data as a function of thickness, reducing the effect of the noise.

Final Thickness Determination

The thickness determination of the nanowire cross section in this chapter is used in Chapter 4 to calibrate the results from electron holography. Two thickness-determination techniques were chosen as useful: the renormalized orthogonal-images routine (RENOIR), and discrete tomography (DT). RENOIR and DT were chosen because both yield calibrated information across the entire thickness range, with no apparent "missing wedge" gaps in interpretability. However, the two techniques do not agree on the final thickness, and closer analysis, carried out in this section, shows that they have several points of disagreement, some of which are beneficial to DT, others to RENOIR. Ultimately, with the available information neither can be said to be clearly better than the other, and both the RENOIR and DT cross-sectional thickness profiles are carried forward into the next chapter, after discussion of their inaccuracy and imprecision.

Disagreement Between RENOIR and Discrete Tomography

As has been briefly discussed earlier, the results of determining the thickness profile of the nanowire using RENOIR and discrete tomography (DT) are inconsistent. This inconsistency can be seen in Figure 3.21. There are slight mismatches near the edges and on the sloping regions, but the largest disagreements come in the high-thickness region around the 70-150 nm position in the line profile in 3.21. This is not likely to be the result of nonlinearity in the RENOIR intensity profile, as that would have an intensity that is sublinear with thickness, leading to an underestimate relative to the "real" thickness. For a more important disagreement, the next section details the effect of choosing different RENOIR width thresholds.

Different RENOIR Thresholds vs. DT

As noted previously, the surprising effect on the entire tilt series of choosing different threshold values can be seen in Figure 3.19. Figure 3.19 shows the nanowire width measured for different intensity threshold values from T = 50 counts to T = 500 counts, with the optimal T = 200 counts level in red and the discrete tomography (DT) width at each orientation plotted in black. Several features are immediately apparent when the DT and RENOIR results are compared. First, the T = 200 level and the DT reconstruction agree exactly at +7°, but the match becomes less consistent with increased distance away. Some of this disagreement is expected due to the STEM probe tails, but some is not.

Around -25° and +35°, the DT width minima are near the T = 400 and T = 500 lines. This might be expected, as the probe tails would generate more intensity from the fall-off of a "house"-type orientation (*e.g.* that seen in Figure 3.4) with high thickness at the edge, as compared to a "plateau"-type orientation (*e.g.* that seen in Figure 3.1) with a low thickness at the edge.

Forming unexpected contrast, in Figure 3.19 around -50° and +60° are the other two plateaus, and both are consistent with very low thresholds; $T < 50$ for -50°, and $50 < T < 100$ for +60°, and DT yields noticeably larger widths for these orientations than RENOIR: DT widths are larger by about 2 nm. This indicates fundamental disagreement between the two techniques, and potentially the missing wedge data is playing a role in the DT reconstruction, or an additional unaddressed factor is leading to the misdetermination. The missing wedge, or some variety of intensity-thickness misdetermination in either RENOIR or DT, or a slight difference arising from the slightly differently aligned input data, are the obvious reasons why RENOIR and DT would agree on the +7° plateau while disagreeing in the same manner on the -50° and +60° plateaus.

Therefore, there appears to be a critical disparity between RENOIR and DT widths at high tilt angles that is not obviously explained by the naïve width-determination thresholding applied here. If this were evidence for a tilt-angle-dependent disparity between RENOIR and DT, it would be at its highest magnitude for the directly projected thickness of interest in this work. Therefore, the effect of the missing wedge on the DT reconstruction is unclear, and the DT reconstruction, despite being nominally easy to interpret, may have silent inaccuracies, including the intensity-thickness ratio determined in the algorithm.

The discussion in the next section focuses on imprecision and inaccuracy in the thickness-determination measurement and in the techniques used in this chapter.

Potential Sources of Imprecision and Inaccuracy

There are several potential sources of imprecision and inaccuracy in the thickness determination. The two most critical ones here are nanowire width determination, and magnification calibration.

Width Determination

Quantification in every example above relies on some form of width or intensity/thickness determination, either pre- or post-processing.

The techniques that require manual width-determination – all the ones discussed here except discrete tomography – require careful thresholding. The width determination to create consistency with the benchmark width measurement for the AT techniques must all be done post-reconstruction, meaning it will include reconstruction artifacts, and is shown above to not be suitable. The width determination for GT and RENOIR occurs before analysis, so it will not include those reconstruction artifacts. However, both are vulnerable to misdetermination of the width, and the GT used here was not suitable for thickness determination.

Both DT and RENOIR arise out of their own sets of sensible assumptions, but both use intensity-thickness calibration. DT with the DART algorithm used here is a tomographic reconstruction technique with a reasonable track record at reconstructing and determining objects, but may have open questions regarding its intensity-thickness determination – see [60] for details. RENOIR is a HAADF-intensity-calibration technique that should work as long as widths can be accurately measured and as long as the HAADF intensity is a linear function of thickness.

There is a difference in how DT and RENOIR use and determine intensity-thickness calibration. While DT uses intensity-thickness calibrations as an intermediate step in

its reconstruction technique, the RENOIR output is an intensity-thickness calibration. However, the question for RENOIR is if the width determination (through setting an intensity threshold) can be improved. Additionally, there is a difference in access. RENOIR is clearly auditable – *e.g.* Figure 3.19, displaying the effect on the width determination of different thresholds – while the author of this dissertation does not have access to the DART reconstruction program used here and thus cannot input different intensity-thickness calibrations to determine their effect.

Magnification calibration is the critical numerical factor in the accuracy and precision of this experiment. As stated earlier in this chapter, the magnification calibration precision is ±1.5%. This is considered to be the random error in the measurements, while the other factors – RENOIR width determination and DT intensity-thickness dependence, among others – are inherent to the techniques involved, and are therefore considered as part of the nominal inaccuracy inherent to the techniques. Therefore, the total precision of the two key techniques used for thickness determination in this chapter – RENOIR and DT - is reliant on the precision of their magnification calibration, and is therefore taken to be the same.

The accuracy of the two techniques is not perfect, but accuracy in this instance is more difficult to define. The accuracy is subject to technique-dependent factors that cannot be audited simply by reconstructing an unknown object, or even by examinations of how the technique would reconstruct known objects, because there may be other factors in the measurement that prove vitally important, but that are not taken into account here (*e.g.* HAADF intensity-thickness non-linearity). The proof of the accuracy or inaccuracy of these thickness-determination techniques must come from another technique – and the holography analysis carried out in Chapter 4 is an ideal candidate. Therefore, while the precision can be stated, the accuracy cannot, at least at this stage.

Conclusions: Thickness Determination Methods

Ultimately, the objective of this chapter is to quantitatively determine the cross-sectional thickness profile of the InAs nanowire used in Chapter 4 to the highest possible precision and accuracy. This thickness determination is used in the analysis of amplitude and phase images from electron holograms. This requires first the best possible magnification calibration, which is done using a bright-field TEM image to determine the core and shell components, and then the high-angle annular dark-field (HAADF) scanning TEM (STEM) images used for further examination are calibrated, and the HAADF STEM images are found to only capture the core of the wire, not the shell. A HAADF STEM tilt series of images is acquired and used analytically to determine the morphology of the wire.

Several techniques for thickness determination are used. First, three algebraic tomography (AT) algorithms are employed: weighted back-projection (WBP), the algebraic reconstruction technique (ART), and the simultaneous iterative reconstruction technique (SIRT). All three AT algorithms are deemed unacceptable for cross-sectional determination; they must be thresholded post-reconstruction. Two other tomographic reconstruction algorithms are applied to this HAADF STEM tilt series, and these algorithms make use of additional known information about the wire. The geometric tomography (GT) algorithm used here applies the principle that the object is convex and disregards the intensity (preferring a pre-input thresholding algorithm), and reconstructs the outside shape of the wire from the shadow outlines.

However, the GT reconstruction has a large missing-wedge artifact due to the incomplete nature of the angular span of images that were taken. This large missing wedge artifact renders the GT reconstruction unsuitable for further analysis. Discrete tomography (DT) assumes that the object is of constant density, and reconstructs the entire shape of the object, without a detectable missing wedge artifact. The DT algorithm used here yields an apparently sensible result, and this reconstructed cross-section is used for thickness determination

The last technique used is a non-tomographic calibration technique named in this work as RENOIR (short for RENormalized Orthogonal-Images Routine). RENOIR operates by calibrating the intensity of the original HAADF profiles by correlating the width and intensity of cross-sectional line profiles from orthogonal images in the HAADF STEM tilt series. The intensity profiles are improved by renormalizing them to a common total cross-sectional value, and the widths are chosen by optimizing the statistical coherence of the width determination across the entire set of orthogonal data points. The result from RENOIR is a direct thickness calibration of the original HAADF profile, free from any reconstruction artifacts, but still not perfect.

The final result is that two different thickness determination techniques are deemed to yield acceptable cross-sectional thickness profiles. They are RENOIR and DT, with the results shown in Figure 3.21. The DT and RENOIR cross-sectional thickness profiles are fundamentally inconsistent with each other, but each is valid in principle. Therefore, they will both be used in the next chapter to remove and compensate for the thickness contribution from the electron holograms. Magnification calibration provides the primary random error of ±1.5%, common to both RENOIR and DT, but the accuracy and systematic errors in RENOIR and DT require comparison with an additional dataset.

The holography data in the following chapter, Chapter 4, uses the RENOIR and DT thickness determinations of the InAs nanowire developed in this chapter to determine and examine physical quantities of the same InAs nanowire, specifically, the mean inner potential, mean free path, and absorption parameter.

Chapter 4: Tilt-Controlled Electron Holography

Outline

In the previous chapter, two techniques – RENOIR and DT – provide thickness determination of the cross-section of an InAs nanowire. In this chapter, this thickness determination is applied to off-axis electron holograms of the same wire to yield mean inner potentials from the phase, and mean free paths from the amplitude. By systematically examining a range of orientations of this nanowire, the effects of dynamical diffraction on off-axis electron holograms are measured and compared with simulations.

In the following sections, electron holography is used to quantitatively analyze the phase and amplitude of the electron beam. The backbone of this analysis is an extensive self-calibrating experimental dataset of off-axis electron holograms taken at tilt-controlled specimen orientations. The specimen orientation is determined throughout the high-angular-resolution tilt series of holograms. Then, the phase and amplitude images are extracted from the holograms, and cross-sections of the wire are taken. Using the thickness determination from Chapter 3, two regions of the nanowire cross-section are examined: the averaged maximum-thickness region, and "synthetic-wedges" calculated from the sloping sides. The shell layer discussed in Chapter 3 is compensated for, and both mean inner potentials and mean free paths are calculated from the synthetic-wedges from weakly diffracting orientations from the tilt series. The phases and amplitudes are then examined in both strongly and weakly diffracting conditions, and the diffraction effects are compared with simulations.

Experimental Data Input and Treatment

The experimental data in this section were acquired on the ATEM, mentioned in Chapter 2 as a FEI Titan 80-300 TEM at the Technical University of Denmark, installed at the Center for Electron Nanoscopy (DTU Cen). For this experiment, the microscope was operated at a primary accelerating voltage of 120 kV; the lower accelerating voltage was chosen to enhance the signal-to-noise ratio in the recorded holograms.

The specimen was the same InAs nanowire detailed in Chapter 3. Both holograms and diffraction patterns of the specimen were acquired and analyzed; all images were taken on a 2048×2048 pixel CCD camera.

The holograms were acquired in the Lorentz TEM mode of operation. As detailed in Chapter 2, this mode does not use the conventional objective lens, allowing for medium-resolution electron holography. A tilt series of holograms of the InAs nanowire were acquired every 0.1° from 2° to 13° relative to the horizontal.[22] As discussed in Chapter 2, reference holograms were taken for each specimen hologram, and 64s exposure times were taken to maximize the signal-to-noise ratio - [109] applies similarly long acquisition times for the same reason. After acquisition, the holograms were processed in a standard manner, as detailed in Chapter 2, using the ASUHolography software package. The specimen and reference holograms, when

[22] The center of the symmetric pattern in the tilt series of holograms is seen at approximately 7° to the horizontal. Therefore, there are two degree ranges for the data discussed here: [2,13]° and [-5,6]°. The former is used when the stage goniometer reading is required; the latter when the orientation of the nanowire specimen is of interest.

processed, yielded phase and amplitude images. These can be seen for different tilt angles in Figure 4.1.

However, the phase unwrapping routines included in the ASUHolography package, while optimized for electron holography, did not yield faithful unwrapped phases due to the rapid thickness changes in the nanowire. Therefore, phase unwrapping was done manually on single-pixel line profiles[23] across the wire, which allowed the original holograms to be used as a reference to determine the correct number of phase wraps that needed to be unwrapped. This phase unwrapping strategy is unavailable to the automatic routine. Figure 4.2 shows one example of this manual unwrapping, where the original line profile, the phase unwrapping, and the final phase profile are seen. Phase and amplitude line profiles were taken from the same region of the wire, and amplitude was converted into amplitude-derived t/λ using the equation given in Chapter 2 ($-2\times\ln(A/A_0)=t/\lambda$). The phase and amplitude line profiles were then processed individually; however, because the phase and amplitude are derived from the same electron holograms; the same steps were carried out on both phase and amplitude line profiles for removing aliased data points and aligning the profiles with each other.

TEM diffraction patterns were acquired in both normal-TEM-mode (defined as with the conventional objective lens on) and Lorentz mode. These allow for determination of the specimen orientation. Analysis of the diffraction patterns post-acquisition to determine the orientation was done by comparing with routines from the commercially available JEMS software package [110].

Orientation Determination

The orientation of the InAs nanowire was determined by analysis of diffraction patterns. Normal-TEM-mode diffraction patterns, when compared with simulated diffraction patterns from JEMS [110], indicated that the experimental data matched orientations near the $[10\bar{1}0]$ wurtzite zone axis. By tilting the InAs nanowire specimen and acquiring diffraction patterns, the axis about which the nanowire was tilted was also determined. However, due to an additional beam tilt between normal-TEM-mode and Lorentz mode, a Lorentz mode diffraction pattern was used to determine the orientations at which imaging took place, and the tilt axis in normal-TEM-mode was applied to these orientations. The tilt axis of the specimen is the same in both Lorentz and normal-TEM-modes because the specimen tilt is a property of the holder and specimen, not the electron beam.

Figure 4.3 shows the orientation determination step-by-step. Figures 4.3a and 4.3b are excerpts from normal-TEM-mode diffraction patterns used to determine the specimen orientation, the nearest low-order crystal zone axis, and the specimen tilt axis, taken

[23] Linescans on the phase image wider than a single pixel involve averaging, which, if done on the phase images, can remove the sharp jump in phase that is a signature of a phase wrap. For example, if a pixel in the noisy wrapped phase image, where all values lie on the interval $[0,2\pi)$, with a value of 6 radians is naïvely averaged with an adjacent pixel that has a value of 1 radian, the result will be 3.5 radians – despite the high likelihood, given their physical proximity, that the two pixels lie on either side of a phase wrap, making the real phase average of these two 6.64 radians. In this way, averaging pixels over the wrapped phase image masks the true value of the phase and phase wraps, and loses the value of the phase wrap. This is wildly suboptimal for this work. One way of averaging would be to average on the complex images generated by electron holography itself (real and imaginary components, instead of amplitude and phase). This was not done in this dissertation, and would be useful in future work.

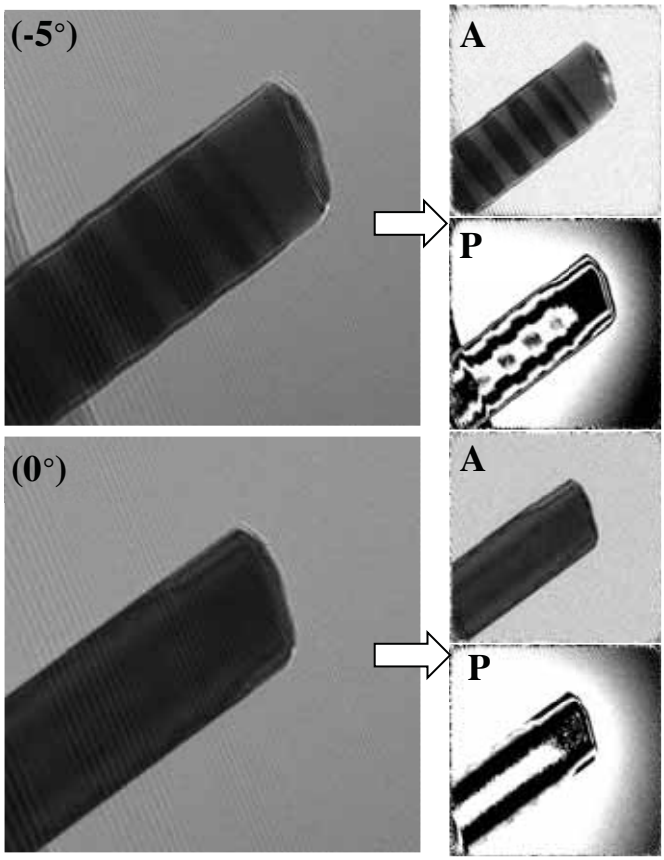


Figure 4.1: Holograms and reconstructions of nanowire. The holograms shown on the left are typical examples from the tilt series of holograms, with their reconstructed non-unwrapped phase (P) and amplitude (A) images shown to the right. This depicts the same wire and region as in Figure 3.1. Diffraction contrast in the images and charging in the phase images are discussed in the text. The phase images show a non-flat vacuum, which is evidence of charging.

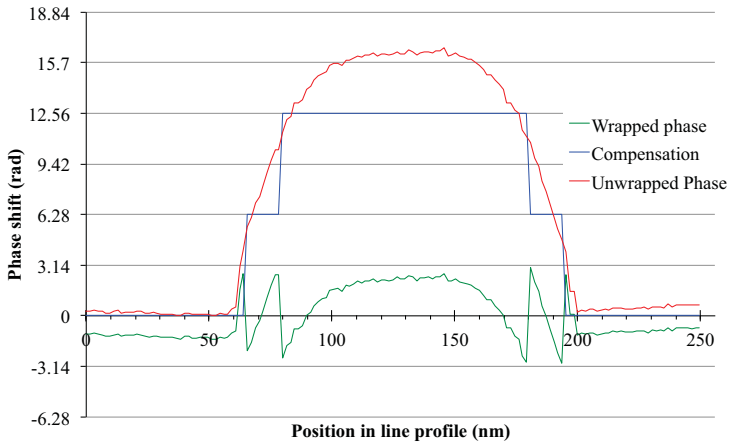


Figure 4.2: Phase profile and unwrapping. Green line shows the original wrapped phase, blue line the manual compensation (multiples of 2π), and the red line the sum of the two. This graph shows the effect of phase unwrapping on a single phase line profile, as well as resetting the 0 of the phase to remove negative phase shifts. This profile is from the red set of profiles in Figure 3.1, at +2° tilt as plotted in Figure 4.4 (9° stage tilt). Note the slope in vacuum, at positions >200nm or <60 nm, also present in the image in Figure 4.1.

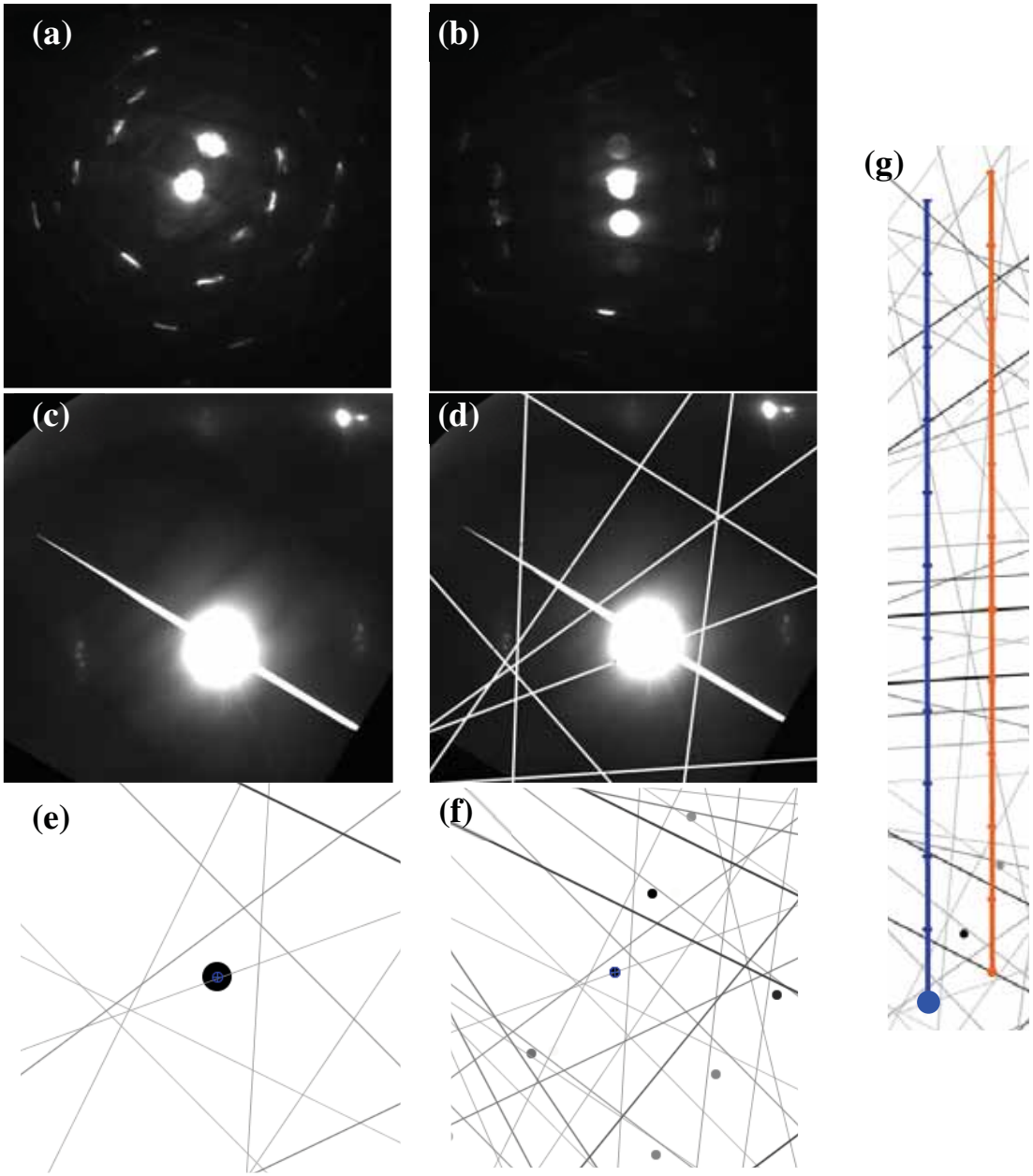


Figure 4.3: Determination of nanowire orientation for holography. (a) and (b) are conventional-TEM-mode (non-Lorentz) diffraction patterns used to determine the crystal zone axis and the specimen tilt axis, taken 5° of stage tilt apart. (c) is a Lorentz-mode diffraction pattern taken to determine the relative beam tilt between conventional-mode and Lorentz-mode; this provides the orientation of the experimental hologram, which differs from the conventional-mode orientation due to a beam tilt, but the direction of the stage tilt axis in reciprocal space is the same. By determining the presence and location of Kikuchi features, shown in (d), the quantitative tilt position is determined, as seen in (e) and (f). By combining this information, the tilt axis and the positions of the holograms can be determined; (g) shows the tilt axis in orange as determined in normal-mode, and the corresponding tilt axis in blue for Lorentz mode, which was used for holography, with tick marks every 1°. The blue line in (g) has a dot at the orientation used as -5° in this chapter (2° stage tilt), where 0° is the tickmark between the strong Kikuchi bands.

from two specimen orientations separated by 5°. Figure 4.3c is a Lorentz mode diffraction pattern, taken to determine the relative beam tilt between normal-TEM-mode and Lorentz mode; this provides the orientation of the experimental hologram. The specimen orientation is precisely determined from identifying Kikuchi features [1], highlighted in Figure 4.3d. These match to the JEMS simulations in Figures 4.3e and 4.3f. By combining this information, the tilt span and the positions of the holograms can be determined; Figure 4.3g shows the tilt span in orange for normal-TEM-mode, with tick marks every 1° and round markers where the experimental diffraction patterns match, and the corresponding tilt span in blue for Lorentz mode, also with tick marks every 1° and a round marker for the location where the experimental pattern was taken from. The tilt span in Lorentz mode matches (-110.6, 111.8) to (86.5, 98.5) mrads. With this information, the tilt series can be quantitatively analyzed.

An Overview of the Holography Tilt Series

There are many different segments to the analysis of the holography data in this work. The data were processed as outlined in Chapter 2 and as discussed above, using the ASUHolography software plugins for Gatan DigitalMicrograph.

The tilt series of holograms is now considered, and the presence of diffraction effects is immediately obvious, but the tilt series itself requires some interpretation. Figure 4.4 shows the results of the tilt series as a whole, and the effect of tilting the nanowire specimen. Figure 4.4 contains phase and amplitude images generated from 1-pixel line profiles, colored blue (a), green (b), and red (c) to match Figure 3.1. Figure 3.1 shows where the line profiles were acquired from, and this blue/green/red color scheme is consistent across all red, green, and blue line profiles and graphs in this chapter. These are not as-acquired images: each vertical 1-pixel row is a *spatial* line profile across the cross-section of the nanowire in a single image, while each horizontal 1-pixel row is an *angular* line profile of the same location in the nanowire cross-section across the tilt series. The graph at the bottom of Figure 4.4 shows the effect of the tilt angle on the average phase and amplitude of the plateau of the wire (uncorrected for charging effects). Each point in the graph in Figure 4.4 corresponds to an average over a 30-vertical-pixel, 1-horizontal-pixel section out of the top and middle images in Figure 4.4, and the *x*-axis of the graph corresponds to the horizontal axis in the top and middle images in this figure.

Two types of orientations can be seen in Figure 4.4: Strongly diffracting orientations and weakly diffracting orientations. Strongly diffracting orientations have phase and amplitude that varies rapidly with tilt angle (*e.g.* between 0° and 1° in Figure 4.4), so the phase and amplitude are very sensitive to tilt angle. Weakly diffracting orientations have phase and amplitude that varies more slowly with tilt angle (*e.g.* between 2° and 3°, or -1° to -2°, in Figure 4.4), so these can be averaged over to improve signal-to-noise ratios.

The effect of the tilt angle on the phase and amplitude of the wire is dramatic in the strongly diffracting conditions. In the graph in Figure 4.4, roughly symmetrical peaks and valleys can be seen, with a center feature and two side features, each approximately the same distance from the center feature. This is not a thickness effect, because that would not show abrupt changes in phase for this wire. This is a diffraction-contrast effect, as reinforced by the location of major Kikuchi bands in

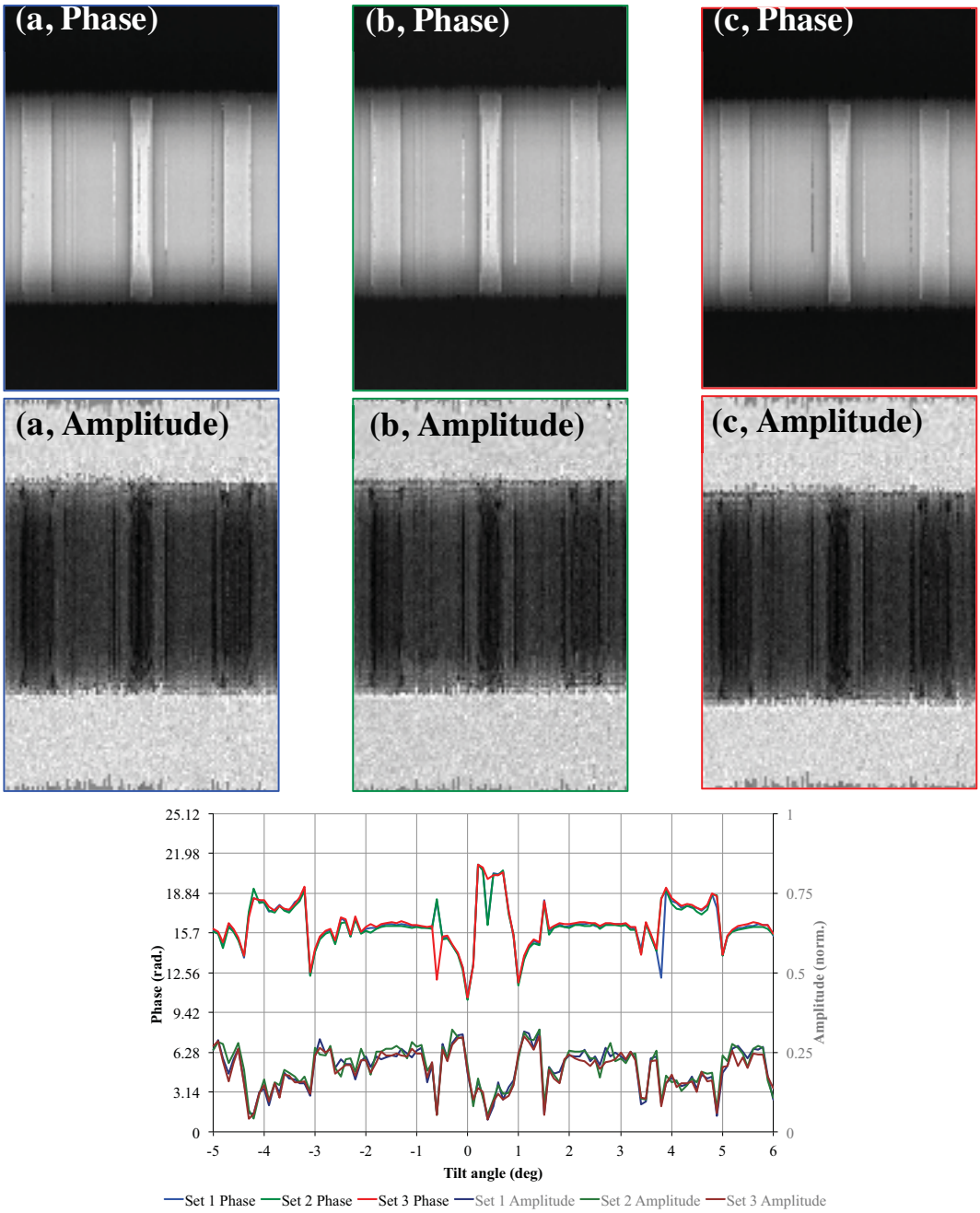


Figure 4.4: Phase and amplitude line profiles as a function of specimen orientation. The phase (top) and amplitude (middle), from individual 1-pixel line profiles across the full tilt series. In each image, each 1-pixel vertical column is a single line profile from the phase and amplitude, and each 1-pixel horizontal row is a single position as a function of tilt angle in 0.1° across the [-5,6]° tilt series. Bottom: Phase and amplitude averaged over the thickest, middle part of the wire. All line profiles are taken from the positions marked in Figure 3.1, and plotted with the same coloring. By averaging over the 30-pixel region of maximum thickness to increase sampling, a correlation can be seen between strongly diffracting conditions and amplitude attenuation. Orientations are as marked with the blue line in Figure 4.3g, with the -5° orientation corresponding to the blue circle in 4.3g.

Figure 4.3g along orientations where the effects occur, showing strongly diffracting conditions. Therefore, to understand this, comparison with simulations is necessary.

Weakly diffracting orientations are also visible in Figure 4.4. The nanowire cross-sections from the line profiles across the angular span between 2.1° and 3° are averaged and shown in Figure 4.6. These cross-sections appear very similar to the nanowire cross-section obtained using high-angle annular dark field in the previous chapter. Therefore, despite the necessity of comparing the results from strongly diffracting conditions with simulations, some results can be obtained from the weakly diffracting condition without relying on simulations. Weakly diffracting conditions are where the analysis of the tilt-controlled holography data starts.

However, there is a non-zero slope in the phase images in vacuum, which is a characteristic signal of charging. This is visible in both the line profile in Figure 4.2 and in the images in Figure 4.1, and must be compensated for before further analysis. Now, two simple mathematical models are applied (their derivation is found in Appendix A).

Charging

The work in this section is reliant on a theoretical framework provided by M. Beleggia and R. Dunin-Borkowski, who are both kindly thanked.

There is a non-zero slope in vacuum in the phase images (Figure 4.1) and phase line profiles (Figure 4.2 is representative). This non-zero slope is a product of charge accumulation on the nanowire under electron irradiation. This charge accumulation will influence measurements of the mean inner potential, because the accumulated charge affects the measured electric potential inside the nanowire. It is possible to quantify and model this charging behavior, using the measured slope in vacuum in *e.g.* Figure 4.2 [111] [112]. Using the vacuum region is good, because the vacuum region will experience no attenuation due to the material, so the fringe contrast in the holograms will be maximized and the phase measurements will be minimally noisy.

The slope seen in Figure 4.2 is sloping up away from the wire. This indicates that the wire is accumulating negative charge; this is seen in all images taken. Once the quantity of charge is determined, the distribution of charge inside the wire must be considered. Two simple models are posited here: the charge is uniformly distributed throughout either the volume ("volume charge") or the surface of the wire ("surface charge"). Appendix A contains the exact calculation for each of these two models.[24] While the details of the charge distribution are not measurable outside the wire, they do change the electric potential inside the wire. However, for both surface and volume charge models, the total charge is the same – only the spatial distribution of the charge in the cross-section of the wire changes. The width of the wire used in the charge calculation is 133.6 nm, corresponding to the width of the wire measured in the Chapter 3.

The results of this charge compensation for each line profile and tilt angle are plotted in Figure 4.5, which shows that the linear charge density is on the order of 4×10^5 electron charges per cm. The charge density shows no broad angularly dependent trend across the series, indicating that the nanowire charging was not significantly

[24] While the calculation of charging effects is relatively straightforward, it is not especially innovative, and lengthy enough to qualify as a diversion from the point of Chapter 4.

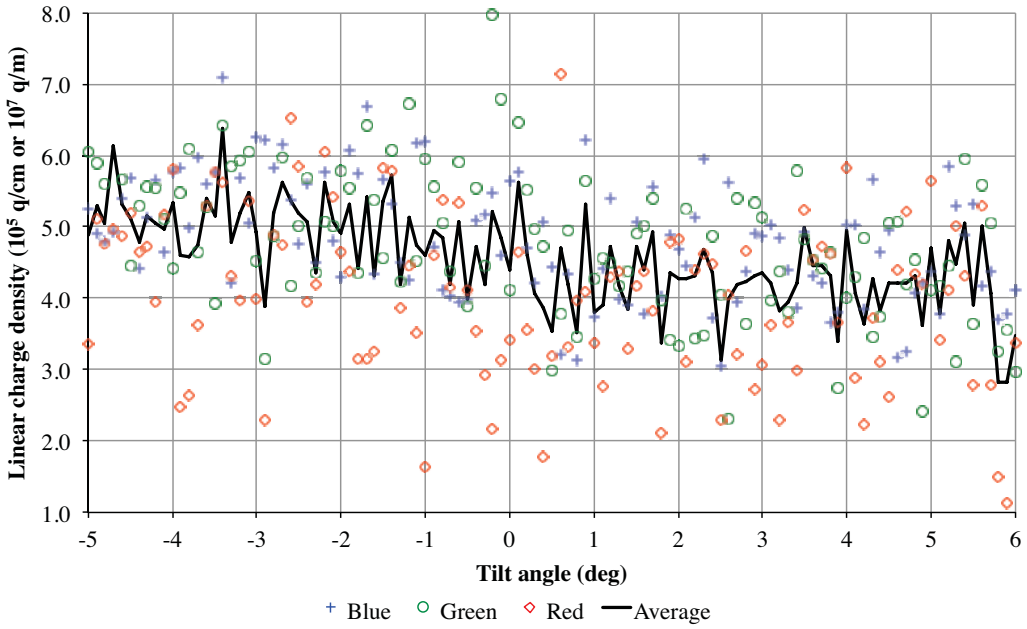


Figure 4.5: Charge compensation for phase images. The charge on the wire, calculated from the phase slopes in vacuum. The calibration was carried out for each profile individually.

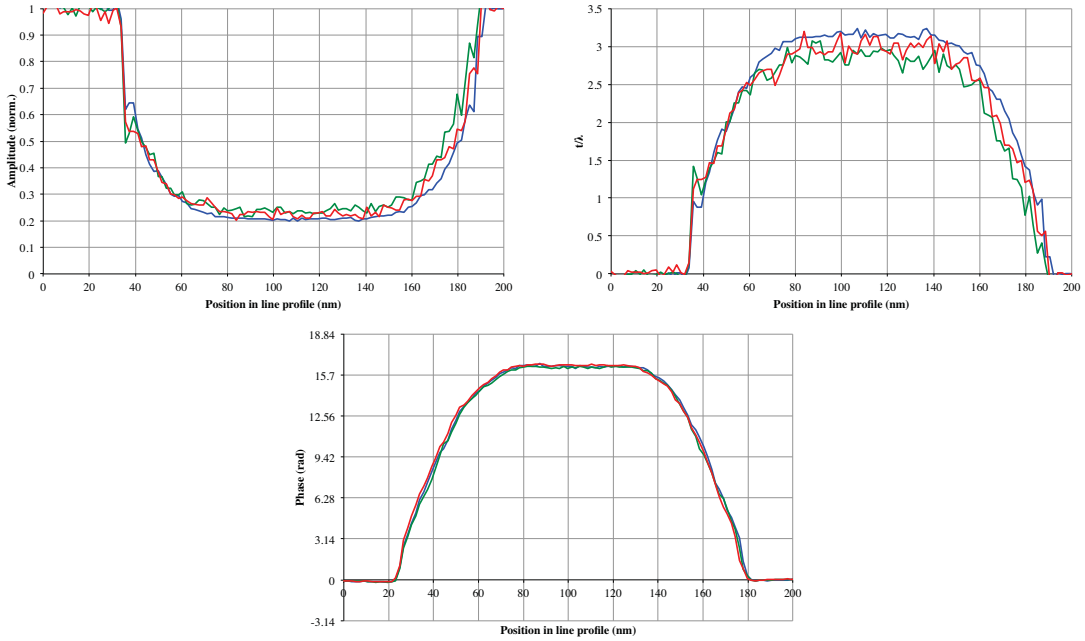


Figure 4.6: Cross-sectional line profiles from holography data. Averaged over the weakly diffracting orientations in the interval [2.1-3.0]°, colored line profiles from colored locations in Figure 3.1 and aligned horizontally for display. With the line profiles from the amplitude (top-left), amplitude-derived t/λ (top-right), and phase (bottom), there are some differences despite the line profiles coming from the same region of the wire. Also, note the effects of charging in the phase.

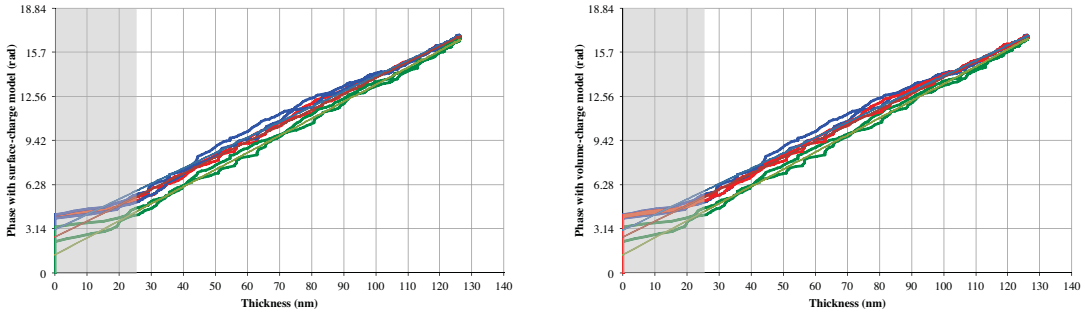


Figure 4.7: Phase versus thickness, weakly diffracting conditions, without linearity correction, using DT thickness data. Grey shows where the abrupt specimen edge means no thickness is present experimentally, leaving spurious data. The same points are connected in Figures 4.6 and 4.7, replotted here as a function of thickness. The phase appears nonlinear at $t >$ ~95 nm, and has a non-zero intercept.

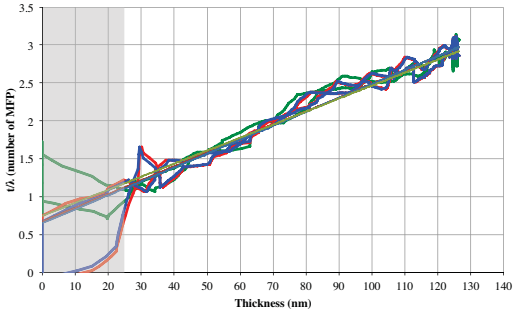


Figure 4.8: t/λ versus thickness. As Figure 4.7, but amplitude-derived data does not have charge-correction, and the same nonlinearity and intercepts are visible.

changing during tilt-series acquisition and was not orientation-dependent. Once the charge compensations are determined, they are applied to each profile from the tilt-controlled series, before other analysis is carried out. While the charge compensation seen in Figure 4.5 is noisy, an average value of the charge cannot be used to compensate. The charge compensation ensures that the vacuum in each *individual* profile is flat and 0, for which an average compensation is insufficient.

The effects of charge compensation are discussed later in this chapter, during mean inner potential determination.

Weakly Diffracting Conditions

With charging compensated for, weakly diffracting specimen orientations are the start of the tilt-controlled analysis. Weakly diffracting conditions are the natural starting point and benchmark for this analysis, because, in principle, diffraction effects are minimized, yielding benchmark measurements for the mean inner potential in the experiment.

The advantage of using an average over several measurements is that it improves the signal-to-noise ratio, decreasing the error in the experiment. The span of the tilt range chosen as "weakly diffracting" is the span [2.1,3.0]° relative to the near-center of the pattern, comprising images from 10 different, consecutive orientations. From Figure 4.4, there appears to be minimal dependence on tilt angle in the phase or amplitude along this range.

Why Is This A Weakly Diffracting Condition?

The criterion for *weakly diffracting condition* in this experiment is *minimal systematic deviation between images taken at adjacent tilt angles, and minimal presence of Kikuchi bands and other rapidly varying reciprocal-space elements*.[25]

This criterion is also the practical criterion generally accepted to avoid diffraction contrast effects for mean inner potential measurements, or measurements that depend on the mean inner potential. The traditional method, used routinely for the mean inner potential measurements detailed in Chapter 2, is to align the specimen near a major zone-axis, then tilt away from the zone-axis a few degrees; this ideally keeps the features (*e.g.* an interface) reasonably well-aligned, but is considered to negate diffraction effects. This nanowire examined here does not have a feature to be aligned, so this is not a concern. The weakly diffracting orientation used here does fit this criterion.

Therefore, this is a valid and generally acceptable specimen orientation to use for mean inner potential, absorption parameter, and mean free path measurements. Now, two different methods will be considered for this weakly diffracting data to obtain these parameters: the maximum-thickness method, which uses the data from the maximum-thickness region of the nanowire, and the synthetic-wedge method, which uses the slope from the sides of the nanowire.

Two Methods: Maximum-Thickness and Synthetic-Wedge

This weakly diffracting amplitude and phase information is analyzed in two ways: by averaging over the single maximum-thickness "plateau" region of the nanowire, and

[25] There will be a small tilt-dependent thickness effect, but it is minimal on this small 1° range.

by using the two regions of the nanowire with changing thickness to generate a "synthetic-wedge". This section provides an overview both of these methods.

The maximum-thickness method uses the average value of the amplitude or phase over the plateau-like central region of the nanowire. The primary advantage of the maximum-thickness method is the average of many points will be much more precise than any one single point, yielding a single scalar output at high precision. However, this amplitude or phase average will include the surface layer, and the surface layer likely has different properties from the core. Therefore, one number (*e.g.* phase averaged over the maximum-thickness region) is a combination of two properties (*e.g.* the nanowire core mean inner potential and the shell mean inner potential), and so cannot be used to determine the properties of the core and the shell. The maximum-thickness method is applied in this chapter to examine how the amplitude- and phase-derived quantities for the nanowire core and the shell depend on each other. However, the maximum-thickness method is used only after the synthetic-wedge method is applied.

The synthetic-wedge method uses the changing-thickness regions of the nanowire to determine the mean inner potential and mean free path, and can compensate for the nanowire shell. The synthetic-wedge method sorts the measured phase and amplitude data by nanowire thickness, and uses the change in phase (or amplitude-derived t/λ) as a function of change in nanowire thickness to determine the mean inner potential (or mean free path). The shell around the nanowire can also be compensated for in the synthetic-wedge method, and the mean inner potential and mean free path can be determined for the nanowire core and shell separately. Therefore, the synthetic-wedge method is examined first.

Synthetic-Wedge Approach to Weakly Diffracting Conditions

In this section, the synthetic-wedge approach is applied to weakly diffracting conditions. The line profiles for the weakly diffracting conditions used here can be seen in Figure 4.6 for the amplitude, amplitude-derived t/λ, and phase. This holography-derived data can be reshaped by applying the thickness determined using discrete tomography and RENOIR and sorting the data by thickness to create a synthetic-wedge. The surface layer shell surrounding the nanowire can be compensated for by using the discrete tomography reconstruction. The nanowire core mean inner potential can be determined from the slope of this shell-compensated phase, and the nanowire shell mean inner potential from the magnitude of the shell compensation. The same technique is applied to amplitude-derived t/λ to determine mean free paths instead of mean inner potentials.

Creating the Wedge

Several general points about data analysis are perhaps relevant. The holography data have been interpolated to improve the quality of the fit between the holography data and the thickness data. To fit the holography and thickness data to each other, the phase and amplitude line profiles are interpolated from their spatial calibration[26] to the 0.21 nm/pixel calibration used for the original HAADF STEM data. Interpolation is

[26] The spatial calibration of the line profiles depends on whether duplicated values arising from aliasing effects in the one-pixel-wide line profiles are removed. If these are removed, then the correct spatial calibration is 1.78 nm/pixel; if these are not removed, then the correct spatial calibration is 1.645 nm/pixel. In the former case, the nanowire is 80 pixels wide; in the latter, it is 86 pixels wide.

done solely to improve the precision of the alignment between the holography and thickness data.

The holography line profiles across the nanowire (the phase data and the amplitude-derived t/λ data, seen in Figure 4.6) and the thickness line profiles across the nanowire (seen in Figure 3.21) are aligned with respect to each other by using the two sloping sides of the nanowire. To understand how this is done, consider the effect of a misalignment between the holography data and the thickness data. If the holography data is *e.g.* shifted slightly to the left relative to the thickness data, then the left side of the wire will show a consistent thickness underestimate, and the right side of the wire will show a corresponding thickness overestimate. With this misalignment, a plot of phase divided by specimen thickness, plotted as a function of specimen thickness for the entire line profile across the nanowire, will show two different lines, corresponding to the thickness-underestimated left side of the nanowire and the thickness-overestimated right side of the nanowire. However, if the holography data were in exact registry with the thickness data, then the left and right sides of the nanowire would have the same phase-thickness relation. Thus, the holography and thickness line profiles minimizing the difference between the linear fits on the left side and the right side, the optimal correlation between the holography and thickness information is determined.

The result of the alignment process between the holography data and the thickness data is seen in Figure 4.7 for the phase data, and 4.8 for the amplitude-derived t/λ data. These are scientifically analyzed in the following section; this section evaluates the quality of their alignment. The connected lines on each graph in Figure 4.7 and Figure 4.8 trace out adjacent points from the original line profiles across the nanowire (seen in Figure 4.6). Figure 4.7 and Figure 4.8 shows a slight mismatch between points at similar thicknesses on the left and right sides of the original line profile. This is due to mostly the effects of noise, but perhaps also slight residual spatial calibration imprecision. However, the spatial calibration used is verified against measurements taken using BFTEM and HAADF STEM. For the linear fits seen in Figure 4.7 and Figure 4.8, the deviations from linearity are generally positive for one side of the wire and negative for the other, leading to an average that continues to lie in the middle.

Due to the truncated-corner nature of the hexagonal cross-sections, there is also no interpretable data for nanowire thicknesses below 25 nm. As can be seen in the reconstruction in Figure 3.9 and the thickness-determination line profiles in Figure 3.21, the truncated corners ensure there is no thickness data in the original on both sides between 0 and 25 nm. Thus, the holography data is not accurate below approximately 25 nm, indicated by the grey-shaded area in Figure 4.7 and Figure 4.8, and regions below 35 nm may not have good thickness-determination, either, due to the slight miscalibrations or misalignments mentioned in the previous paragraph. Therefore, all of the analysis in this chapter focuses on the results from the regions of the nanowire that are at least ≥35 nm thick.

However, the non-linearity and non-zero *y*-axis intercepts shown in Figure 4.7 and Figure 4.8 indicate that the surface layer must be considered. Neglecting diffraction contrast, the phase and the amplitude-derived t/λ of this nanowire are linear functions of thickness. However, the linear fits seen in Figure 4.7 and Figure 4.8 show significant deviations from linearity at higher thicknesses for thickness data, and also show non-zero *y*-axis intercepts. The deviation from linearity indicates that the

specimen thickness for holography is not completely modeled by the thickness-determination approach in Chapter 3, and the non-zero *y*-axis intercept indicates that the thickness is systematically underestimated across the entire cross-section. Both of these effects are shows in the following sections to be due to the shell layer surrounding the nanowire, which is not captured in the HAADF-derived thickness determination. While another possible explanation might be sublinear image intensity as a function of thickness in HAADF imaging of thicker materials, this concept is examined in Appendix C. However, compensating for the shell layer surrounding the nanowire improves the linear fit of the weakly diffracting condition holography data substantially.

Phase and Amplitude-Derived t/λ Without Shell-Layer Correction

The synthetic-wedge data for the phase, including linear fits, are seen in Figure 4.7, with both surface- and volume-charge models (see Appendix A) applied. The synthetic-wedge data, including linear fits, for the amplitude-derived t/λ data are shown in Figure 4.8. Now, the differences between these individual line profiles are briefly discussed. Each line profile appears to yield a slightly different linear fit in Figure 4.7. One possible explanation might be because the line profiles were taken from slightly different regions, as seen in Figure 3.1. While the blue, green, and red line profiles were intended to sample regions with the same thickness profile[27], the green profile may be too close to the tip of the nanowire; *ergo*, the green profile may have a different thickness cross-section, and this may explain the green profile's decreased phase in Figure 4.6 in the maximum-thickness region. These effects serve as a demonstration of how important thickness determination is for this experiment – even two regions that appear rather similar in cross-sectional profile can yield quite different results.

The phase synthetic-wedge graphs in Figure 4.7 include linear fits that evince the linearity problem mentioned in the previous section. Figure 4.7 shows the phase as a function of thickness, using the DT thickness data. Both volume- and surface-charge effects are compensated for (in Figure 4.7, right and left graphs, respectively) but the differences between the two charging models appear minor compared to the differences between the line profiles. The <25 nm thicknesses are unreliable and marked as such with a grey box, but the linear fits established by the >25 nm region are extrapolated down to 0 nm thickness, yielding non-zero *y*-axis intercept phases. Non-zero *y*-axis intercept phases in the phase-versus-thickness graphs indicate that the specimen thickness is systematically underestimated for the whole phase line profile. As discussed in the previous section, this non-zero intercept is the result of the HAADF-derived thickness determinations only accounting for the core, not the shell, but the phase (and amplitude) measured using electron holography includes both the core and the shell. The shell compensation is discussed in the next section.

In comparison with the phase graphs in Figure 4.7, the amplitude-derived t/λ graphs in Figure 4.8 show the same trends. The amplitude-derived t/λ graphs in Figure 4.8 have noisier data, but the same fundamental problem applies: the data appear to not be linear above approximately 95nm, which indicates that the thickness data is not a complete model of specimen thickness, and shows a non-zero *y*-axis intercept, which

[27] The region between the red/blue and green profiles has a crystallographic defect, visible in Figure 3.4, so was not analyzable in this context. It is considered in the spatially-averaged, orientation-averaged t/λ plots from weakly diffracting conditions considered later in this chapter because the defect was not visible in the amplitude images over the span.

indicates that the specimen thickness is consistently underestimated. Two potential sources of this nonlinearity are nonlinear imaging methods or surface layer contributions. The effect of non-linear HAADF intensity is discussed in Appendix C, but this model is not as good as the surface-layer model. The next section discusses surface layer contributions, which are found to model the effects very well.

Direct Surface-Layer Compensation

Directly compensating for the shell layers around the core of the nanowire is possible by calculating the thickness of the shell at each point in the cross-section from the discrete-tomography (DT) reconstruction. This is done by scaling the DT reconstruction by (133.6+8.4)/133.6, as the diameter of the core is 133.6 nm, and the shell diameter is larger by 8.4 nm. Then, the original reconstruction is subtracted, and the shell remains. Figure 4.9 shows the shell in grey and the InAs core in white. The grey part of Figure 4.9 is forward-projected, and the result is quantified into shell thickness across the line profile. An overlay of the core and shell line profiles is in Figure 4.10a. This shows a high projected shell thickness near the edges, and a constant projected shell thickness in the plateau region.

A co-plot of the projected shell thickness as a function of projected core thickness reveals the way forward. This plot, seen in Figure 4.10b, shows a relatively smooth decrease in projected shell thickness as a function of projected core thickness for core thicknesses >95 nm. If the nanowire facets were clearly defined, then the transition between different projected shell thicknesses would be abrupt, but the continual decrease is a symptom of the rounded subfacets, seen in reconstruction in Figure 3.9, where the corners of the hexagonal cross-section would be. As noted in Chapter 3, the presence of rounded nanowire corners is also confirmed from the original HAADF intensity cross-sectional line profiles seen in Figure 3.1. Therefore, the plot in Figure 4.10b is a good plot of shell thickness as a function of core thickness.

The amplitude-derived t/λ data in Figure 4.8 has thickness nonlinearity, but this single-pixel-line-profile dataset is noisy. To provide the best possible amplitude-derived t/λ data to test the shell compensation, spatially averaged line profiles were used. The amplitude-derived t/λ line profiles seen in Figure 4.6 are rather noisy when compared to the phase line profiles in Figure 4.6. Because there are no wrapping effects in the amplitude, spatial averages can be taken. Spatially averaged (20-pixel-wide) amplitude-derived t/λ line profiles were extracted from the reconstructed amplitude images and angularly averaged across the same weakly diffracting orientations discussed earlier ([2.1,3]°). This spatially and angularly averaged line profile is then turned into a "synthetic-wedge" following the procedure in the previous section. The synthetic-wedge is seen in Figure 4.11, on the left, and the linearity of this synthetic-wedge is examined in the linearity-test graph on the right in Figure 4.11. The linearity-test graph shows the results of a linear fit to the data within a 15 nm (thin), 20 nm (medium), and 30 nm (thick) rolling window. The linearity-test results in Figure 4.11 (right graph) reinforces the idea that there is a change in the linearity of the synthetic-wedge in Figure 4.11 (left graph) above 95 nm, so the shell-compensation should, if it is doing well, make the synthetic-wedge in Figure 4.11 (left) appear much more linear.

The next step is to apply the shell-thickness-as-a-function-of-core-thickness plot seen in 4.10b to the line profiles from the holography data in Figure 4.7, Figure 4.8, and Figure 4.11. Figure 4.11 (left) is used as an example, but the same procedure is

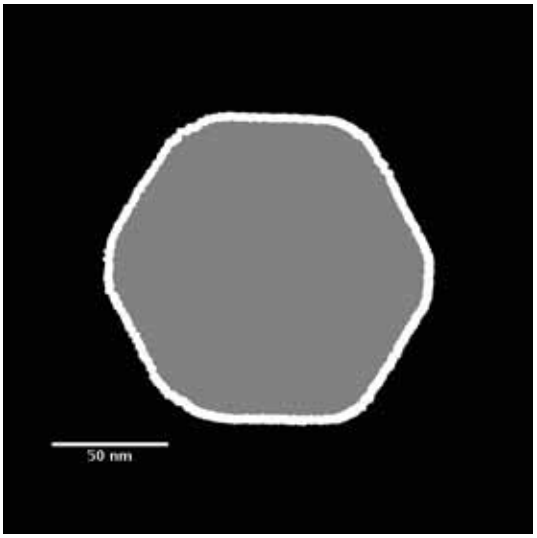


Figure 4.9: Hypothesized shell from discrete tomography reconstruction. The InAs core of the nanowire is grey; the shell white. The DT reconstruction seen in Figure 3.9 is simply scaled up (see text) to match the shell width.

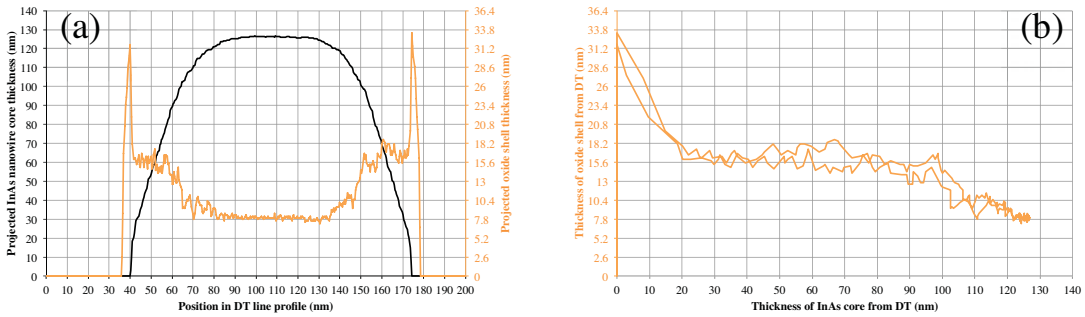


Figure 4.10: Shell thickness as a function of core thickness. The shell, calculated from the DT reconstruction in Figure 4.9, is plotted as a function of (left) position in cross-sectional profile and (right) core thickness. The change seen at t_{core} > ~95 nm explains the sublinearity seen in Figures 4.7 and 4.8. The data in (b) is used to compensate for the shell, and to determine shell parameters in following sections.

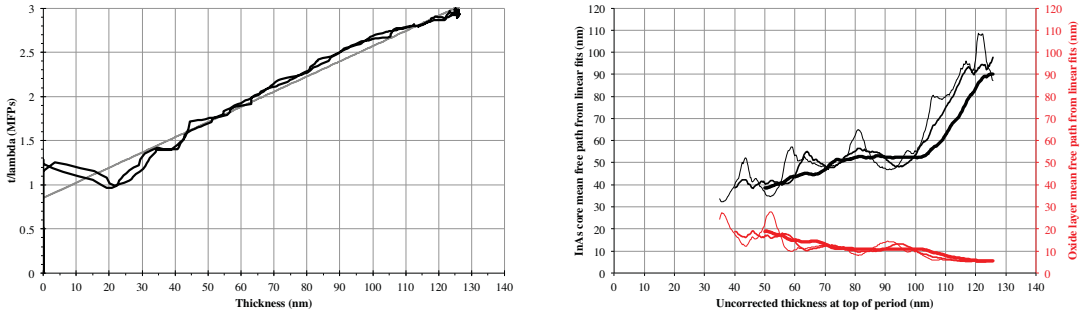


Figure 4.11: Spatially averaged t/λ as a function of thickness. Derived from 20-pixel spatial average from weakly diffracting conditions on the interval [2.1,3.0]°. Left: Nonlinearity is clearly visible, as well as a non-zero intercept. Right: Mean free paths calculated from the slope (black) and intercept (red) of linear fits over 15 (thin), 20 (medium) and 30 (thick) nm thickness windows from left graph, showing the changes in the linear fit above t_{core} > ~95 nm. This should be invariant with thickness, neglecting diffraction effects due to the weakly diffracting condition.

followed for all of the holography synthetic-wedge data. Figure 4.10b (shell thickness as a function of core thickness) is multiplied by a shell compensation factor *F*, and is subtracted from *e.g.* Figure 4.11 [left graph = amplitude-derived t/λ as a function of core thickness]. *F* is fitted such that the resulting linear fit to the shell-compensated amplitude-derived t/λ data has a *y*-axis intercept of 0. If the shell compensation is good, then zeroing the *y*-axis intercept removes the entire shell contribution to the plot, because this *y*-axis intercept is a sign of thickness underestimation due to no shell in the HAADF-derived thickness determination. Performing this shell-compensation procedure yields the plot seen in Figure 4.12, which is a plot of the shell-compensated InAs core t/λ against the DT-derived thickness. The graph in Figure 4.12 shows strong linearity across the entire thickness range, and the sublinearity seen in 4.11 (left) is absent. While the linearity seen below ~25 nm is an artifact, as there is no specimen thickness data there, the rest of the specimen thickness range can be used to determine the linear fit. In practice, the region from 45 nm to the maximum thickness was used for the linear fit in Figure 4.12 to avoid potential thickness artifacts at the lowest measured thicknesses. Both the shell and core mean free paths can be determined from this information. The weighting of the shell compensation and the knowledge of the thickness from Figure 4.10b allow for computation of the shell mean free path, and the slope of the linear fit in 4.12 yields the core mean free path.

Figure 4.12 shows that the shell compensation approach works well for linearizing holography data. The non-linearity in Figures 4.7, 4.8, and 4.11 (left) is most easily explained by this shell compensation.

However, this only uses the DT reconstruction to determine the shell compensation. The effect of using the DT shell-thickness-as-a-function-of-core-thickness on RENOIR cross-sectional thickness profiles is discusses in the next section.

Using DT Shell Compensation on RENOIR Thickness Profiles

Unfortunately, applying the DT shell-compensation data to the RENOIR cross-sectional thickness profile mixes the two thickness-determination methods. The entire principle that "the shell is uniform around the nanowire" is itself an assumption. RENOIR shell determination – through scaling the RENOIR cross-section – is possible, but would be noisier than the DT data, because RENOIR relies on re-calibrated HAADF intensity profiles. Thus, not only does this make the RENOIR data analysis reliant on the DT data, but also there is an additional influence stemming from the different maximum thicknesses from DT and RENOIR.

As noted in Figure 3.21, DT and RENOIR determine different maximum thicknesses of the nanowire – 129.8 nm for RENOIR, 125.6 nm for DT. Therefore, the DT shell compensation has a maximum thickness of 125.6 nm – no shell compensation data exists above this thickness from DT. When applying the DT shell compensation to RENOIR, it would be possible to linearly scale the shell compensation accordingly, such that the maximum became 129.8 nm instead. However, the differences between the DT shell and the "real" RENOIR shell are likely minimal and below the noise level in the shell determination, and this introduces an additional arbitrary scaling factor.

Therefore, the approach taken here is to simply use the DT shell compensation data as-is. This solution is suboptimal, but a DT-shell compared to a RENOIR-shell likely

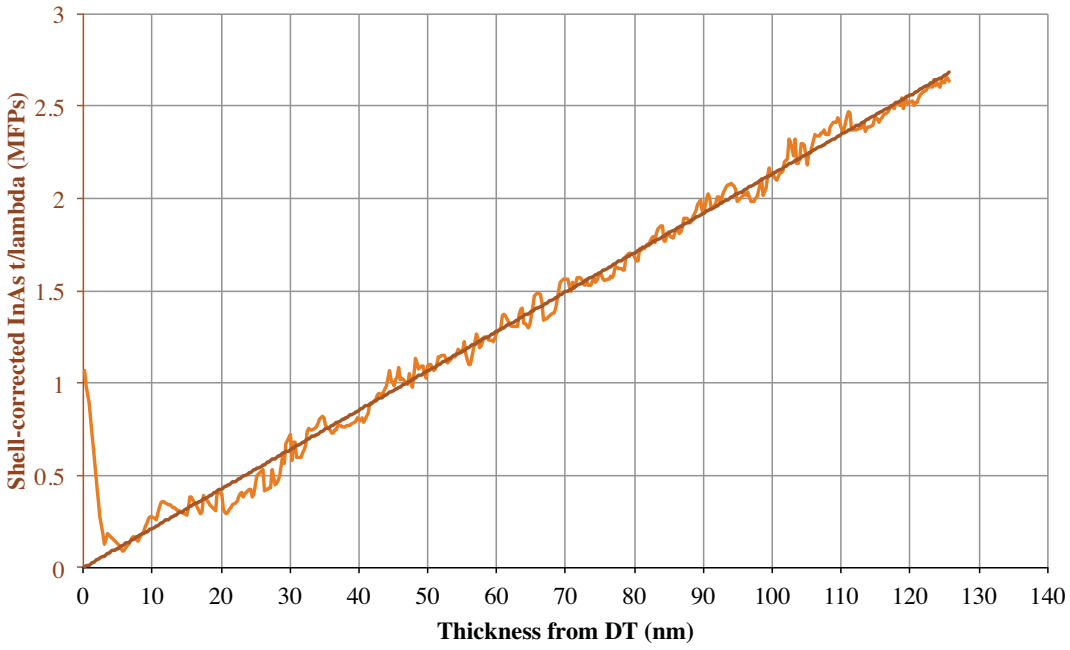


Figure 4.12: Spatially averaged t/λ as a function of thickness from Figure 4.11 with shell correction from Figure 4.10b. This graph corrects for the shell surrounding the wire, and thus improves linearity. This correction is applied by subtracting off the thickness function seen in Figure 4.10b through determining the weight of the shell thickness function by minimizing the linear fit's *y*-axis intercept. This renders the weakly diffracting condition core InAs t/λ linear with thickness – compare to Figure 4.11. This approach is also used on the other phase and amplitude-derived t/λ synthetic wedges. The magnitude of the shell correction to t/λ is related to the shell t/λ.

has minimal influence on RENOIR-derived mean free paths and mean inner potentials – after all, even using the DT data to model the shell in this way is an approximation, as is assuming the shell has a constant density.

Now, this DT-derived shell compensation is applied to the synthetic-wedges, and mean free paths and mean inner potentials for the core of the nanowire and the shell surrounding the nanowire are determined.

The Synthetic-Wedge Results with Shell Compensation

In this section, the blue/green/red synthetic-wedges in Figure 4.7 (phase) and Figure 4.8 (amplitude-derived t/λ) are shell-compensated, and the shell-compensated results analyzed. For both the core and the shell, mean inner potentials are determined from the shell-compensated phase synthetic-wedges, and mean free paths from the shell-compensated amplitude-derived t/λ synthetic-wedges.

The shell compensation procedure from the previous section (Figure 4.10b), when applied to the blue/green/red amplitude-derived t/λ synthetic-wedges shown in Figure 4.8, yields the mean free path results seen in Figure 4.13b using the DT thickness determination – dark blue/green/red for the core mean free path, light blue/green/red for the shell mean free path – for each synthetic-wedge. Analogously, the same shell-compensation routine is carried out on the phase data in Figure 4.7, with the results seen in Figure 4.13a. Both Figures 4.13a and 4.13b are using the DT thickness determination for the line profile and for the shell determination. The analogous graphs using RENOIR thickness data are in 4.14b for the t/λ lines, and 4.14a for the mean inner potential, both of which use the DT shell determination. The black and grey bars in Figure 4.13b and 4.14b come from the spatially averaged synthetic-wedge in Figure 4.12: the black bar in Figures 4.13b and 4.14b is the core and the grey bar is the shell. As noted above, Figure 4.13b is using the DT thickness determination, and Figure 4.14b is using the RENOIR thickness determination.

First the mean free path results (Figures 4.13b, 4.14b) are discussed in the following section, and then the mean inner potential results (Figures 4.13a, 4.14a) are discussed in the section after. Because the 20-pixel spatially averaged mean free path data is the best example of the amplitude-derived t/λ data, it will be examined first, with the individual blue/green/red 1-pixel mean free path synthetic-wedges discussed after. This section includes a comparison between the synthetic-wedge generated from the 20-pixel spatially averaged data and the synthetic-wedges generated from the 1-pixel blue/green/red line profiles. In the section after, the mean inner potential data will be discussed. The mean inner potential data only has 1-pixel (non-spatially-averaged) synthetic-wedges, but the analysis of Figures 4.13a and 4.14a is not impeded by only having 1-pixel synthetic-wedges.

Mean Free Path Determination

From the nanowire core and shell mean free path graphs in 4.13b and 4.14b, several characteristics are discussed in this section. First, the short mean free path of the shell as compared to the InAs core is discussed. Then, the wide scatter in the shell data is compared to the minimal scatter in the core data. Finally, the disparity between the DT results (Figure 4.13b) and the RENOIR results (Figure 4.14b) will be examined. Another amplitude-derived component, the absorption parameter, is discussed in the simulation section later in this chapter.

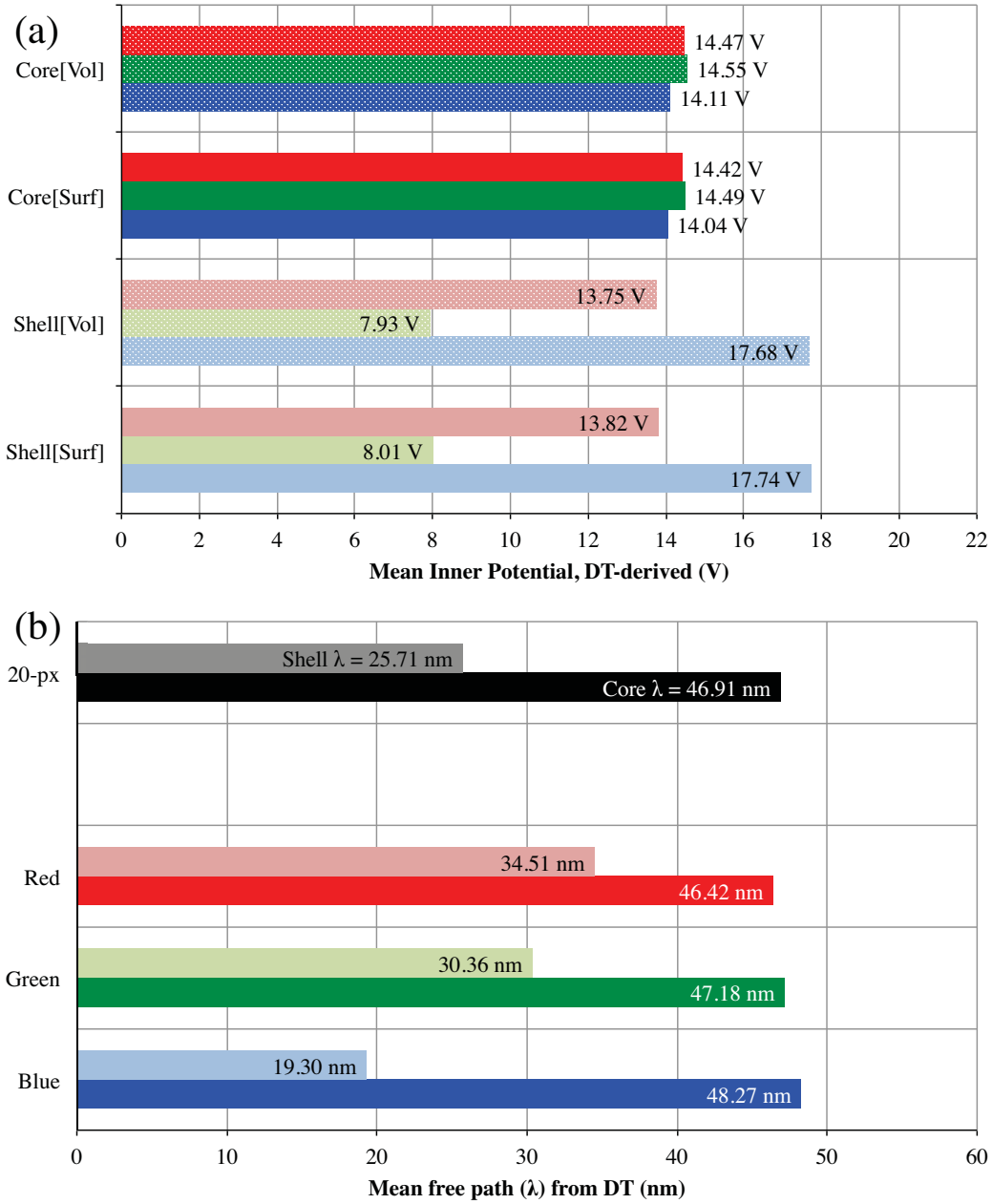


Figure 4.13: Mean inner potentials and mean free paths from weakly diffracting conditions using discrete tomography thickness determination. By using the approach outlined in Figure 4.12, the mean inner potentials (top) and mean free paths (bottom) are determined for weakly diffracting conditions in the tilt series. InAs core values are determined using the slope of linear fits like the one shown in *e.g.* Figure 4.12, and shell values are determined by the magnitude of the shell compensation like that applied to Figure 4.11 to yield Figure 4.12. Mean inner potential values are using volume "[Vol]" and surface "[Surf]" charge calculations. The grey and black "20-px" lines in the mean free path come from the 20-pixel spatial average shown in Figure 4.12, instead of the red, green, and blue data that come from the 1-pixel line profiles shown in Figure 4.6.

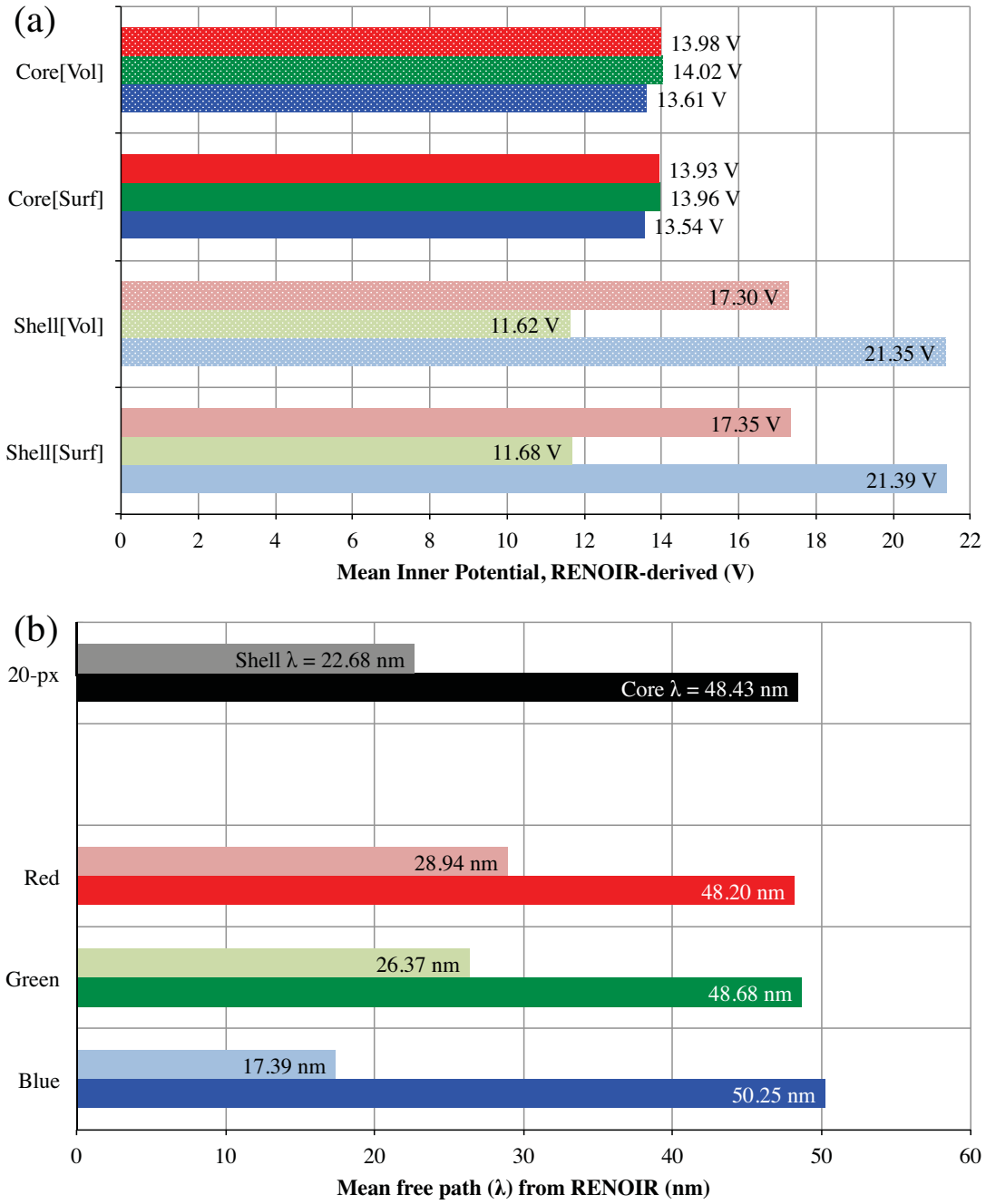


Figure 4.14: Mean inner potentials and mean free paths from weakly diffracting conditions using RENOIR thickness determination. As figure 4.13, but using RENOIR thickness determination instead of discrete tomography, on the same holography data. The differences in thickness determination lead to differences in determined quantities (mean inner potential and mean free path).

The first result of the mean free path determination is the short mean free path of the shell surrounding the wire as compared to the InAs core. This is unexpected, as the shell shows less intensity in the HAADF images than the core, which would intuitively indicate that the shell scatters less strongly and has a longer mean free path than the core. However, this result is consistent with a back-of-the-envelope calculation from the intensity measured using conventional BFTEM, which, when treated in the same way as the amplitude images considered here, yields the results in Figure 4.15. Comparing Figure 4.15 to the projected shell thickness near the edge of the nanowire cross-section in Figure 4.10a, the ~32 nm shell thickness near the edge of the nanowire cross-section shown in Figure 4.10a combined with the ~1.6 maximum mean free paths in BFTEM shown in Figure 4.15 compare favorably with the DT and RENOIR mean free paths of 25.71 nm and 22.68 nm, respectively, from the spatially-averaged data in Figures 4.13b and 4.14b. Therefore, while it is not intuitive that the shell mean free path should be so short, given the HAADF data, the holography data indicates that the shell mean free path is short, a conclusion reinforced by analysis of the BFTEM data. This might result from the shell not yielding much high-angle scattering, but much more low-angle scattering – possibly linked to the atomic number dependence of mass-thickness contrast (see Chapter 2 for more details).

The imprecision of the shell mean free paths is apparent when comparing the single-pixel line profiles with the spatially averaged line profiles. The precision is more easily examined for the amplitude-derived mean free path data rather than the phase-derived mean inner potential data. This is because the amplitude-derived mean free path data includes a spatially averaged (20-pixel) line profile that has improved noise-reduction as compared to the 1-pixel line profiles, and the phase-derived mean inner potential data does not have this. The phase-derived mean inner potential data does not have this spatially averaged line profile, making it somewhat more difficult to analyze the imprecision of the individual phase line profiles. The single-pixel blue/green/red line scans in Figure 4.13b show a variation in shell mean free path from the spatially averaged, high-quality data of +34%, +18%, and -25% for DT thickness data. This is true for the RENOIR thickness determination as well: the variation in shell mean free paths is +28%, +16%, and -23% for RENOIR thickness data (blue/green/red relative to grey in Figure 4.14b). These are wide variations in the shell mean free path, and the core has much smaller variations: for comparison, from the same figures, the variations in the core mean free path from the spatially averaged high-quality data are -1.0%, +0.6%, and +2.9% for DT thickness data, and -0.5%, +0.5%, and +3.7% for RENOIR. The deviations in mean free paths of the core and the shell are inversely correlated – a point that is addressed in the maximum-thickness section.

The disparity between the RENOIR and DT data is attributable to their different thickness determinations. For the spatially averaged data in Figured 4.13b and 4.14b, the RENOIR shell mean free path is 12% below the DT shell mean free path, and the RENOIR core mean free path is 3.2% above the DT core mean free path. Again, as with the single line profiles, the deviations as a function of thickness determination technique for the core and shell MFP are inversely correlated. The results from applying different thickness determination techniques show the criticality of the thickness determination – while the shell is more sensitive, the core is also sensitive to a change in thickness determination.

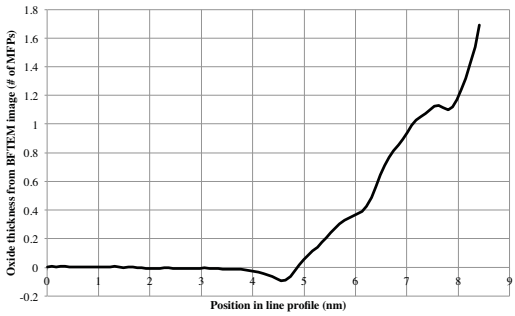


Figure 4.15: Bright-field TEM line profile at nanowire edge. The short surface layer mean free path in Figures 4.13b and 4.14b is possibly reinforced by this rough calculation. This is from bright-field TEM data that is processed in the same way as the holography amplitude images. It shows the ~4 nm wide surface shell (probably a native oxide) as having a short mean free path. If the surface layer thickness near the edge is ~20nm as given in Figure 4.10, then this is potentially consistent with the short mean free path given in Figures 4.13b and 4.14b. This is not meant to be conclusive (note defocus fringe visible around 4.5 nm), but to indicate that the short shell mean free path is plausible.

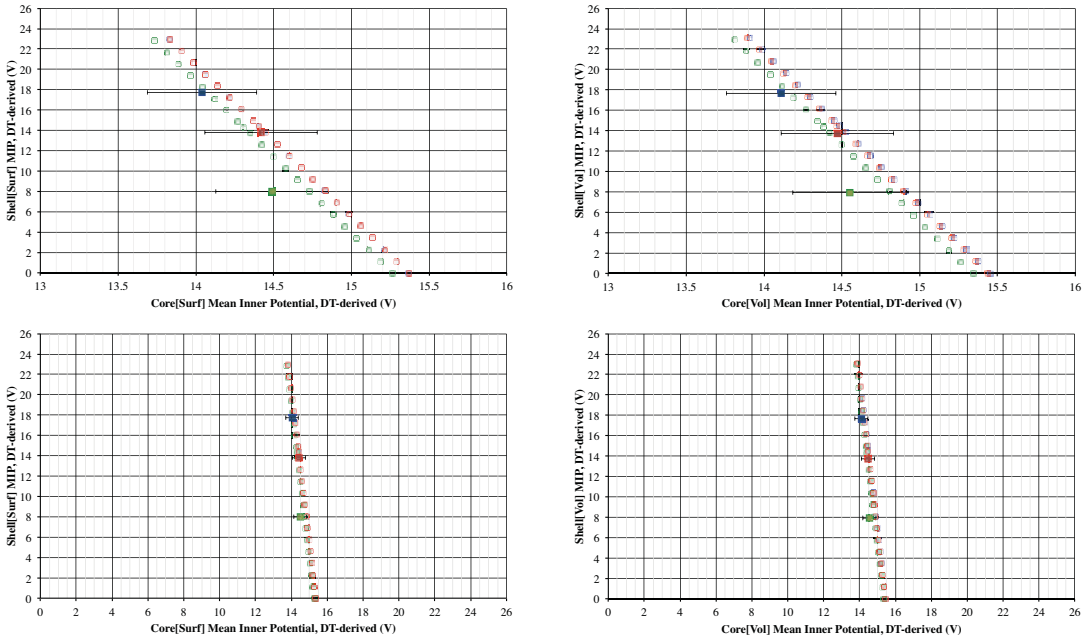


Figure 4.16: Interdependence of core and shell mean inner potentials for DT thickness. The wider spread in shell data than core data, seen in Figures 4.13 and 4.14, and the interdependent nature of the fitting process to determine the core and shell mean inner potential (and mean free path) data raises the question of how strong this interdependence is. Would it be more plausible to have wide variation in the shell than the core? By examining the effects of a slight misdetermination of the core mean inner potential on the shell mean inner potential (i.e. fixing the core mean inner potential, and determining the shell mean inner potential given this fixed core, fixed thicknesses, and fixed measured phase shift), the wide variations seen in Figures 4.13 and 4.14 are partially explained. By plotting the shell mean inner potentials as a function of the core (for surface [Surf] and volume [Vol] charging models), the relative insensitivity of the core to the shell can be seen. The only difference between the top and bottom graphs is the x-axis scaling.

Mean Inner Potential Determination

The mean inner potential V_0 is determined from the phase data in the same way that the mean free path is determined from the amplitude-derived t/λ data. The subtraction of the Figure 4.10b shell-thickness-as-a-function-of-core-thickness synthetic-wedge is fitted to the phase synthetic-wedge to bring the intercept to 0. This yields a plot of only the phase shift due to the InAs core of the nanowire. The slope of this phase shift ($d\phi/dt$) provides the mean inner potential:

$$\phi = C_E V_0 t \quad \Rightarrow \quad \frac{d\phi}{dt} = C_E V_0 \quad \Rightarrow \quad V_0 = \frac{1}{C_E}\frac{d\phi}{dt}$$

The weight of the oxide/surface-layer shell correction provides the mean inner potential of the shell using the first equation above, without the differential, because the shell thickness is known.

The mean inner potentials derived from the DT and RENOIR thickness measurements are shown in Figures 4.13a and 4.14a, respectively. These figures graph the mean inner potential for both the core and shell sections, and apply to each both the surface-charge-compensation (solid bars) and volume-charge-compensation models (dotted bars) discussed earlier and in Appendix A. Several observations are apparent, one of which is unique to the phase shift section: whether the surface- and volume-charge models described in Appendix A is better cannot be determined from the current data – the difference between the charging models is too small. The other observations are the same as the mean free path analysis in the previous section: the shell mean inner potential has much more scatter than the core mean inner potential, but appears to likely be lower; the core and the shell mean inner potentials show the same inverse relationship as the mean free paths; the mean inner potentials from the RENOIR and DT thickness models do not lie within each others' confidence intervals.

The choice of charge-compensation model makes an order of magnitude less difference in the mean inner potential than the choice of thickness models. As seen in Figures 4.13a and 4.14a, the difference in mean inner potential between the volume- and surface-charge models is about 0.05V, while the difference in mean inner potential between the two thickness models is about 0.5V. This means that, while the charge compensation in general is necessary to impose a flat vacuum, the exact nature of that charge compensation is not determinable without thickness knowledge whose accuracy and precision is much better than current techniques.

The shell mean inner potential, like the shell mean free path in the previous section, has much higher scatter than the core mean inner potential. The bar graphs in Figures 4.13a and 4.14a starkly show this difference in scatter. Even given this scatter, 2 of the 3 measurements show a mean inner potential of the shell less than that of the core. They also appear inversely correlated – the highest shell mean inner potential is correlated with the lowest core mean inner potential, and vice versa. This is the same relationship as the one described for the mean free path data in the previous section. This is likely due to the same reason, which will be discussed in the following sections. Therefore, the mean inner potential of the shell is indeterminate due to the high imprecision of the data, but the shell mean inner potential is possibly lower than the mean inner potential of the core.

Most critically, the two thickness-determination methods do not yield mean inner potential values of the core of the nanowire that overlap. The two differ by approximately 0.5 V – with the DT-determined mean inner potentials lying around 14.5 V, and the RENOIR-determined mean inner potentials around 14.0 V, both excluding the outlier blue point, which generally lies about 0.45V under the green and red points (for comparison, note the difference in Figure 4.6, largely suppressed in Figure 4.7, but reappearing in Figures 4.13a and 4.14a). This difference between the RENOIR and DT mean inner potentials indicates that, like the mean free path, the two thickness-determination methods do not agree on the mean inner potential, with an approximately 3.5% disparity between the RENOIR and DT mean inner potentials.

Conclusions From the Synthetic-Wedge Information

From this section, several conclusions can be drawn. First, the synthetic-wedge method provides phase-derived mean inner potentials and amplitude-derived mean free paths. The mean free paths and mean inner potentials are both analyzed in the same way, and many of the same conclusions hold. It is a useful way to decompose the data into core and shell components. The core measurements appear to be more precise than they are accurate, due to the systematic uncertainty over the thickness determination technique. The shell measurements are very noisy, and thus unreliable. From the mean inner potential data, it is clear that the effect of specimen charging can be removed, but whether the surface- or volume-charge data is more reliable cannot be answered from the current dataset.

The mean inner potentials and mean free paths from the core appear generally well behaved. For the mean free path of the nanowire core, the number obtained from the averaged-line-profile and the individual profile measurements agree relatively closely. This improves confidence in the core mean inner potential determination, which does not have the benefit of comparison with an average. However, one of the three core mean inner potential data points – the “blue” one in Figures 4.13a and 4.14a – is below the other two (“green” and “red”). The high shell mean inner potential might help illuminate the reason, and the inter-relationship between the core and the shell mean inner potentials for a given profile is examined in the maximum-thickness section.

Compared to the core data, extracting either accuracy or precision from the shell mean inner potential or mean free path data is less plausible. The mean free path of the shell appears to be shorter than in the core, and the mean inner potential possibly lower in the shell than in the core. This is a counterintuitive result, but likely related to the amorphous structure of the shell. Further work on this question would be necessary to determine the exact parameters.

The core measurements of mean inner potential and mean free path are relatively precise, but of questionable accuracy, due to systematic thickness uncertainty. The two thickness determination techniques, when applied to this nanowire yield fundamentally different results. This is expected, as they determine different maximum thicknesses and different thickness profiles. However, as the “true” maximum thickness is not known, this is not determinable.

The effects of miscalibration and error estimation are treated in the following section. Miscalibration and error estimation are important factors for understanding why the

mean inner potential and mean free path data are precise for the core and not for the shell.

Precision and Accuracy: Effect of Miscalibration and Error Estimation

Both systematic and random errors affect this electron holography experiment on the InAs nanowire. The systematic errors (*e.g.* the results from different thickness determination techniques) are more troublesome than the random ones. There are lateral errors – those affecting the x-axis of the synthetic-wedge – and intensity errors – those affecting the y-axis of the synthetic-wedge. However, one set of possibly interlocking errors combines both lateral and intensity error, potentially providing the explanation for the shell layer's mean inner potential and mean free path imprecision.

In this determination of the InAs nanowire mean inner potential and mean free path, each individual value is subject to approximately the same lateral error influences. ±1.5% error due to magnification calibration added in quadrature with ±2% error if the width of the wire is ±2 pixels wider in the holography images than estimated leads to a ±2.5% error in lateral determination from the holography images. This is true for both DT and RENOIR.

Estimation of the intensity errors in this instance is perhaps simpler, but the intensity error cannot be applied to the mean free path and mean inner potential data. Intensity error estimation is perhaps simpler, because there are both 1-pixel wide and 20-pixel wide amplitude-derived t/λ synthetic-wedges. The results from the mean free path comparison between spatial averages and single line profiles, stated earlier (determined from Figures 4.13b, Figure 4.14b), would indicate that the total intensity error in the measured mean free paths is possibly around 2%. However, caution must be taken due to the limited number of data points.

As the amplitude-derived t/λ shows, the intensity error could be measured and improved by increased spatial averaging. However, as discussed earlier, spatial averaging was not performed in this work with the phase data due to phase wraps. The spatially averaged results appeared much more precise than the individual line profiles. However, as the noise level in the phase synthetic-wedges (Figure 4.7) is much lower than that in the amplitude-derived t/λ synthetic-wedges (Figure 4.8), this is not likely to help the phase profiles much. The intensity error could also be improved by noise reduction in the shell calculation. These are calculated from the discrete tomography reconstruction – which leads to noise enhancement in the DT outline of the shell, as can be seen from Figure 4.10b.

However, the largest single contributing factor to the inaccuracy is systematic error. The pair of thickness-determination techniques, RENOIR and DT, both determined from logical premises, disagree with each other by about 3.5% systematically. This affects accuracy more than precision. The two thickness determination techniques, therefore, are mutually incompatible quantitatively, but are internally self-consistent.

However, a subtler miscalibration effect might also lead to inaccuracy in determining core and shell mean inner potentials and mean free paths. If the width of the wire were underestimated on *e.g.* the amplitude-derived t/λ cross-sectional profile, then the intercept of the line would be increased, but the maximum-thickness region would not change. Thus, a width underestimation would decrease the slope of the linear fit while simultaneously raising the intercept. Thus, the mean free path of the core would decrease and the mean free path of the shell would increase. The same effect would

arise if the magnification calibrations were incorrect, or if the line profiles drawn on the amplitude and phase images were slightly rotated relative to the wire. This is part of the fundamental reason why correct magnification calibration is critically important for this experiment – it forms the basis for all of the thickness quantification in nanometers. In addition, this miscalibration would not necessarily be detectable if it were present. Since, at the magnification used during this experiment, the nanowire itself is under a hundred pixels wide in the amplitude and phase images reconstructed from holograms, this miscalibration could be difficult to detect. Miscalibration effects were minimized to the greatest possible extent, but there is no way to verify their complete removal. Miscalibration effects would be the greatest on the shell layer, because that derives from the intercept of the synthetic-wedge. Hence, subtle differences, including miscalibrations, of the holography images of the nanowire is one reason why the shell mean free path and mean inner potentials may be noisier than the nanowire core mean free paths and mean inner potentials.

Therefore, to avoid these miscalibration effects to the maximum extent possible, and rely minimally on the width determination, this discussion proceeds to address the interdependence of the nanowire core and shell from the wide, flat maximum-thickness region of the wire.

At Maximum Thickness

Evaluating the maximum-thickness region (129.8 nm thick *via* RENOIR, 125.6 nm *via* DT) of the weakly diffracting condition of the InAs nanowire provides different information from the synthetic-wedge method. The advantages of the maximum-thickness method are an increased signal-to-noise ratio, and using only one point, instead of a more complicated differential and linear fit. The principal disadvantage is the lack of a shell reference function to fit against. Without a shell reference function, determining *e.g.* the mean inner potential from the maximum-thickness region would mean one data point (average maximum phase) being used to determine two parameters (core and shell mean inner potentials). Therefore, the maximum-thickness region is most useful for examining the interdependency of the core and shell regions.

The major advantage of the maximum-thickness method over the synthetic-wedge is the spatial averaging that it allows, with a significant additional advantage provided by the reduced undetectable sensitivity to exact calibration detailed two sections previously. By using the wide, flat top of the nanowire in the flat-topped “plateau” orientation (see Chapter 3), a region of the nanowire with a single thickness can be analyzed quantitatively. By averaging over the entire maximum-thickness region, the signal-to-noise ratio of the measurement can be improved. The maximum-thickness method, by avoiding a wedge, also avoids the calibration issues in the previous section. While the maximum-thickness method does rely on the spatial calibrations to determine the maximum thickness – either directly from the discrete-tomography reconstruction or through the width-determination in RENOIR - it does not invoke the subtle inverse correlation between core and shell that the synthetic-wedge method uses, and which was detailed in the second-previous section, where misdetermination of the width undetectably modifies the relationship between the core and shell parameters. It does not invoke this correlation because it does not determine the core and shell mean inner potentials, only the relationship between them.

As mentioned above, the major disadvantage of the maximum-thickness method is the reliance on a single data point to determine two parameters. The one phase value at

maximum thickness is used to determine both the core and the shell mean inner potentials, and the one maximum amplitude-derived t/λ value is used to determine the mean free paths of both the core and the shell. This generates one data point, and two unknowns – *ergo*, the mean inner potential or mean free path of the core and the shell cannot be determined individually. However, the relationship between them can be determined, and this interrelationship is now examined.

Mean inner potentials of the nanowire core and the nanowire shell have different sensitivities. The resolution to the problem of two unknowns and one data point is through picking a ratio, and partitioning the total phase shift into "core" and "shell" components accordingly. To take a simple example (with numbers chosen purely for example), if the total phase shift is 10 radians in the thickest part of the wire, and 90% of this is due to the 90 nm core, making 10% of this total phase shift the result of the 10 nm shell, then both the core and the shell have the same mean inner potential. However, if 95% of this phase shift is due to the core, and 5% to the shell, the mean inner potential of the shell is reduced by 50% compared to the "10% shell" case, while the mean inner potential of the core increases by about 6% compared to the "90% core" case. Varying this core/shell ratio leads to a plot of the dependence of the shell mean inner potential on the core mean inner potential. This same procedure is also considered for the amplitude-derived mean free paths.

To be clear, the maximum-thickness regions from the phase and amplitude-derived t/λ data does not determine the mean inner potential or the mean free paths of the specimen – only the interdependence between the core and the shell mean inner potentials or the core and the shell mean free paths, and possible ranges for the core mean inner potential under a set of assumptions about the shell mean inner potential.

Maximum-Thickness Mean Inner Potentials

There is an inverse linear relationship between the core and the shell mean inner potentials for both DT and RENOIR thickness-determination methods. The shell mean inner potential as a function of the core mean inner potential is plotted in Figure 4.16 for the DT-derived maximum thickness (125.6 nm), and Figure 4.17 for the RENOIR-derived maximum thickness (129.8 nm). Figures 4.16 and 4.17 show a linear relationship between the core and the shell mean inner potentials. The filled points in Figures 4.16 and 4.17 correspond to the values from the synthetic-wedge method, detailed in the previous section, and RGB coloring maintains consistency. In Figures 4.16 and 4.17, both charge models are shown, indicated with [Surf] for surface charge (on the left, in both Figure 4.16 and Figure 4.17), and [Vol] for volume charge (on the right, in both Figures 4.16 and Figure 4.17), but this makes a general difference of ~0.1 V in the core MIP, with MIP[Surf] > MIP[Vol]. The effect of surface-versus-volume charge is a greater difference for the maximum-thickness region than the synthetic-wedges discussed earlier. This greater influence is expected because the maximum effect of charge compensation is in the middle of the wire.

As expected, Figure 4.16 and Figure 4.17 show that small changes in the core mean inner potential lead to large changes in the shell mean inner potential, explaining the difficulty in determining shell properties. If the shell of the nanowire is assumed to have a mean inner potential in the range of 8-16 V, then, according to the DT-thickness, the core should have a mean inner potential of 14.2-14.9 V, and, according to RENOIR-thickness, the core should have a mean inner potential of 13.7-14.3 V.

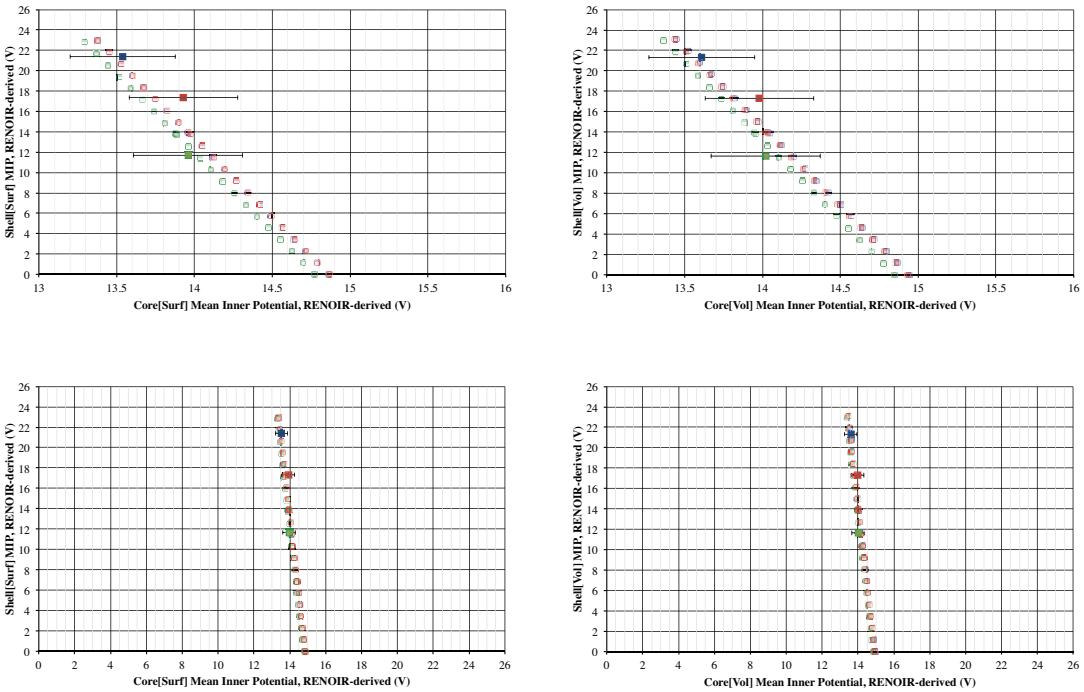


Figure 4.17: Interdependence of core and shell mean inner potentials for RENOIR thickness. The same effects are present for RENOIR and DT (see Figure 4.16). The differences in the two thickness determination results lead to the same behavior, but different absolute numbers for the mean inner potential.

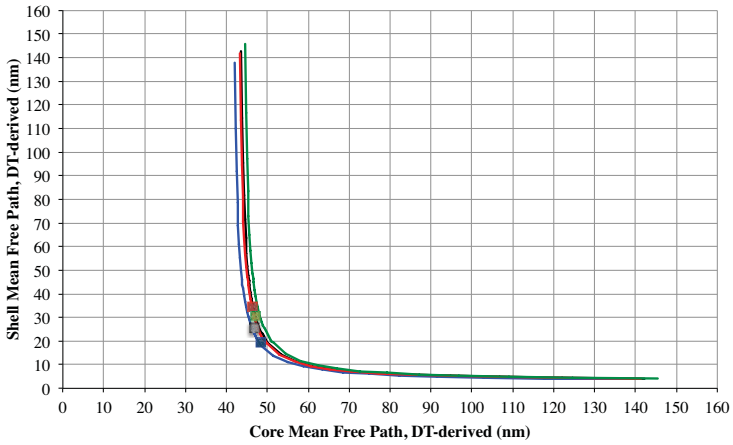


Figure 4.18: Interdependence of core and shell mean free paths for DT thickness. This is carried out in the same way as Figure 4.16, and for the same reason – to determine the interdependency between the core and shell. This graph takes the known t/λ experimental data and the known thicknesses, and determines the shell mean free path as a function of the core mean free path. The shape of this dependence, in contrast to that in Figures 4.16 and 4.17, can be either sensitive to fluctuations in the core mean free path or in the shell mean free path, and sets asymptotic lower limits to the mean free paths of the core and the shell.

Thus, the core mean inner potential only slightly changes when the shell mean inner potential is doubled.

The core mean inner potentials calculated using this method are consistent with the synthetic-wedge determinations. Using the DT thickness determination, the single core mean inner potential determined from the synthetic-wedge method is comparable to the range of core mean inner potentials estimated using the maximum-thickness method. When considering the above 8-16 V wide range for the mean inner potential of the shell, the midpoint of the range of core mean inner potentials is ~14.6 V for the DT-determined maximum-thickness plots, near the ~14.5 V DT synthetic-wedge mean inner potential. The same is true when the same procedure is carried out using the RENOIR thickness determination: the midpoint of the range of is ~14.0 V, near the ~14.0 V RENOIR synthetic-wedge mean inner potential. Comparing these numbers, the maximum-thickness method shows the same disagreement between RENOIR and DT as the synthetic-wedge method. In Figure 4.16 and Figure 4.17, the individual data points from the synthetic-wedge method are given as filled squares of the appropriate color, with 2.5% core error bars indicated.

While the mean inner potentials of the nanowire core and the shell influence each other, both are still dependent on the total phase. Figure 4.16 and Figure 4.17 show "green" points with a lower total phase than the "red" or "blue" points. Comparing the green points to the red and blue points, this lower total phase yields lower mean inner potential values for both the core and the shell. This lower total phase of the green points can be detected on the original phase graphs in Figure 4.6, but it is a subtle effect in Figure 4.6. This is in comparison with the synthetic-wedge method seen in Figure 4.13a and Figure 4.14a, where the green data points determined the largest core mean inner potential and the lowest shell mean inner potential. However, all of the maximum-thickness mean inner potential results in Figures 4.16 and 4.17 are in agreement with the synthetic-wedge mean inner potential determinations in Figures 4.13a and 4.14a. In summary, the maximum-thickness method is useful for determining acceptable ranges for the mean inner potential.

Errors in the thickness determined here are likely to be only the magnification-calibration error of 1.5% discussed in Chapter 3, giving the maximum-thickness mean inner potential determination a higher precision than the synthetic-wedge data. However, the accuracy is still affected by the conflicting RENOIR and DT thickness determinations.

The maximum-thickness method reinforces the synthetic-wedge method's determination of the core and shell mean inner potentials of the nanowire. As previously discussed, because there is no shell compensation, meaning the maximum-thickness method cannot yield a single value for the mean inner potential. However, the maximum-thickness method is used to examine the interdependence of the core and shell mean inner potential measurements. The conclusions are that the shell mean inner potential is very sensitive to the core mean inner potential, but not *vice versa*, and that the synthetic-wedge method's mean inner potential determinations are within the margin of error. In the next section, the relationship between the core and shell amplitude-derived mean free paths is discussed.

Maximum-Thickness Holography-Derived Mean Free Paths

The maximum-thickness method, when applied to the amplitude-derived t/λ holography data, shows that inverse relationship between the core and the shell mean free paths is nonlinear, but still consistent with the synthetic-wedge data. In Figures 4.18 and 4.19, the DT (4.18) and RENOIR (4.19) shell mean free path as a function of core mean free path is plotted in the same manner as the mean inner potentials in Figures 4.16 and 4.17, with the addition of the 20-pixel spatially averaged data. However, because the mean free path depends inversely on the thickness (instead of directly, like the mean inner potential), the maximum-thickness mean free path curves in Figures 4.16 and 4.17 appear quite different to the mean inner potential lines.

The maximum-thickness mean free path core/shell interdependence, in Figures 4.18 and 4.19 is consistent with the mean free path determinations from the synthetic-wedge in Figures 4.13a and 4.14a. All three curves of shell mean free path as a function of core mean free path in Figures 4.18 and 4.19 show similar behavior – asymptotic approaches to a mean free path <10nm for the shell and a mean free path between 40 and 50nm for the core. The individual linear fits from the synthetic-wedge method are plotted as individual points in Figures 4.18 and 4.19. The individual linear fits are all relatively consistent with each other and with the interdependence graphs. However, the core mean free paths determined from the synthetic-wedge method is near the asymptotic-minimum, which, according to Figures 4.18 and 4.19, makes the shell mean free path difficult to precisely determine. This explains why, in Figures 4.13a and 4.14a, the shell mean free paths are considerably less precise than the core mean free paths.

The mean free paths core/shell relationships determined *via* the two different thickness-determination methods – DT (Figure 4.18) and RENOIR (Figure 4.19) – show minimal disagreement. Comparing Figures 4.18 and 4.19 directly, the curves of shell mean free path as a function of core mean free path are similar in form, magnitude, and range. The core and shell synthetic-wedge mean free paths, in Figures 4.13a and 4.14a, do show a difference with thickness-determination method. One likely reason why the maximum-thickness method does not show this thickness-determination difference is the non-linear relationship between the core and shell mean free paths. This non-linear core/shell relationship means the shell mean free path is difficult to precisely determine, affecting the ultimate precision of the core mean free path determination. The agreement between the two thickness-determination methods on the maximum-thickness mean free path is unlike the mean inner potential, discussed in the previous section.

Therefore, from the maximum-thickness mean free path plots in Figures 4.18 and 4.19, it can be determined that the synthetic-wedge mean free path determinations are plausible, but that the core mean free path is close to the asymptotic limit that would make the shell mean free path difficult to quantify precisely. Because there is a relatively thin shell around the nanowire compared to the thick core, it is difficult to characterize the shell.

Conclusions From Maximum-Thickness Analysis

The shell around the nanowire means the maximum-thickness method cannot be used on this holography data to determine mean inner potentials or mean free paths. It cannot be used because the maximum-thickness method has one known value (*e.g.* average maximum phase in the nanowire cross-section) but two unknowns (*e.g.* core

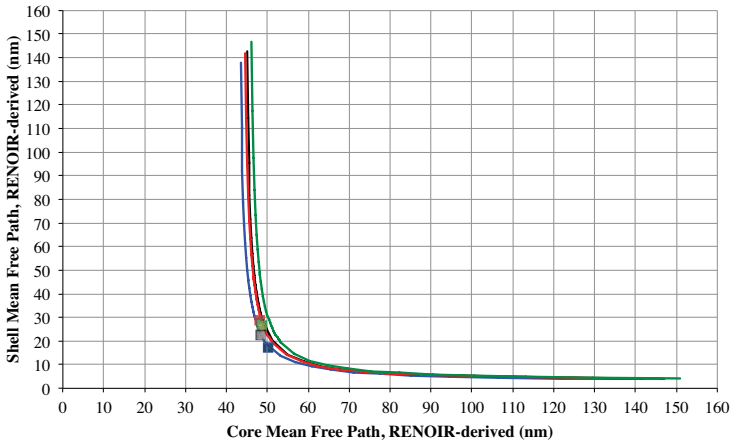


Figure 4.19: Interdependence of core and shell mean free paths for RENOIR thickness. The same behavior is observed for RENOIR thickness-determination results as DT thickness-determination results in Figure 4.18. It would not be possible from this data to distinguish clearly between the two thickness determination methods.

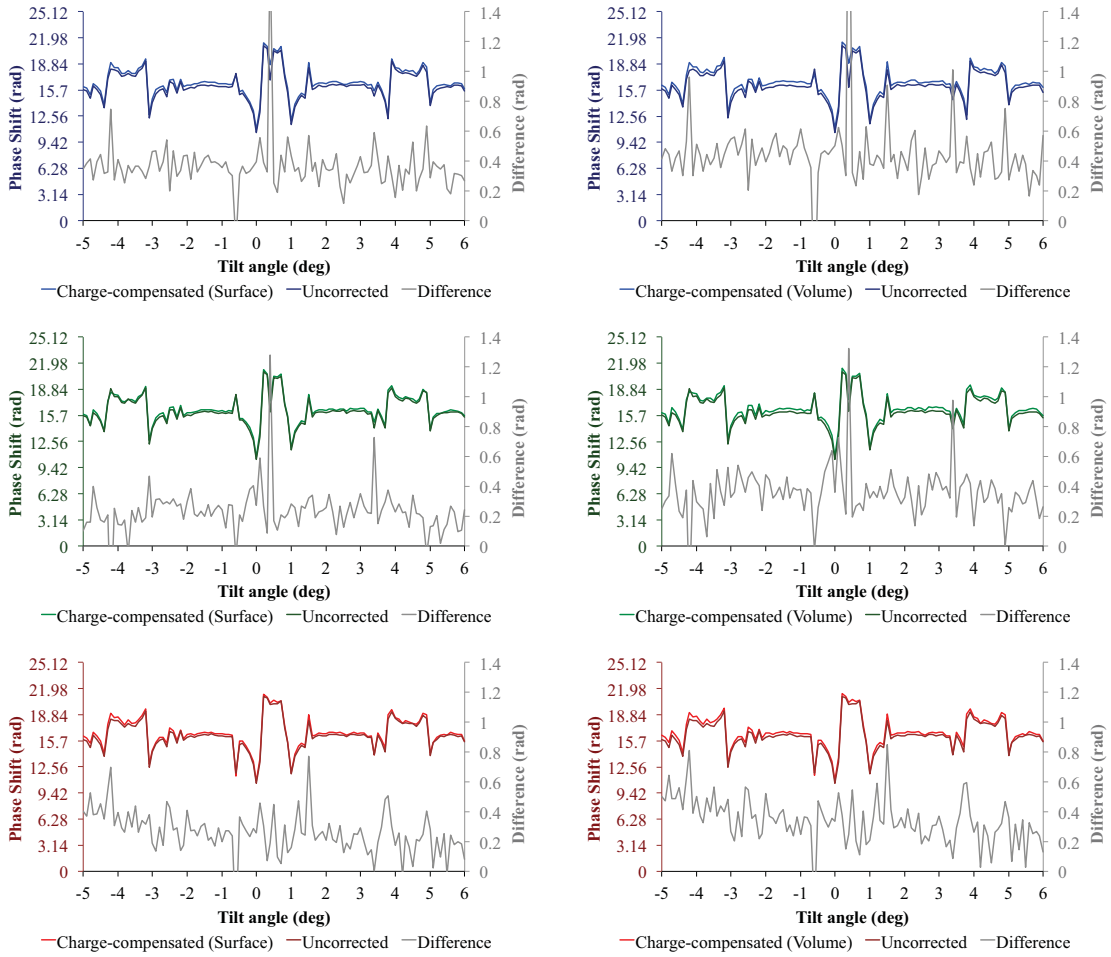


Figure 4.20: Charging as a function of tilt angle. For surface charge (left) and volume charge (right), the differences are minor between the two charging models, and the phase data in this figure is too noisy to be able to distinguish between them. However, the charge compensation is still ~0.3 V relative to as-acquired phase data.

and shell mean inner potentials). The synthetic-wedge method, discussed in the previous sections, does allow for quantitative mean inner potential and mean free path determination. However, the maximum-thickness method, as used here, allows for examination of the relationship between the core mean free path (or mean inner potential) and the shell mean free path (or mean inner potential). This relationship provides a check on the synthetic-wedge method's mean inner potentials and mean free path. According to the maximum-thickness analysis, both mean free path and mean inner potential synthetic-wedge determinations, from the previous section, are acceptable, even if the shell values are imprecise. For both mean inner potentials and mean free paths, the core quantities are much more precise than the shell quantities.

The phase is used to show a linear dependence of the shell mean inner potential on the core mean inner potential. The synthetic-wedge mean inner potential quantities are consistent with the relationships determined using the maximum-thickness method. Different thickness-determination methods, however, lead to relatively different mean inner potentials: for a shell mean inner potential of 12.0 ± 4.0 V, the core mean inner potential is around 14.6 ± 0.3 V if DT thickness-determination is used, and 14.0 ± 0.3 V is the RENOIR thickness determination is used. There is possible, if minimal, overlap between the two thickness models. Therefore, the two different thickness models from Chapter 3 pose an intractable problem for mean inner potential analysis in this chapter.

In contrast to the mean inner potential, the mean free path has an inverse-linear thickness dependence, and the thickness model has less of an effect on the mean free path using the maximum-thickness model. The interdependence curves show that, if the mean free path of the nanowire core is less than approximately 50 nm, small shifts in the core mean free path lead to large, nonlinear shifts in the nanowire shell mean free path. The mean free paths determined using the synthetic-wedge method fall within the interdependence curves determined by the maximum-thickness method. However, because the core mean free path from the synthetic-wedge method is around 50nm, the shell mean free path is difficult to determine. In contrast, the differences between the two thickness models are minimal. As a result, the mean free path of the core is relatively precise, while the mean free path of the shell is not, irrespective of thickness-determination method.

The maximum-thickness method allows for analysis of the sensitivity of the shell mean free path (or mean inner potential) as a function of core mean free path (or mean inner potential). For both the mean inner potential and the mean free path, the shell values are very sensitive to the core values. This is probably due to the relative thicknesses of the nanowire core (~125 nm) and the shell (~8 nm) in this experiment. However, this makes it difficult to determine the exact mean inner potential or mean free path values for the core or the shell; the shell values are very dependent on the core values.

In summary, the maximum-thickness analysis allows for the relationship between the core and shell parameters to be examined. If either was precisely known, then the other could be easily determined. The linear fit parameters lie within the acceptable range of parametric variation, and the maximum-thickness data is subject to fewer errors. The choice of thickness model (DT or RENOIR) makes a significant difference for mean inner potentials, but less so for mean free paths. Exact determination of the mean free paths might be more difficult, due to asymptotic

behavior, but the mean inner potentials could, in principle, be determined more easily, given additional information.

However, all of the analysis in the synthetic-wedge and maximum-thickness sections of this chapter is dependent on the amplitude-derived t/λ and phase data being linear with thickness. It was taken as axiomatic at the start of the synthetic-wedge section that the specimen orientations examined here are weakly diffracting. However, from the phase and amplitude as a function of specimen tilt angle plotted in Figure 4.4, it is also clear that many specimen orientations are not weakly diffracting. Therefore, to analyze diffraction effects on the measured phase and amplitude, the next section provides a systematic examination of the full tilt series of holography data: phase, amplitude, and amplitude-derived t/λ.

Experimental Tilt-Controlled Phase, Amplitude and t/λ Data

The full results from tilt-controlled holographic experiments are now examined in several different ways. As discussed above, the original holography data – phase, amplitude, and amplitude-derived t/λ – is a set of cross-sectional line profiles across the nanowire. When combined with specimen thickness information, these line profiles become phase, amplitude and amplitude-derived t/λ as a function of specimen thickness. Finally, by taking these line profiles from a series of holograms from systematically acquired tilt angles, the same cross-section can be examined at different specimen orientation. This leaves a quantity (*e.g.* phase) that can be examined as a function of thickness and/or specimen orientation. The previous sections only analyze weakly diffracting conditions – averaging together a small subset of the tilt series – and plot the holography data (phase, amplitude, and amplitude-derived t/λ) as a function of thickness.

In this section, the full tilt series of holograms is analyzed. The holography data (phase, amplitude, and amplitude-derived t/λ) are examined as a function of tilt angle first in the maximum-thickness region (where averaging will allow for the best signal-to-noise ratio), then in selected non-maximum-thickness regions, generated by the synthetic-wedge method. Then, the full tilt-controlled synthetic-wedge is examined, which plots the holography data as a function of both thickness and tilt angle. The role for simulations is then clarified, and these simulation results are detailed more clearly in the section after.

Tilt-Controlled Maximum-Thickness Measurements

In this section, the phase (charge-compensated), amplitude, and amplitude-derived t/λ data from the tilt series are examined in the maximum-thickness "plateau" region of the nanowire as a function of tilt angle. First, the analysis from earlier in this chapter that led to the identification of weakly diffracting conditions is recalled, and then the correlation between diffraction effects in the amplitude and phase is discussed. Several features arising from diffraction effects are then discussed. Note that, in this section, the phase, amplitude, and amplitude-derived t/λ are examined without determining the magnitude of the specimen thickness.

Phase and amplitude plots at the maximum thickness are now examined. In Figure 4.4, the amplitude and phase at the maximum thickness as a function of tilt angle are plotted. The phase in Figure 4.4 is not corrected for charging effects; Figure 4.20 shows the phase as a function of tilt angle with both surface and volume charge correction.

The starting point for the analysis of Figures 4.4 and 4.20 is to recall the earlier discussion of weakly diffracting conditions. As discussed earlier in this chapter, there is clear evidence of diffraction effects in the maximum-thickness graphs as a function of tilt angle in Figures 4.4 and 4.20. The expected variation in amplitude and phase with specimen tilt angle, if there were no diffraction effects, would be smooth. Also as discussed earlier in this chapter, the "weakly diffracting" region considered in the previous sections spans the angular range from +2° to +3°, and is, indeed, relatively invariant with tilt angle. However, there are large variations in amplitude and phase as a function of tilt angle. These, neglected before, are now discussed.

Another clear aspect of Figures 4.4 is the correlation between amplitude and phase. In Figure 4.4, it is at the same specimen tilt angles that both the amplitude and phase deviate from the weakly diffracting condition. This indicates that modeling this behavior using simulations would require a match between simulation and experiment for both amplitude and phase. Additionally, while the amplitude variations appear relatively small, as discussed in Chapter 2, the amplitude attenuation is exponentially dependent on thickness, meaning that small shifts due to noise are interpreted as relatively drastic changes in mean free path.

There is also symmetry visible in Figures 4.4 and 4.20, especially in the phase as a function of tilt angle. The phase graph in Figure 4.4 appears symmetric about approximately +0.5°. There are local minima at ±0° and +1°, a plateau between them and local extrema around +1.5° and -0.5°. There are also local minima around +4° and -3° starting U-shaped plateaus that end at +5° and -4.5°. Comparing these orientations with the Kikuchi-band diffraction features in Figure 4.3, where most of these locations would be at or near Kikuchi bands, gives another strong indication that diffraction effects are visible in Figure 4.4 and 4.20. Features in the amplitude are also visible in Figure 4.4 that are roughly symmetric about +0.5°, reinforcing the idea that features are present in both the amplitude and phase as a function of tilt angle. One consequence of this is that the symmetry of the pattern decreases the likelihood of specimen drift during acquisition and reinforces the orientation determination and relation to simulations.

One feature in particular demonstrates why automatic phase unwrapping would be unsuitable for this specimen. In Figures 4.4 and 4.20, the phase as a function of tilt angle undergoes phase shifts on the order of 2π from the local minima to the immediately adjacent maxima. These phase shifts were checked manually by comparing with hologram fringe shift in the original holograms. However, in a few cases (notably at -0.6°, but also at +3.8°), the original holograms were difficult enough to interpret due to high amplitude attenuation that even the manual phase-unwrapping result disagreed for different line profiles. Given that 2π phase shifts between images in two adjacent tilt angles are part of the data, not comparing with the original holograms could lead to misinterpretation. Automatic phase unwrapping routines in the ASUHolography package did not verify that the unwrapped phase agreed with the original images. Verifying the automatic phase unwrapping would have required the same fringe-comparison operations as it took to unwrap the phase manually.

Figure 4.20 shows relatively few differences between the two charging models. In Figure 4.20, the grey lines show that the effect of using either surface- or volume-charge as compared to no charging model is ~0.3 radians. Figure 4.20 also shows that

the difference between the surface and the volume charging models is less than the noise level. Therefore, while compensating for the charging effect is required, it is not possible to determine which model is a better representation of the specimen's physical charge distribution.

The amplitude-derived t/λ graphed as a function of specimen tilt angle in Figure 4.21 provide a useful counterpoint to the phase measurements in Figures 4.4 and 4.20. In Figure 4.21, diffraction-contrast effects occur around the same specimen tilt angles where there are deviations from the "weakly diffracting" phase value. While the diffraction-contrast effect on the phase in Figure 4.4 is both positive and negative, the diffraction-contrast effect on the amplitude-derived t/λ in Figure 4.21 is generally towards increasing t/λ, meaning increased amplitude attenuation. These diffraction-contrast effects are occasionally dramatic: Figure 4.21 shows peaks up to almost 7 t/λ, as compared to ~3 t/λ in the weakly diffracting condition. The amplitude-derived t/λ as a function of tilt angle highlights the presence of diffraction effects: the phase in Figure 4.4 varies both up and down, while the amplitude-derived t/λ in Figure 4.21 generally only increases.

Therefore, it can be said that, at the maximum specimen thicknesses examined here, Figure 4.4, Figure 4.20, and Figure 4.21 show multiple unambiguous signs of diffraction contrast in phase, amplitude, and amplitude-derived t/λ. In the next section, tilt-controlled synthetic-wedges are created, analogous to the weakly diffracting condition synthetic-wedges generated earlier.

Tilt-Controlled Synthetic-Wedges

This section generates tilt-controlled synthetic-wedges from the phase, amplitude, and amplitude-derived t/λ data, and briefly discusses the effect of DT shell compensation on the display of RENOIR data.

The procedure in this section is fundamentally the same one used in previous sections for generating "weakly-diffracting-condition" synthetic-wedges, but with the additional effect of specimen tilt angle. First, the thickness profile determined using either thickness-determination method (DT or RENOIR) is applied to each cross-sectional profile at all tilt angles. Then, the profiles are reordered from smallest to largest thickness and linearly interpolated in the thickness direction to a resolution of 0.25 nm into an equally-spaced grid to form a synthetic-wedge greyscale image, where one spatial dimension ("*x*") is tilt angle, another ("*y*") is thickness, and the grey value is the parameter (phase, amplitude, or amplitude-derived t/λ). The RENOIR wedges, like the DT wedges, will be presented both with and without shell compensation. The magnitude of the shell compensation used is calculated from the data discussed earlier at the weakly diffracting condition, because the weakly-diffracting-condition data is less noisy due to its angular average, and the shell, if it is amorphous, should not show orientation-dependent diffraction effects. However, when the shell compensation is applied, the RENOIR wedges are truncated at the DT maximum thickness. Therefore, the RENOIR raw-synthetic-wedges will be slightly longer in the *y*-direction than either the DT raw-synthetic-wedges or the DT or RENOIR shell-compensated synthetic-wedges; the latter three types of synthetic-wedge will have the same dimensions.

In the next section, thinner regions of the specimen are now examined to see if the diffraction effects detailed in this section persists at lower thicknesses, and to

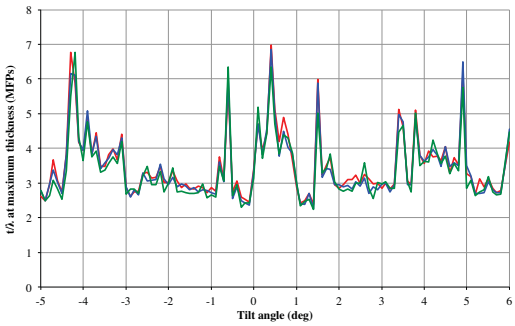


Figure 4.21: Angular dependence of amplitude-derived t/λ at maximum thickness. This plots the amplitude-derived t/λ in the maximum-thickness region as a function of tilt angle across the tilt series for all three profiles (blue, green, red) used elsewhere in this work. The amplitude shows shifts as a function of tilt angle, indicating that orientation effects seen in the phase are not restricted to only the phase.

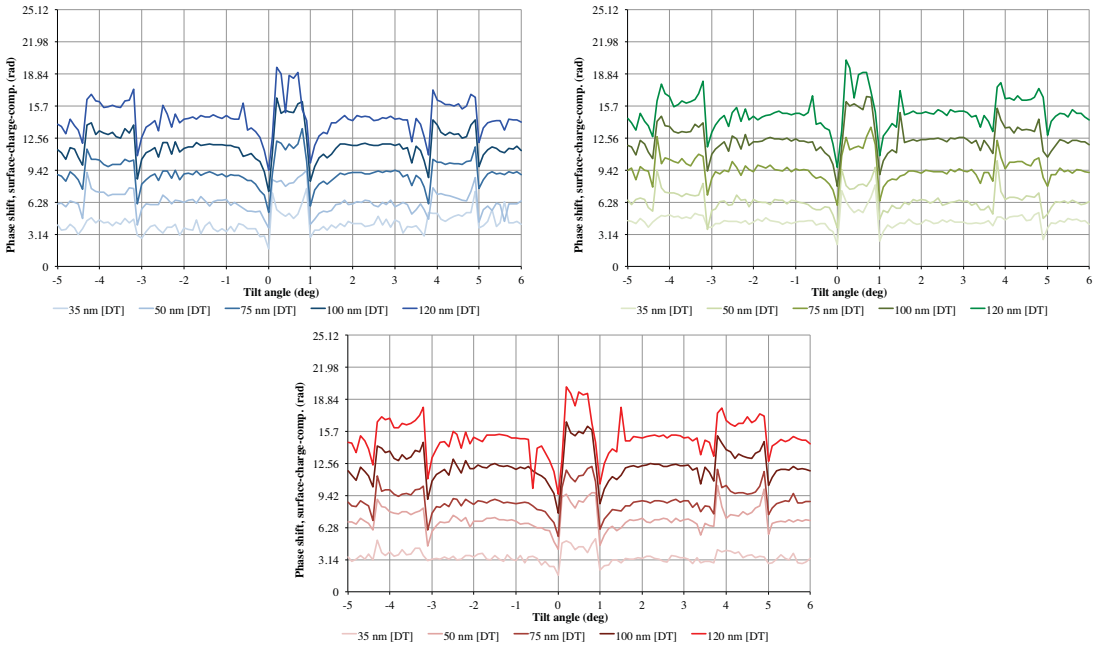


Figure 4.22: Phase as a function of tilt angle, selected thicknesses from a synthetic wedge created using DT thickness information, surface-charge compensation. These extracts from the synthetic wedges – at 35 nm, 50 nm, 75 nm, 100 nm, and 125 nm - show that the differences between the strongly and weakly diffracting conditions are present at all observed thicknesses. There are slight differences at some lower thicknesses (*e.g.* red, 35 nm), but the profiles agree with each other more at higher thicknesses.

investigate if there is a critical onset thickness for diffraction effects. The greyscale synthetic-wedges themselves are displayed and discussed in their own section after.

Tilt-Controlled Selected-Thickness Measurements

This section examines if the diffraction contrast seen in the maximum-thickness regions of the nanowire is also present in the thinner regions of the nanowire. This section examines phase, amplitude, and amplitude-derived t/λ as a function of specimen tilt angle at different thicknesses. However, to examine regions of selected thickness, a thickness model for the nanowire cross-section must be applied. This section uses only the DT thickness model with DT shell compensation, because the same thickness-onset details would be present for DT and RENOIR.

The selected-thickness plots each plot one holography parameter as a function of tilt angle. The plots are seen in Figures 4.22 (surface-charge-compensated phase), 4.23 (volume-charge-compensated phase), 4.24 (amplitude), and 4.25 (amplitude-derived t/λ), all as a function of tilt angle. The thicknesses used are 35-36 nm, 50-51 nm, 75-76 nm, 100-101 nm, and 120-121 nm, as determined using DT. Therefore, each graph plots the holography parameter against tilt angle at fixed thickness. They provide a relatively representative sampling of thicknesses from the minimum clearly resolvable thickness (35 nm) to a high non-maximum thickness (120 nm) with three middle thickness values (50 nm, 75 nm, 100 nm) for clarity. Charge compensation models are calculated for each tilt angle separately.

The selected-thickness results in Figures 4.22, 4.23, 4.24, and 4.25 provide clear evidence that the diffraction effects persist even at the lowest measurable thicknesses. At the lowest thicknesses, the phase as a function of tilt angle – shown in Figure 4.22 for surface-charge compensation and Figure 4.23 for volume-charge compensation – clearly shows diffraction effects at low thicknesses. The phase shows more diffraction effects than the amplitude (Figure 4.24) and amplitude-derived t/λ (Figure 4.25) information, but the middle feature from +0.0° to +1.0° is especially clear on the amplitude and amplitude-derived t/λ. None of the selected-thickness graphs is as clear as the corresponding spatially averaged data from the previous maximum-thickness section (*e.g.* Figure 4.20 compared to Figure 4.22), due both to the decreased noise after averaging, and the increased noise from applying the thickness cross-sectional information.

Some differences occur between individual profiles, especially visible in the phase profiles (Figures 4.22 and 4.23): in the 35 nm phase data, the green profile disagrees with the red and blue profiles, and, in the 35 nm amplitude data (Figure 4.24), the red profile disagrees with the green and blue profiles. However, while one profile might disagree with the other two in the thin regions, the line profiles in the thicker regions are generally closer to each other, with some noise from the "left-side/right-side" effect discussed earlier and visible in Figure 4.7 and 4.8. The magnitude of this diffraction contrast is thickness dependent: comparing the phase profiles in Figure 4.22 at 50 nm vs. 120 nm, the difference between the plateau at +0.5° and the flat region at +2.5° is $0<\pi$ at 50nm, and $>\pi$ at 120 nm, and it is unclear from this information if this is a linear dependence. Therefore, simulations are needed to fully understand this effect.

This section discusses selected-thickness cross-sections of the synthetic-wedges. In the next section, the full tilt-controlled synthetic-wedges are analyzed.

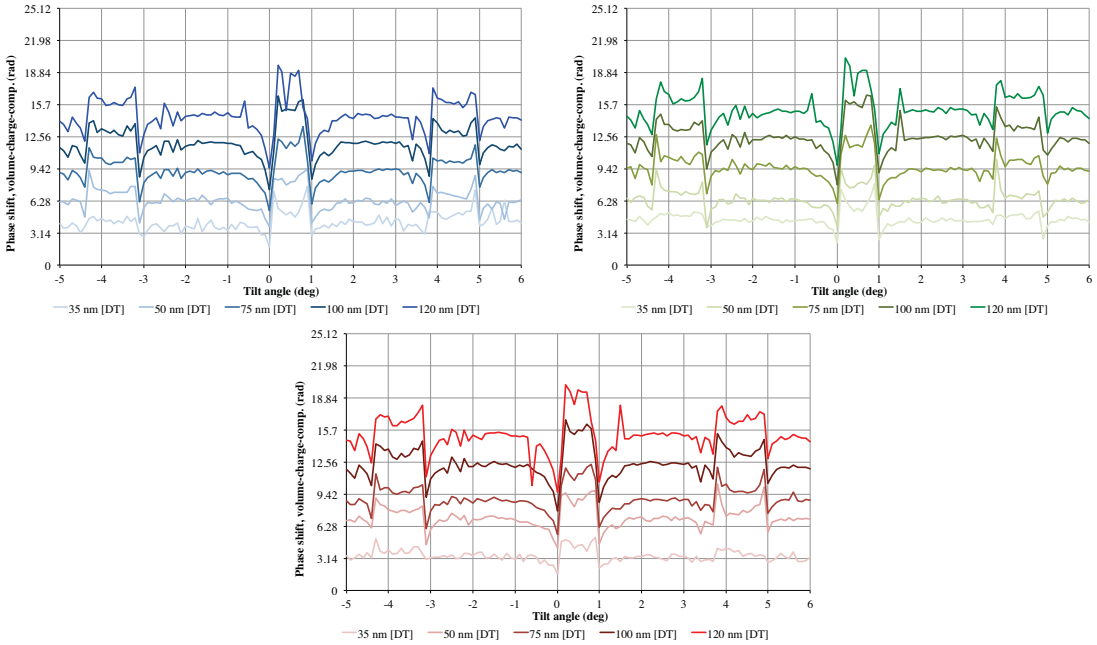


Figure 4.23: Phase as a function of tilt angle, selected thicknesses from a synthetic wedge created using DT thickness information, volume-charge compensation. The minimal differences between this figure and Figure 4.22 show the infeasibility of distinguishing between the two models for charging used in this work.

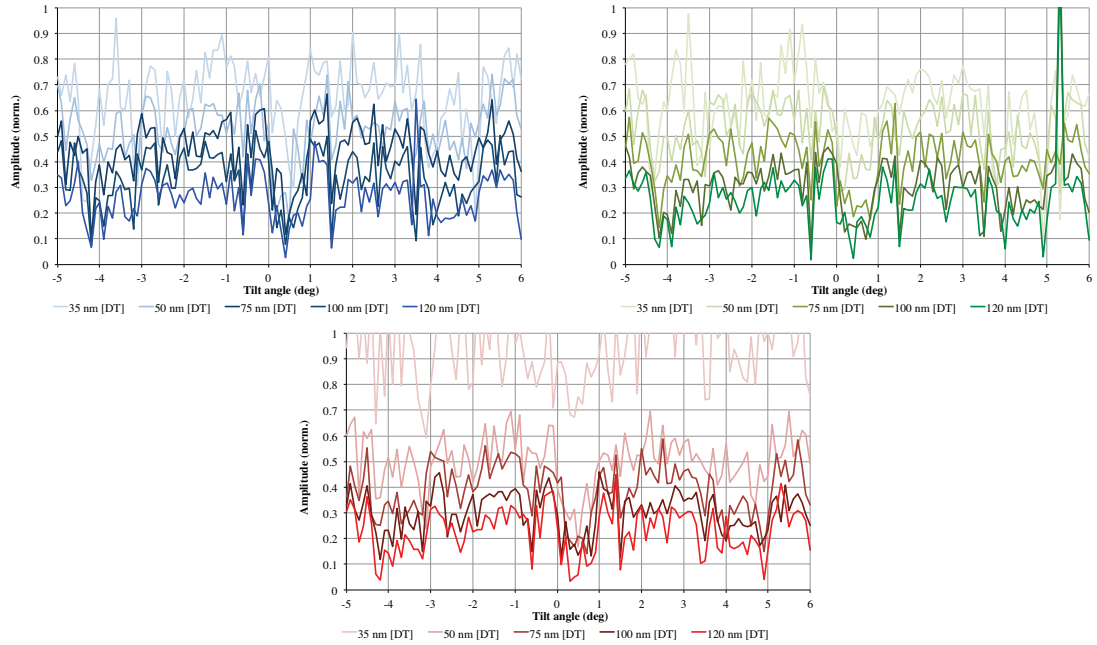


Figure 4.24: Amplitude as a function of tilt angle, selected thicknesses from a synthetic wedge created using DT thickness information. The amplitude is much noisier than the phase seen in Figures 4.22 and 4.23. There are differences between the red profiles and the other two colors at low thicknesses, but these differences do not appear to persist at higher thicknesses. Despite the noise, there are clear signs of diffraction effects. Note these are from the single-pixel profiles discussed earlier.

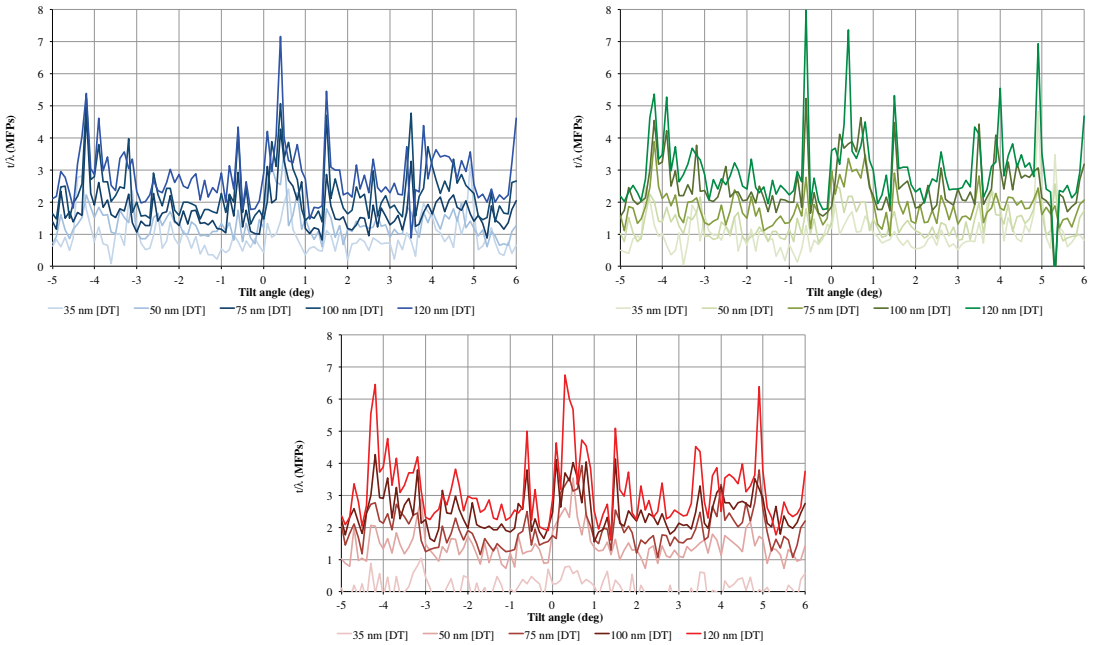


Figure 4.25: Amplitude-derived t/λ as a function of tilt angle, selected thicknesses from a synthetic wedge created using DT thickness information. This, like its counterparts in Figures 4.22, 4.23, and 4.24, shows the strong diffraction effects that are possible in the electron holograms across the tilt series and at both low and high thicknesses.

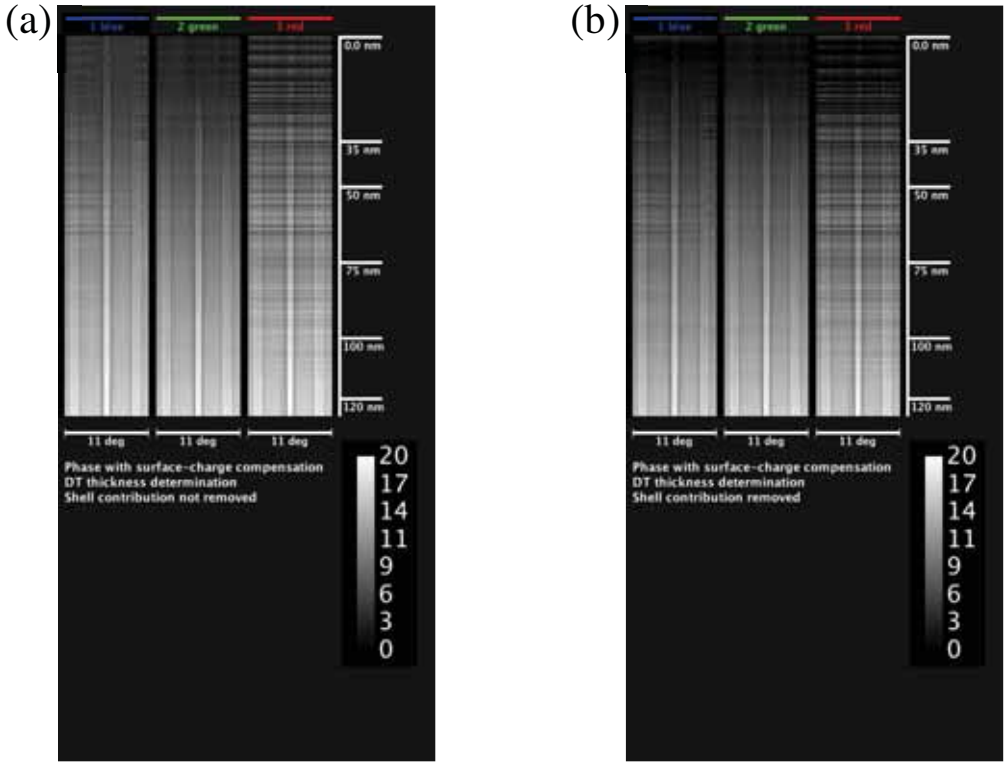


Figure 4.26: DT synthetic wedges for phase with surface-charge compensation with and without shell compensation. These are synthetic wedges, where the x-coordinate corresponds to the tilt (tick=0.1°), and y-coordinate to thickness (tick=1 nm). The data in Figure 4.22 is taken from the marked locations in the shell-removed image. The phase is measured here in radians and plotted in greyscale (key on right). The banding in *e.g.* the red profiles is due to the two-side effect detailed in Figure 4.7.

Analysis of the Full Tilt-Controlled Synthetic-Wedge

In this section, the full tilt-controlled synthetic-wedges are analyzed. The full tilt-controlled synthetic-wedges are examined both with and without shell corrections, and applying both DT and RENOIR thickness determination methods. The parameters examined are, as above, the phase with surface- and volume-charge corrections, the amplitude, and the amplitude-derived t/λ. The extent of dynamical diffraction, the broader trends in the tilt-controlled synthetic-wedge data, and the limiting factors in this determination are all detailed in this section, leading up to discussing the requirements from image simulations to correspond with this data.

First, the tilt-controlled synthetic-wedges from the experimental data all have the same form, and this form should be explained. The first synthetic-wedge figure is used as an example – the phase with surface-charge compensation and DT thickness, seen in Figure 4.26. Each of the tilt-controlled synthetic-wedge figures has two images; the image on the left (*e.g.* Figure 4.26a) does not have compensation for the effects of the nanowire shell, and the image on the right (*e.g.* Figure 4.26b) includes shell compensation. In each black-background image (*e.g.* Figure 4.26a), there are three similar greyscale images, with "blue", "green", or "red" indicated above them. These greyscale images are each a synthetic-wedge. Each synthetic-wedge corresponds to either the blue, green, or red profiles, using the color-coding established as conventional in Chapters 3 and 4. To the right is a labeled thickness axis, with markers indicating 0 nm thickness, 35 nm, 50 nm, 75 nm, 100 nm, and 120 nm. The bottom of the DT profile is 125.5 nm. The markers are each 1 nm (4 pixels) wide. Along the bottom of each synthetic-wedge is the indication of their angular span – 11° in total. The tick marks on the side are 0.1° (1 pixel) wide. Therefore, each pixel in the synthetic-wedge has a size of 0.1° in the *x*-direction and 0.25 nm in the *y*-direction. The intensity scale in the lower right indicates the grey-value calibration in each black-background image. The same grey-value calibration is used for the synthetic-wedges of each kind. For the phase synthetic-wedges, in radians; for the amplitude synthetic-wedges, normalized between 0 and 1; and for the t/λ synthetic-wedges, between 0 and 8 t/λ. A brief label detailing the key processing characteristics in the lower left of each image.

The DT wedges are seen in Figures 4.26, 4.27, 4.28, and 4.29, and the RENOIR wedges in Figures 4.30, 4.31, 4.32, and 4.33. Each of these figures has two black-background images like this – the left image without shell-compensation, and the right with shell-compensation using the DT shell profile and the magnitudes calculated from the weakly diffracting condition, as detailed earlier. Note that the RENOIR wedges in Figures 4.30, 4.31, 4.32, and 4.33 are limited to the DT maximum thickness only for the shell-compensated synthetic-wedges.

Now, the features of the synthetic-wedges are examined, starting with the synthetic-wedges using the DT thickness determination. The phase with surface-charge correction is in Figure 4.26 and with volume correction is in 4.27. Several features are notable. The phase shows clear diffraction contrast at all thicknesses, although some of the wedges are noisier than others. The horizontal stripes, especially visible in the "red" synthetic-wedge in Figure 4.26 and Figure 4.27, is due to the "left-side/right-side" miscalibration effect discussed earlier in this chapter. However, this horizontal striping does not obstruct the general conclusions, and is a general feature of the synthetic-wedges, and so will not be further noted unless it blocks critical analytical features – although, as noted, this striping likely does play a role in the single-

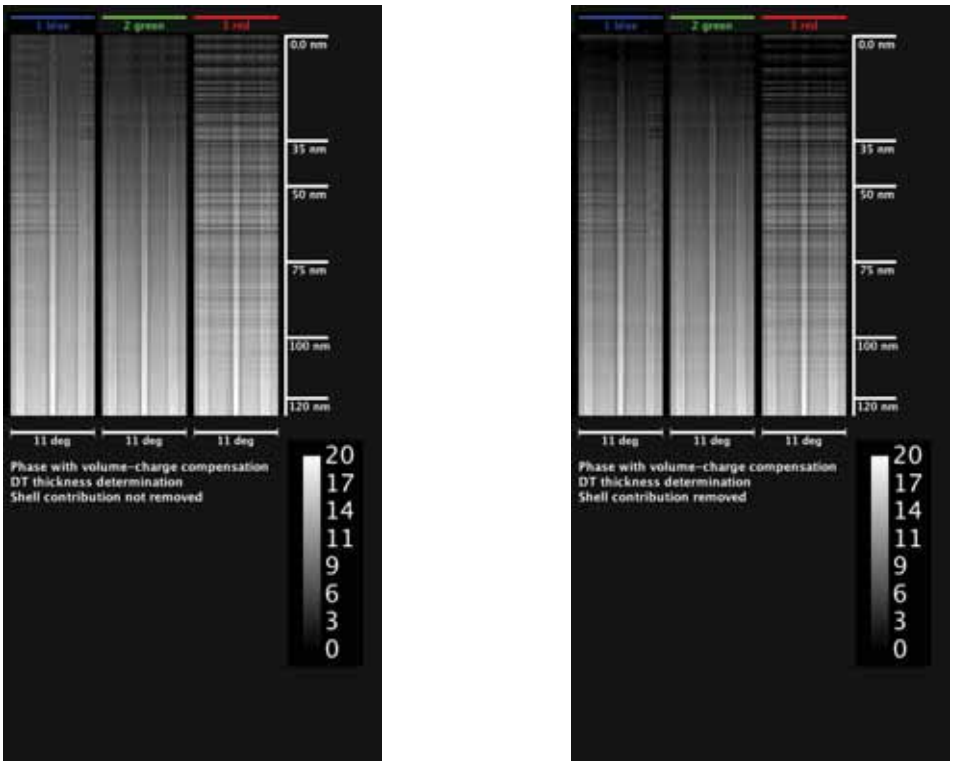


Figure 4.27: DT synthetic wedges for phase with volume-charge compensation with and without shell. The archetype is Figure 4.26. This shows the same effects as Figure 4.26, and the two charge models are not distinguishable. The effect of the shell is visible, especially at low thicknesses.

Figure 4.28: DT synthetic wedges for amplitude with and without shell. The archetype is Figure 4.26. The shell effects are especially visible, as is the strong diffraction contrast. Intensity is normalized to vacuum level.

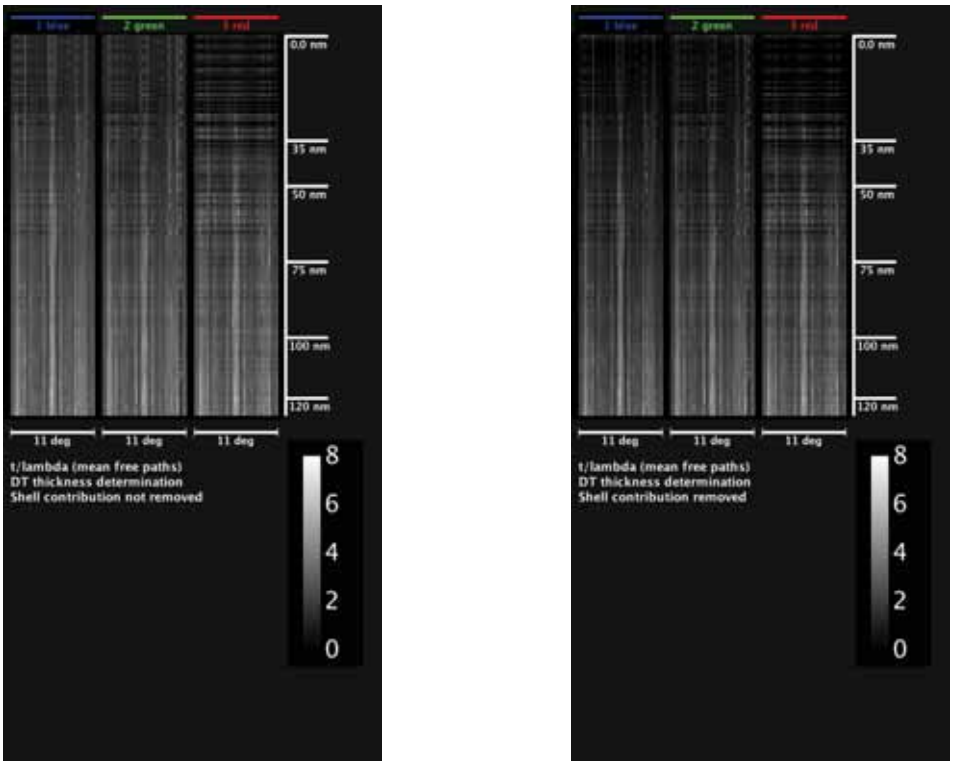


Figure 4.29: DT synthetic wedges for amplitude-derived t/λ with and without shell. The archetype is Figure 4.26. Diffraction effects are especially visible, and persistent across even the lowest thicknesses. Intensity is in mean-free-paths.

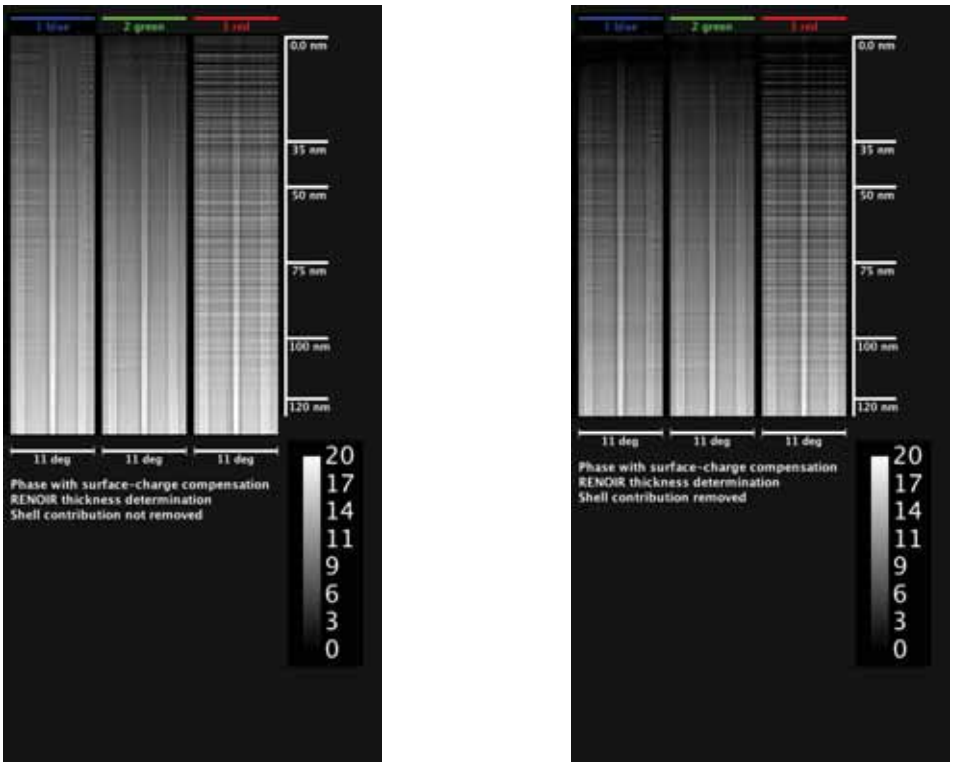


Figure 4.30: RENOIR synthetic wedges for phase with surface-charge compensation with and without shell. The archetype is Figure 4.26. The shell-compensated thickness is limited to the DT maximum thickness (see figure 4.10). Otherwise, the features are very similar to those in Figure 4.26.

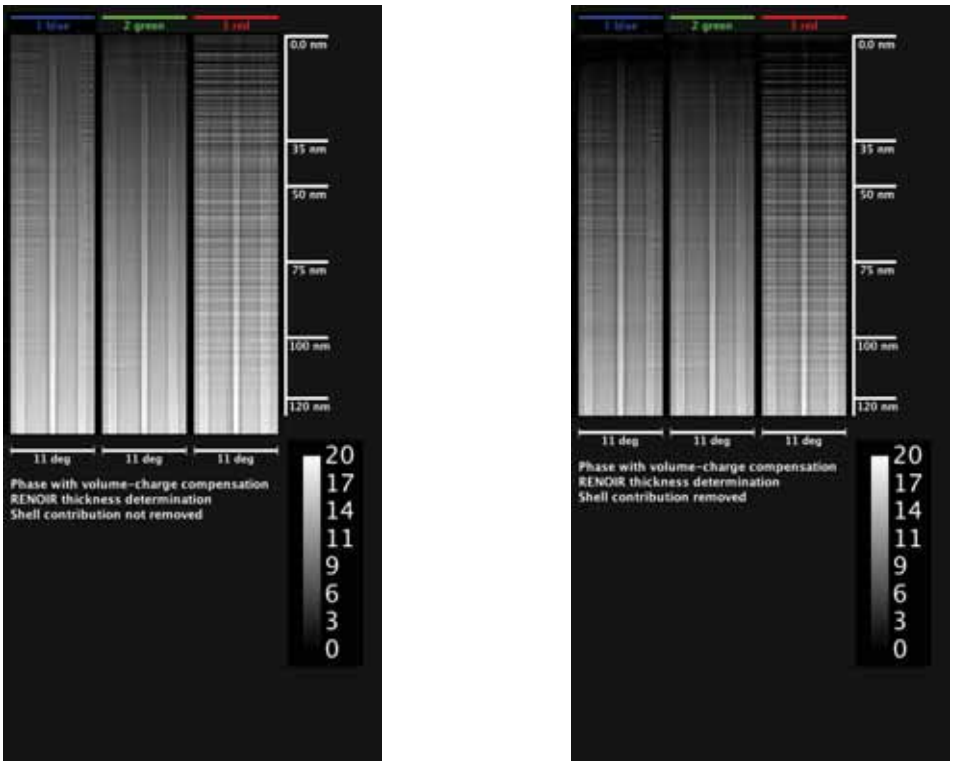


Figure 4.31: RENOIR synthetic wedges for phase with volume-charge compensation with and without shell. The archetype is Figure 4.26. Compared to Figure 4.27, the key difference is the maximum thickness change due to RENOIR.

Figure 4.32: RENOIR synthetic wedges for amplitude with and without shell. The archetype is Figure 4.26. As with Figure 4.31, the key difference when compared to the DT counterpart (Figure 4.28) is the maximum thickness change due to RENOIR thickness determination.

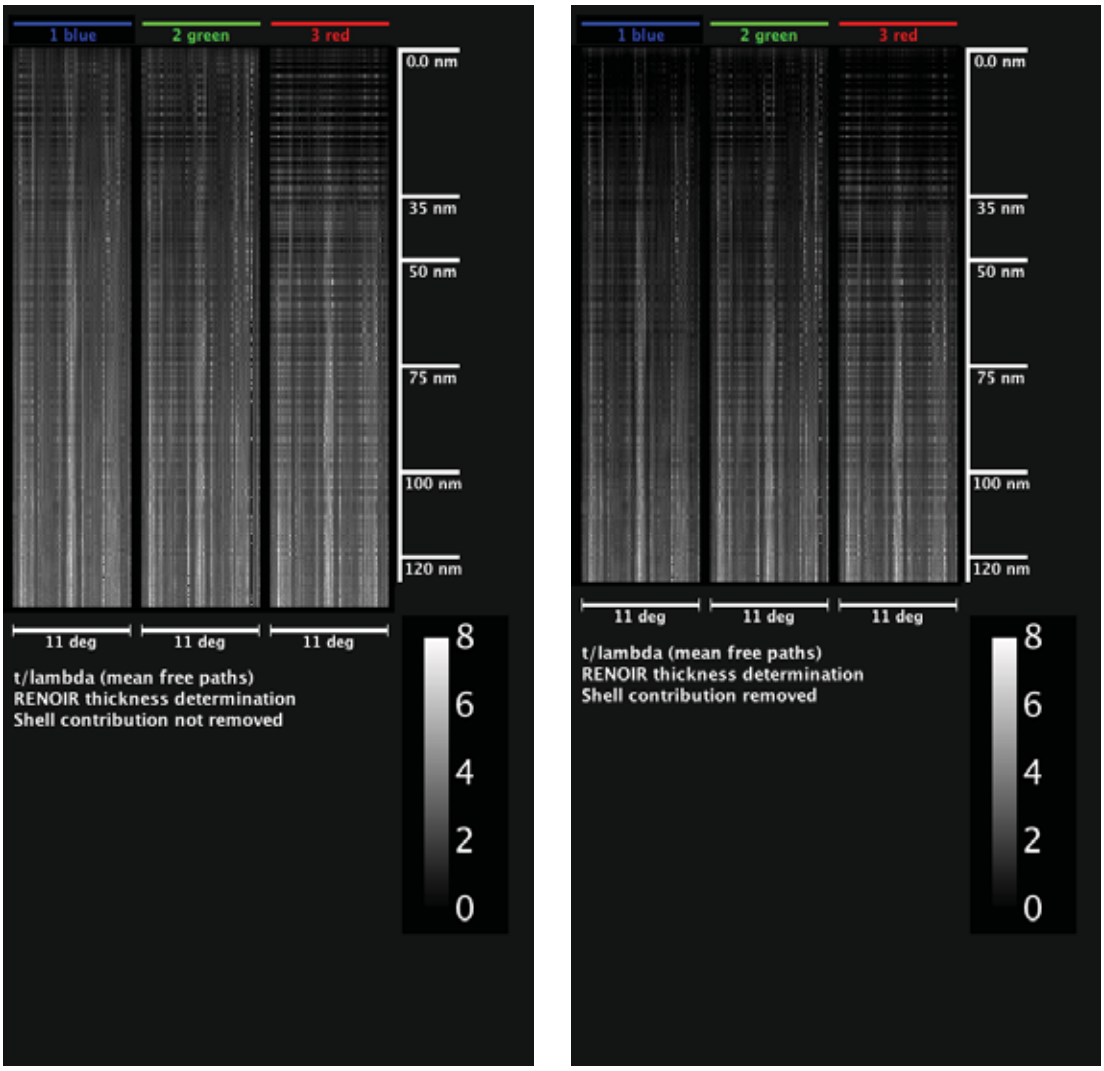


Figure 4.33: RENOIR synthetic wedges for amplitude-derived t/λ with and without shell. The archetype is Figure 4.26. As with Figure 4.31, the key difference when compared to the DT counterpart (Figure 4.29) is the maximum thickness change due to RENOIR thickness determination. Diffraction effects are clearly visible.

thickness measurements in the previous section. Second, in the phase synthetic-wedges (Figure 4.26 and Figure 4.27), the diffraction contrast features are strongly evident – at all thicknesses in the wedge, the feature in the middle of the tilt range is clearly seen and the two side features are usually seen. This indicates that the diffraction contrast features are, indeed, consistent for each individual tilt angle relative to others. Third, in Figure 4.26 and Figure 4.27, the two narrow peaks on either side of the central band are occasionally difficult to resolve, likely due to the strong amplitude attenuation and high noise levels at these orientations. In Figure 4.26b and Figure 4.27b, the removal of the shell contribution has the most obvious effect on very low thicknesses (<35 nm), but this is not an interpretable region of the graph. Visually, the effect of the shell contribution removal is minimal – this does not mean the shell contribution itself is minimal, given its effect on quantification but because this linearity is not clearly visible on a greyscale image.

The amplitude and amplitude-derived t/λ DT synthetic-wedges are seen in Figures 4.28 and 4.29. These show the same diffraction contrast trends as the phase in Figures 4.26 and Figure 4.27: there is a clear middle diffraction feature and two side features, present along the entire thickness examined here. The amplitude synthetic-wedges in Figure 4.28 show attenuation increasing with thickness, and the attenuation increases further in the strongly diffracting conditions in the middle feature and the two side features. Accordingly, the amplitude-derived t/λ synthetic-wedges in Figure 4.29 shows increased t/λ in the middle feature and the two side features. However, the amplitude and t/λ information in Figure 4.28 and Figure 4.29 is much noisier than the relatively smooth phase information in Figure 4.26 and Figure 4.27. Comparing Figures 4.28b to 4.28a and Figure 4.29b with 4.29a, the shell compensation appears more visually detectable in the amplitude and amplitude-derived t/λ than it does in the phase. This reinforces the importance of the shell compensation.

The RENOIR wedges, in Figures 4.30, 4.31, 4.32, and 4.33, tell the same story as the DT wedges. Visually, very little distinguishes the RENOIR synthetic-wedges from the DT synthetic-wedges. As previously discussed, the RENOIR synthetic-wedges without the shell removed are shown as longer than the shell-compensated ones, due to their higher maximum thickness; the thickness-axis bar on the right side of these images is deliberately not lengthened, to highlight the difference. The key distinctions – the diffraction contrast and its differences between phase, amplitude, and amplitude-derived t/λ – are all equally seen in RENOIR as in DT. The critical difference between RENOIR and DT, then, is not the general interrelationship of the phase, amplitude, and amplitude-derived t/λ, but in quantification.

However, to understand the diffraction-contrast effects seen in these synthetic-wedge images, experimental holography data must be compared with simulated data.

Relating Experiment to Dynamical Diffraction Simulations

This section briefly details how image simulations have been carried out to understand the experimental results from electron holography performed in this dissertation. The primary focus of this section is not to delve deeply into the simulations themselves, but to detail how they are a necessary and possible adjunct to this genre of experiment, to discuss how to use them to interpret the experimental data.

The multislice image simulations were carried out by Rafal E. Dunin-Borkowski for specified orientations. According to him, improved multislice propagators were applied to correctly represent phase as measured using electron holography for tilted crystals. The necessary inputs and assumptions from experiment were the chemical composition and the crystallographic parameters for the wurtzite InAs core [98], and the region of reciprocal space nominally spanned by the holography tilt series at 120kV primary accelerating voltage, determined above. Debye-Waller factors for In (0.009461 nm^2) and As (0.006412 nm^2) were provided by Chris Boothroyd. Absorption was modeled with a frequency-invariant amplitude-attenuation absorption parameter a, and several absorption values were examined (a=[0,0.05,0.1,0.15,0.2]). The shell surrounding the wire was not modeled.

The image simulations were performed with a thickness resolution of 0.742 nm/pixel and a tilt resolution of 0.0113°. The simulated thickness used went from 0 nm up to a maximum thickness of 150.63 nm. Both 2-D and 3-D simulation potentials were examined, but only the more physically representative 3-D potentials are presented here.

The shell around the InAs nanowire core was not a part of the simulation, but should or could potentially be a part of future simulations. This also means that all of the experimental data that could be compared with simulations must have the shell contribution removed.

Methods of Comparison Between Simulation and Experiment

In this section, methods of comparing the simulated phase and amplitude data with experimental phase and amplitude data are discussed. While the phase-derived mean inner potential and amplitude-derived mean free path are the focus of the experimental analysis, the simulations provide direct phase and amplitude data.

The mean inner potential is modeled in the simulation as a simple multiplier to the phase. A simulation with a different mean inner potential would simply have all of the phase values offset by this multiplicative factor. Therefore, in principle, the mean inner potential could be determined from the simulations by fitting a multiplicative factor to the simulated phase. The mean inner potential used in the simulations themselves is that from isolated-atom scattering factors [66], about which much more will be said in Chapter 5.

In contrast to the single-phase values, a range of amplitudes must be examined, due to the absorption input parameter. The absorption parameter must be given as input to each individual simulation, and, as will be discussed later in this chapter, the phase results are sensitive to this absorption parameter such that the amplitude results of the simulation cannot be easily tweaked by applying a multiplier. The most useful approach is direct comparison between the simulated amplitude results from individual absorption parameters and the experimental amplitudes, bypassing the mean free path.

With this in mind, the discussion moves to the outcome of comparison between image simulation and experimental holography data.

Comparison Between Simulation and Experiment

In this section, the image simulations are presented and compared with experimental holography data. To compare experimental holography data with image simulation, in

concordance with the work in the previous sections, the tilt calibration between the simulated and experimental data is first examined, followed by several points of comparison: the weakly diffracting synthetic-wedge, a strongly-diffracting synthetic-wedge, and finally the single-thickness measurements from non-maximal thicknesses. Experimental phase data is compared with simulated phase data, but full analysis is not possible.

Global Factors Affecting Comparison

The simulations are most easily visualized by using the synthetic-wedge approach detailed earlier (*e.g.* Figure 4.26). Representative amplitude and unwrapped phase simulations are seen in Figure 4.34, using an absorption parameter a of a=0.05 in 4.34a (phase) and 4.34b (amplitude), and a=0.1 in 4.34c (phase) and 4.34d (amplitude).

The simulated synthetic-wedge data in Figure 4.34 do not appear to be very similar to the synthetic-wedge data in *e.g.* Figure 4.26, but this is a result of the difference in visual scaling. Comparing Figure 4.34 to *e.g.* Figure 4.26, it is intuitively obvious that the visual scales are different: Figure 4.34 is "wide", while Figure 4.26 is "tall". This is because the pixels in the synthetic-wedge in Figure 4.26 are 9 times more coarsely sampled in the angular direction, and 3 times more finely sampled in the thickness direction. Figure 4.34 uses the same scale-marks as the synthetic-wedges in *e.g.* Figure 4.26: the end-marks on the tilt axis are 0.1° (9 pixels), and the width of the thickness markings is 1 pixel = 0.74 nm, which is as close as possible to 1 nm without resampling. The specimen thickness sampling difference is due to the resampling process applied to the experimental data to generate the experimental synthetic-wedges. However, the dramatic difference in tilt sampling between the experimental and simulated data plays a role in the following sections.

Resampling the angular resolution of the simulated synthetic-wedge in Figure 4.34 to the angular resolution of the experimental synthetic-wedges implies an additional assumption. The simulated data needs to be downsampled from 0.0113°/point to 0.1°/point. Resampling the data in Figure 4.34 along the tilt axis will create an average of the points. However, some of the features in Figure 4.34 are very sensitive to tilt angle, especially in the strongly diffracting regions, so averaging will smooth out these extrema. Unfortunately, the experimental data is taken from specific specimen orientations, not averaged ones – it is the *location* of that tilt angle which is somewhat uncertain, not its angular spread. One example of a rapidly varying feature is the narrow streak left of the middle in the experimental phase synthetic-wedge (Figure 4.26), which appears to be present only at a limited tilt range in the simulated synthetic-wedge (Figure 4.34), smaller in width than the horizontal tick mark and would be smoothed out by the angular averaging. Therefore, as a result of the averaging process, features that are very sensitive to specimen orientation are not likely to be reproduced when the simulated synthetic-wedges are rescaled to the experimental data's synthetic-wedges. This is a critical concern for the amplitude data, which is rapidly varying in the simulated synthetic-wedge in Figure 4.34, and noisy in the experimental synthetic-wedge in Figure 4.28; therefore, interpretation of the amplitude data will be necessarily limited in strongly diffracting regions.

The next obvious feature of the simulated synthetic-wedges in Figure 4.34 as compared to the experimental synthetic-wedges is the mismatch in total angular range. The middle feature and the two side features from the experimental synthetic-

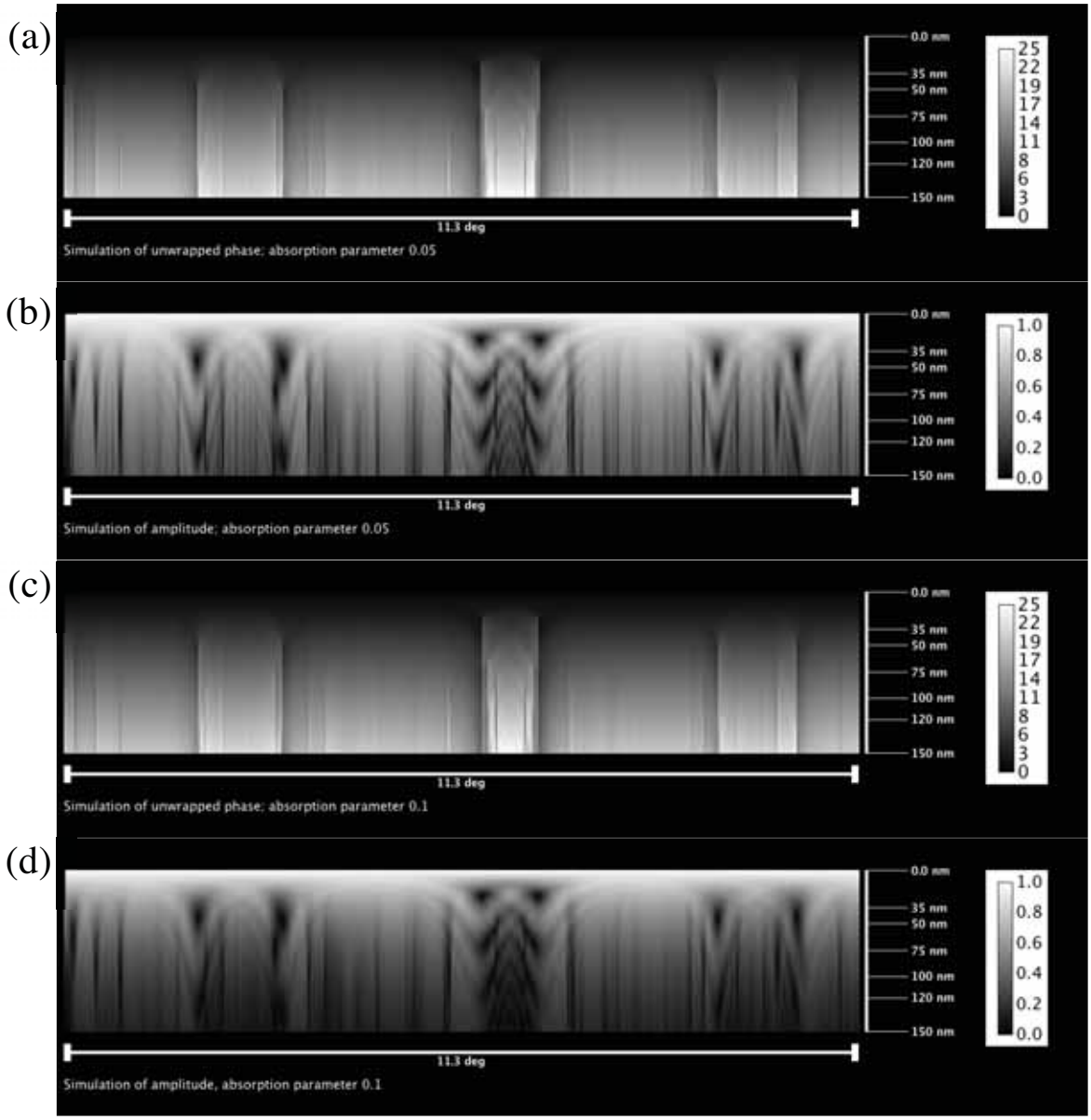


Figure 4.34: Simulations of unwrapped phase and amplitude in synthetic-wedge form. Simulations are courtesy of Rafal Dunin-Borkowski. Compare to Figures e.g. for DT thickness 4.26 and 4.27 for phase, 4.28 for amplitude which are for RENOIR thickness 4.30, 4.31 and 4.32, respectively. The tickmarks on the bars and conventions are the same as in Figure 4.26 – x ticks are 0.1°, y ticks are 1 nm. This shows the difference visually between the highly-spatially-sampled, poorly-angularly-sampled experimental synthetic wedges and vice versa for the simulated synthetic wedges, meaning that the features seen here would not be displayed in the same way in the experimental synthetic wedges.

wedge in *e.g.* Figure 4.26 are reproduced in the simulated synthetic-wedge in Figure 4.34, but the additional features on the left side of the simulated wedges in Figure 4.34 are not present in the experimental synthetic-wedges in Figure 4.26. This indicates that the angular calibration between the experimental images and the simulated ones is fundamentally different. Either the simulated synthetic-wedge or the experimental synthetic-wedge should be taken as a reference. This calibration is performed by scaling the low-thickness diffraction-effect features to match the size of the reference feature[28]. The features chosen for this are the farthest-apart minima in the two side features in the phase at 35 nm (see Figure 4.35 and later discussion). Using the simulated synthetic-wedge in Figure 4.34 as a reference, the experimental synthetic-wedge in Figure 4.26 would need to be scaled down by 9% - from 11.0° of total angular span to 10.0°. Alternatively, using the experimental synthetic-wedge as a reference, the simulated synthetic-wedge could be scaled from 11.3° of total angular span to 14.5°. The most likely explanations are either steady specimen drift during data acquisition, or that the tilt step during acquisition, instead of being the nominal 0.1°, was instead 0.091° on average. A less likely possibility is if the specimen orientation determination is incorrect, and the specimen orientations sampled experimentally are all a different distance from the nearby low-order zone axis. In any case, the source of this angular inaccuracy cannot be determined from the experimental data. The simulated synthetic-wedge is more likely accurate about the angular span than the experimental synthetic-wedge. However, this analysis will proceed by continuing to assume that the experimental angular span is accurate, and will scale the simulated data accordingly. The experimental angular span is used as a reference for continuity with the ongoing analysis, aiding comparison between simulated data in this section and experimental data in previous sections. Importantly, choosing the experimental angular span as the reference does not affect the numerical conclusions, and, for future work, it would be wisest to use the simulated tilt range as a reference for comparison between experimental and simulated holograms.

Two regions of the synthetic-wedges generated by the simulations in Figure 4.34 are not valid for the experimental data. The regions of the simulated synthetic-wedges with thicknesses above 130 nm and below ~25-35 nm are not present in the experimental specimen, as discussed earlier in this chapter.

Analyzing the Simulated Phase, Amplitude and Amplitude-Derived t/λ

Synthetic-wedges are then made from the simulated amplitude and phase, with the same angular calibration as the synthetic-wedges made from the experimental data, and the same angular and thickness sampling. The simulated synthetic-wedges for different simulated absorption parameters can be seen in Figure 4.36 for the phase, and Figure 4.37 for the amplitude and the amplitude-derived t/λ. These are on the same intensity scales as their experimental synthetic-wedge counterparts: 4.26, 4.27, 4.30, and 4.31 for the phase, 4.28 and 4.32 for the amplitude, and 4.29 and 4.33 for the amplitude-derived t/λ. While many of the simulated features in Figures 4.36 and 4.37 lie below the interpretable-thickness threshold, there are several points of comparison. The simulated phase has only one clear feature that varies with absorption parameter in the synthetic wedges: the width of the central plateau. This was used for angular-span calibration in the previous section. The synthetic wedges generated from the amplitude – and especially the amplitude-derived t/λ – varies

[28] Low-thickness phase features are chosen for the calibration, as discussed in the following sections, because they are less affected by absorption parameter than high-thickness phase features.

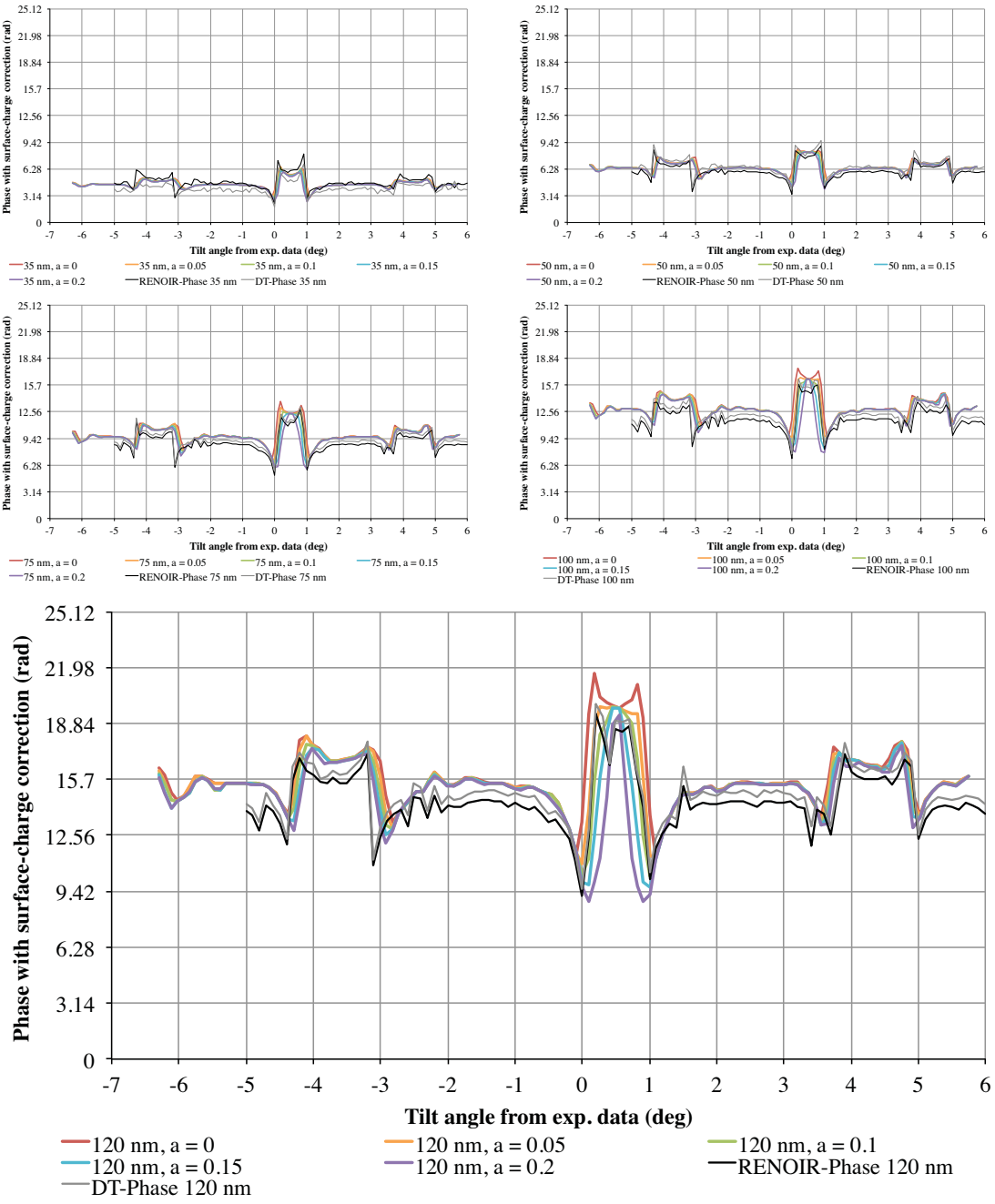


Figure 4.35: Comparison between simulation and experiment of phase as a function of tilt angle at selected thicknesses (surface-charge compensation). The features are reproduced well between experiment and simulation. The simulated data is angularly averaged over the tilt angles (because it is more finely sampled than the experimental data). In the 120nm-thickness graph, at bottom, the clearest effect of the absorption parameter *a* is to change the width of the diffraction features. Otherwise, there are few differences between different phase simulations at different absorption parameters. The angular span between the simulated and experimental data is calibrated with reference to the 35nm diffraction-contrast features, where the effects of the absorption parameter are minimized.

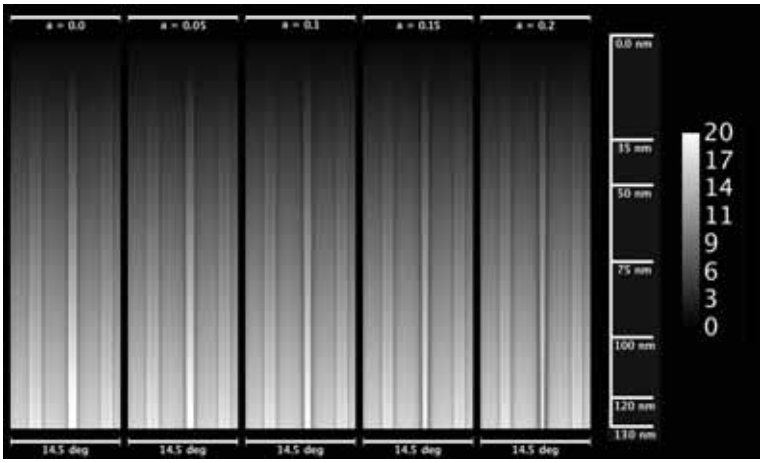


Figure 4.36: Simulated phase wedges, resampled for comparison with experimental data. By resampling the simulated phase data from Figure 4.34 – averaging down to an angular resolution of 0.1° per pixel and interpolating to a spatial resolution of 0.25 nm/pixel, the simulated phase wedges can be compared with the experimental data in Figures 4.26, 4.27, 4.30, and 4.31. Qualitatively, the absorption parameter values a=0.05 and a=0.1 appear closest to the experimental data in *e.g.* Figure 4.26.

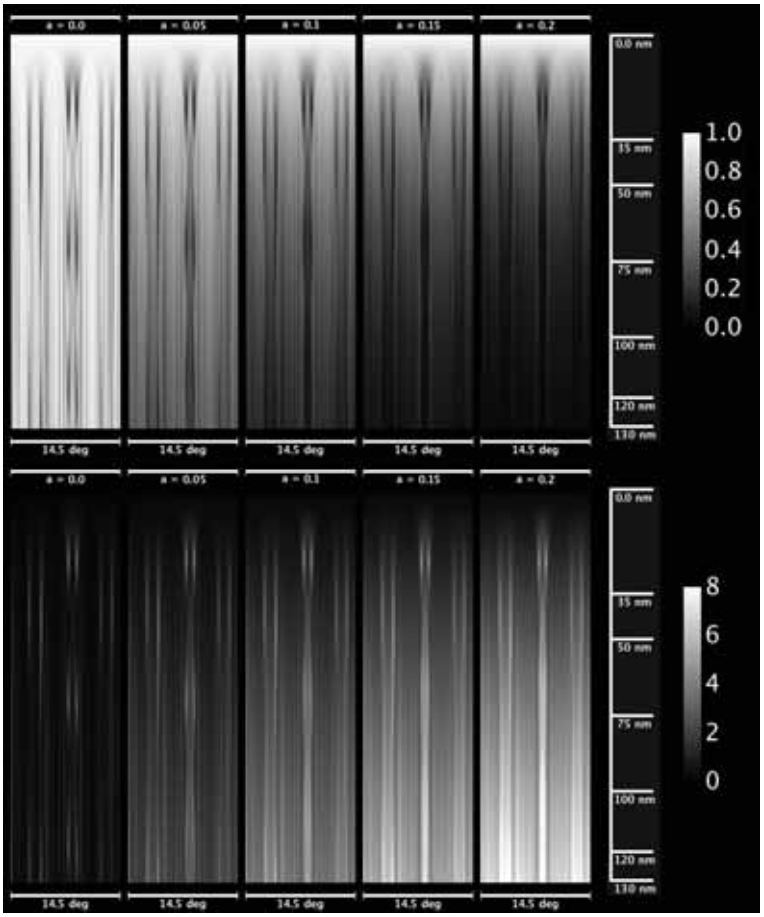


Figure 4.37: Simulated amplitude and amplitude-derived t/λ wedges, resampled for comparison with experimental data. The top image shows amplitude wedges, the bottom amplitude-derived t/λ wedges. These are comparable to the experimental data in Figures 4.28, 4.29, 4.32, and 4.33. The effect of absorption parameter is much stronger than in the phase wedges (Figure 4.36).

noticeably with absorption parameter: higher absorption parameters lead to more amplitude attenuation, as expected. The t/λ synthetic wedge in Figure 4.37 shows that while the presence or non-presence of diffraction contrast features is relatively invariant with absorption parameter, the increase in amplitude attenuation with thickness is clearly noticeable, moreso than on the greyscale amplitude itself.

In the following sections, the simulated and experimental phase, amplitude, and amplitude-derived t/λ data are compared as functions of three parameters: As a function of specimen tilt angle at fixed thicknesses (35, 50, 75, 100, and 120 nm), as a function of specimen thickness in a weakly diffracting condition (+2° to +3°, as used previously), and finally as a function of specimen thickness in a strongly diffracting condition (Laue-condition).

Orientation Dependence at Selected Thicknesses

In this section, the simulated and experimental phase, amplitude, and amplitude-derived t/λ data are compared as a function of tilt angle at selected thicknesses (35, 50, 75, 100, and 120 nm). The individual line profiles, given before as red, green, and blue, are here averaged together in the interest of clarity and brevity. However, the distinction is still retained between thicknesses determined using DT and those determined using RENOIR. Only the surface-charge-corrected phase is used in this section, as the difference in phase between the surface-charge and volume-charge models is much less than the difference between the thickness models. Amplitude and amplitude-derived t/λ are considered first, followed by the phase.

The comparison between simulated and experimental amplitude and amplitude-derived t/λ considered as a function of tilt angle at selected thicknesses allows for analysis of the absorption parameter *a*. This comparison is shown in Figure 4.38 for the amplitude and Figure 4.39 for amplitude-derived t/λ, with different colored lines for different simulated values of *a*, a grey line for values determined using DT, and a black line for values determined using RENOIR. The bottom graph in Figure 4.38, at 120 nm thickness, shows the most attenuation for all values of the absorption parameter and, correspondingly, the highest t/λ values in Figure 4.39 are seen in the 120 nm t/λ graph. As seen in Figure 4.38 and Figure 4.39, irrespective of the thickness determination model used, comparison between experiment and simulation indicates *a* lies experimentally in the range $0.05<a<0.1$ for the weakly diffracting conditions. However, simulations of $a=0.1$ appear to be a better fit to the experimental data for the strongly diffracting conditions. This is reinforced by the other graphs in these two figures, which are more ambiguous due to their lower thicknesses.

Thus, Figure 4.38 and Figure 4.39 indicate that the absorption parameter may be orientation-dependent. Perhaps improved modeling and/or an orientation-dependent absorption parameter are required to sufficiently capture highly diffracting conditions. Also, a heuristic fitting routine might need to be applied to determine the exact tilt angle the experimental data is acquired from, given the angular imprecision inherent in the experimental specimen orientations.

The simulated and experimental phases are more robust against noise than the amplitude and amplitude-derived t/λ, but the absorption parameter poses a problem. The graphs of simulated and experimental phase as a function of specimen tilt angle can be seen in Figure 4.35, with the 120 nm fixed-thickness graph at the bottom. This graph shows two large mismatches between experimental data and simulation. First,

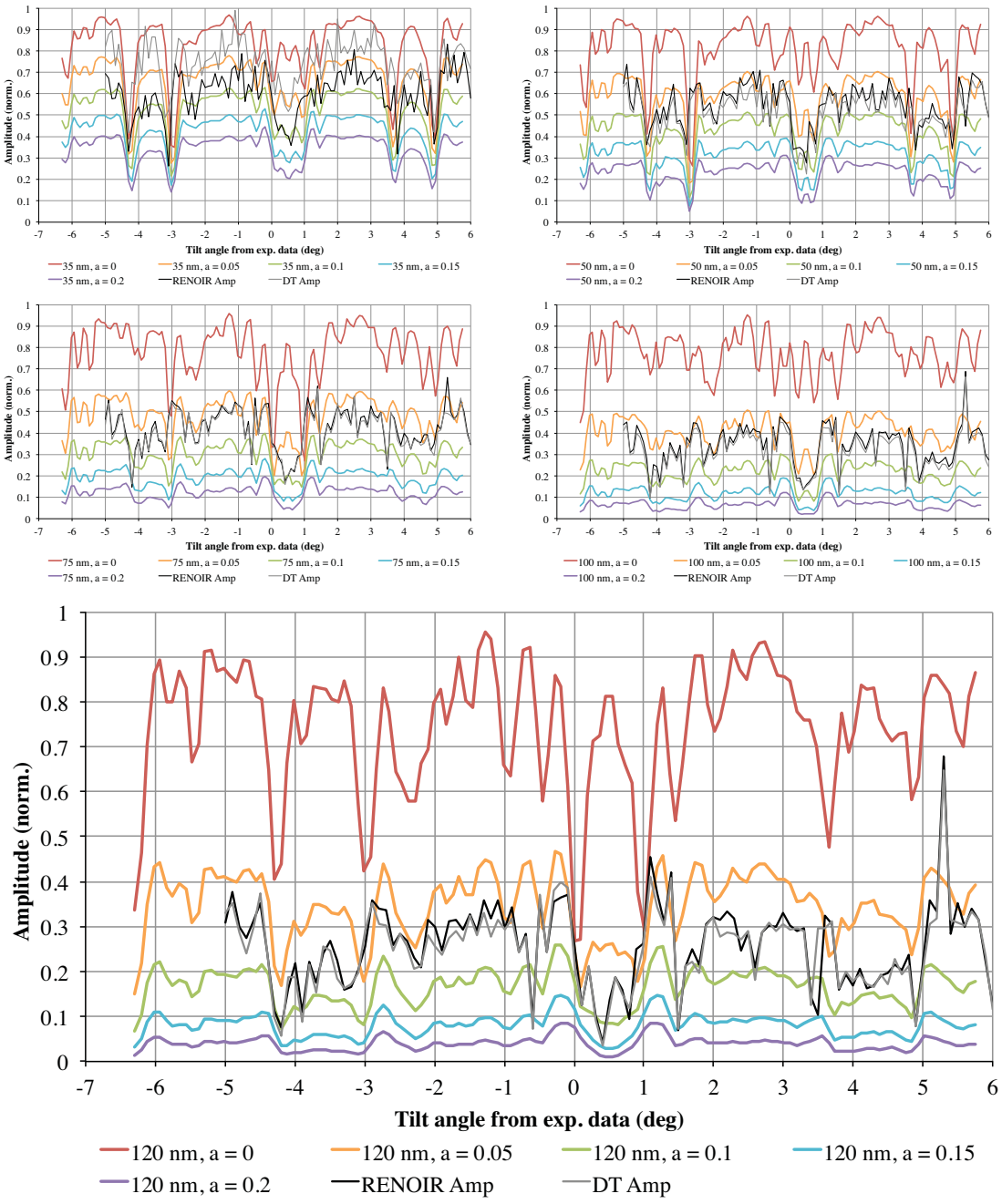


Figure 4.38: Comparison between simulation and experiment for amplitude as a function of tilt angle at selected thicknesses. As with figure 4.35, the features are reproduced well between experiment and simulation. The value of absorption parameter *a* that is the best match seems to be *a*=~0.07-0.08 for weakly diffracting conditions, and *a*=0.1 for strongly diffracting conditions.

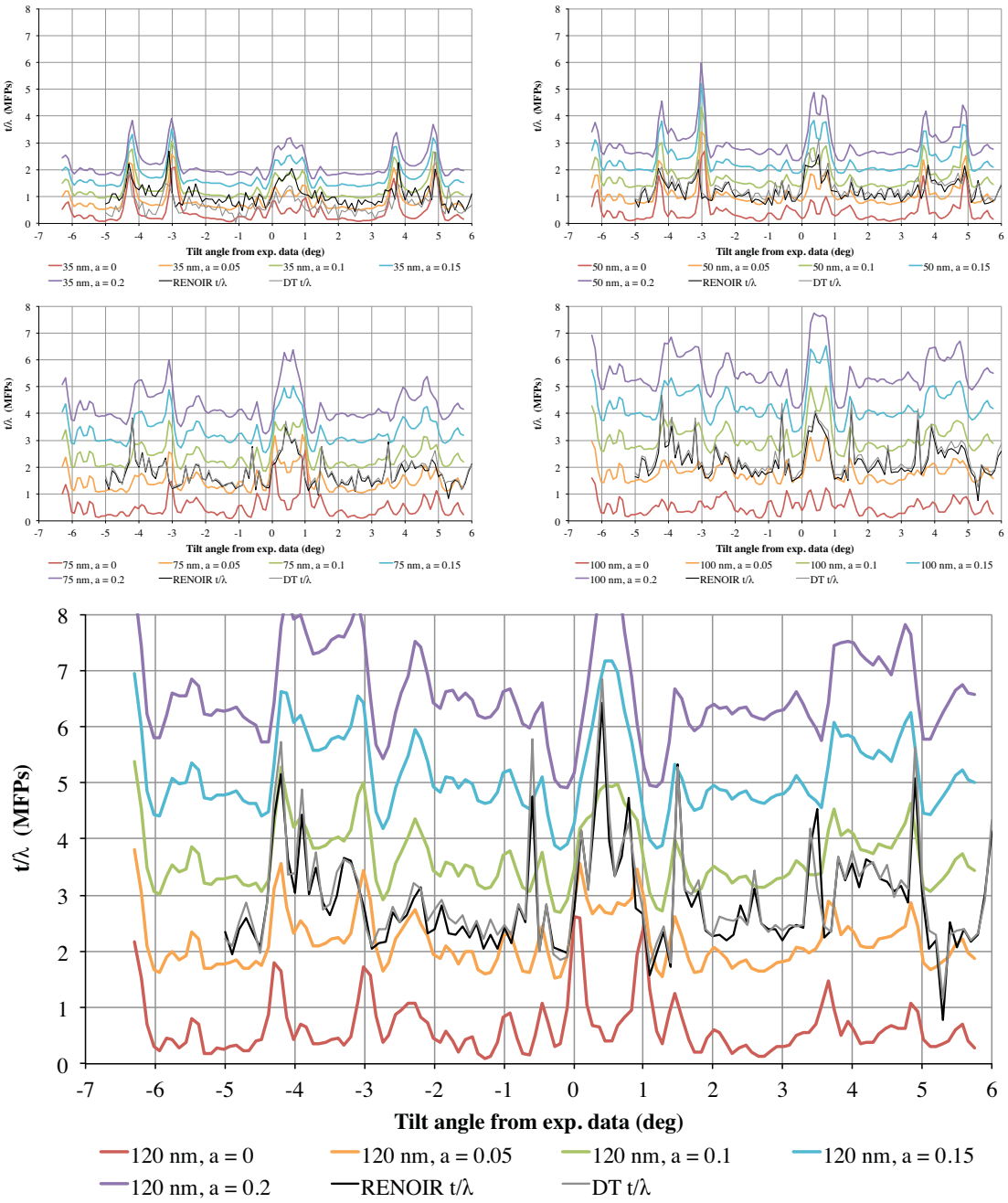


Figure 4.39: Comparison between simulation and experiment for amplitude-derived t/λ as a function of tilt angle at selected thicknesses. As with Figures 4.35 and 4.38, the features are reproduced well between experiment and simulation. As with Figure 4.38, the value of absorption parameter *a* that is the best match seems to be *a*=~0.07-0.08 for weakly diffracting conditions, and *a*=0.1 for strongly diffracting conditions.

the angular span is slightly mismatched, which is especially notable on the leftmost experiment side feature. This is the result of the imperfect angular calibration. The mismatch between the experimental data and the simulated data was determined by taking the width from the outside-most minima on the 35nm phase images. This cannot be calibrated from the higher-thickness data, because there is an absorption-parameter dependence of the width of the features visible – such as the width between the central minima. However, the width of the central plateau – between the bounding minima at +0° and +1° - is consistent at 120 nm with the absorption-parameter determination in the previous paragraph of $0.05<a<0.1$. A simulation at *e.g.* $a=0.07$ might allow for angular calibration between experiment and simulation using thicker regions, improving the angular calibration.

The second notable discrepancy between the experimental and simulated phases in Figure 4.35 is in the general magnitude of the phase, as expected. This discrepancy is especially visible in the weakly diffracting conditions (*e.g.* between +2° and +3°) in Figure 4.35, where the experimental phases are consistently lower than the simulated ones. As stated earlier, this is because the mean inner potential in the simulations used was 15.47V, calculated in the conventional way from isolated-atom scattering factors (IASF) [66] and neglecting any bonding or interaction effects; this will be addressed more in the next section. However, the difference in this case for the simulations between the experimental mean inner potentials of 14.0 ± 0.3 V (RENOIR) and 14.6 ± 0.3 V (DT) and the IASF mean inner potential of 15.47 V is simply a multiplicative factor. In principle, tilt-controlled simulations and measurements of the phase can be matched to determine the phase from a wide variety of specimen orientations.

In future work, the multiplicative scaling of the profile would be a way to determine the mean inner potential. However, determining the mean inner potential by matching tilt-controlled experimental and simulated phases relies on two conditions that are not true for this experimental and simulated dataset: a maximally accurate determination of the orientation at every point, and an accurate absorption parameter being used. Neither of these are the case in this work. However, this matching between simulated and experimental phase is attempted in the next section to show the difficulties.

Quantitative Fitting Between Experimental and Simulated Phase Data

In this short section, the objective is to obtain the mean inner potential by selecting a thickness, and quantitatively fitting the simulated phase and the experimental phase across the tilt series. The results of quantitatively fitting the simulated phase data to the experimental phase data are outlined in this section, but this match is poor due to tilt calibration and not having simulations at the proper absorption parameter.

In principle, the procedure to determine the mean inner potential by matching the simulated and experimental data is simple. First, an absorption parameter a is chosen from the ones graphed in Figure 4.35 – ideally, the closest one as determined from the amplitude information. Then, the simulated phase as a function of tilt angle at the chosen value of a is multiplied by a factor W. W is determined by finding the value that minimizes the root-mean-square (RMS) difference r_{RMS} between the simulated and experimental data. In the data as presented in Figure 4.35, the mean inner potential is set at 15.47 V from isolated-atom scattering factors – see Chapter 5 for more details on why isolated atom scattering factors are suboptimal. Therefore, W/15.47 will be the mean inner potential in volts from the fitted simulated and experimental phase as a function of specimen tilt angle.

Two fixed-thickness datasets are used are used: unweighted-averaging over all five thicknesses shown in Figure 4.35 (35, 50, 75, 100, and 120 nm), and only averaging over the 120 nm tilt series in Figure 4.35; both methods are carried out for DT and RENOIR thicknesses, and for a = 0.05 and a = 0.1 absorption parameters, yielding 8 mean inner potential numbers. The first method is called "5T" below, and the second method is called "1T".

First, the fitting procedure was performed on the tilt-series dataset "5T", with the best-fit RMS residuals r_{RMS} providing a measure of how strong the mismatch is between the simulated data and the experimental data. For the DT thickness data, a=0.05 yields a MIP of 14.81 V (r_{RMS} = 0.72 rad), a=0.1 yields a MIP of 14.88 V (r_{RMS} = 0.73 rad). For the RENOIR thickness data, a=0.05 yields a MIP of 14.3 V (r_{RMS} = 0.72 rad) and a=0.1 yields a MIP of 14.36 V (r_{RMS} =0.73 rad). All of these RMS residuals are very high, compared to the target phases around 14 rad. The high residuals are because the peaks and troughs seen in the experimental data and the simulated data do not line up, due to a combination of the angular width of the strongly diffracting features being absorption-parameter dependent, and the angular width calibration itself not being ideal. If the strongly diffracting features in the simulation are incorrectly fitted to the experimental strongly diffracting features, they will cause large r_{RMS}.

The same high residuals are seen when the fitting procedure is carried out on the tilt-series dataset "1T". For the DT thickness data, a=0.05 yields a MIP of 14.95 V (r_{RMS} = 0.86 rad), a=0.1 yields a MIP of 14.99 V (r_{RMS} = 0.87 rad). For the RENOIR thickness data, a=0.05 yields a MIP of 14.44 V (r_{RMS} = 0.82 rad) and a=0.1 yields a MIP of 14.47 V (r_{RMS} =0.83 rad). Again, the same problem arises with the "1T" dataset as the "5T" dataset: high residuals are the result of a suboptimal fit.

This section details an attempt to obtain mean inner potential values by fitting the simulated phase data to experimental phase data across a range of tilt angles. However, the mismatch between the experimental and simulated data is too high because of a combination of angular width miscalibration and absorption-parameter-dependent angular width of strong diffraction features. The angular width calibration could perhaps be further optimized in this work, but, without the proper absorption parameter to match the angular width calibration to, it would be of dubious value. Therefore, in future work, this approach to mean inner potential determination might be useful if it is coupled with simulations that have a close match between the measured experimental absorption parameter and the simulated one. As noted above, this may require a more complex absorption parameter model.

Moving on from selected-thickness comparisons between simulation and experiment as a function of tilt angle, the next section examines selected-tilt measurements as a function of thickness.

Comparisons of Selected Specimen Orientations

In this section, simulated and experimental phases, amplitudes, and amplitude-derived t/λ are examined as a function of thickness at selected orientations. A weakly diffracting condition and a strongly diffracting condition are both extracted from the synthetic wedges discussed earlier and examined. The weakly diffracting condition is the same as the one used in the previous experimental analysis: the interval [+2.1, +3.0]°, comprising 10 data points. The strongly diffracting condition is selected from

a narrower span near the center of the central diffraction feature: the interval [+0.3, +0.4]°, comprising 2 data points near the Laue condition. As before, the experimental data corresponding to thicknesses below 25-35 nm is not clearly interpretable.

All of the simulated and experimental phases, amplitudes, and amplitude-derived t/λ plots as a function of specimen thickness in this section use the same color scheme, shown in *e.g.* Figure 4.40. The different absorption parameters are plotted in color, the DT thickness determination is plotted in grey, and the RENOIR thickness determination is plotted in black.

From the weakly diffracting condition, comparing simulated and experimental amplitudes allows for determination of the absorption parameter, and comparing simulated and experimental phases is comparatively uninformative. Plotted as a function of specimen thickness, the amplitude is shown in Figure 4.40a, the amplitude-derived t/λ in Figure 4.40b, and the phase is in Figure 4.41. Both the amplitude and t/λ plots in Figure 4.40 reinforce the conclusion from the previous sections that the absorption parameter a lies in the range $0.05<a<0.1$ for weakly diffracting conditions. There are some fluctuations in the experimental data, likely due to noise and thickness calibration imprecision (as detailed in previous sections). The comparison between simulated and experimental phase, seen in Figure 4.41, is relatively uninformative by comparison – the simulations all appear the same, and the discrepancy between the simulated and experimental data is due to the difference in mean inner potentials. By taking the slope from the amplitude-derived t/λ data in Figure 4.40b, the mean free path can be calculated from the simulated data. This is graphed in Figure 4.42. From this, given the experimentally determined mean free path of ~48 nm, the expected absorption parameter would be approximately 0.08.

From the strongly diffracting condition, comparing simulated and experimental amplitudes yields a different absorption parameter than the weakly diffracting condition, and comparing simulated and experimental phases is potentially useful, but not in this instance. The strongly diffracting condition amplitude plotted as a function of specimen thickness is seen in Figure 4.43a, with the corresponding t/λ data in Figure 4.43b, and phase data in Figure 4.44. The experimental amplitude data in Figure 4.43 is much noisier than in Figure 4.40, even above 35 nm, likely due to the difference in averaging windows (10 points *vs.* 2) and the noisier input data. From Figure 4.43a and Figure 4.43b, the experimental amplitude and amplitude-derived t/λ tracks the a=0.1 line. Thus, the absorption parameter in the strongly diffracting condition is higher than in the weakly diffracting condition, indicating that the absorption parameter is orientation-dependent in strongly diffracting conditions. If this were true, as noted earlier in this chapter, the model for absorption parameter should take orientation-dependence into account.[29] The amplitude and amplitude-derived t/λ in Figure 4.43 also shows wavy variations in the simulated data that may or may not be reproduced in the experimental data – the experimental amplitudes are too noisy to tell, for both the amplitude and amplitude-derived t/λ - these wavy variations are not interpretable in this data, but suggest experimental and simulated amplitudes may share subtle identifiable features. From Figure 4.44, the phase as a function of thickness in the strongly diffracting condition is similar to that in the

[29] Another source of this shift in absorption parameter between strongly and weakly diffracting conditions may be if the angularly averaged simulation amplitude is not representative of the amplitude at the actual specimen orientation from where the experimental data was acquired.

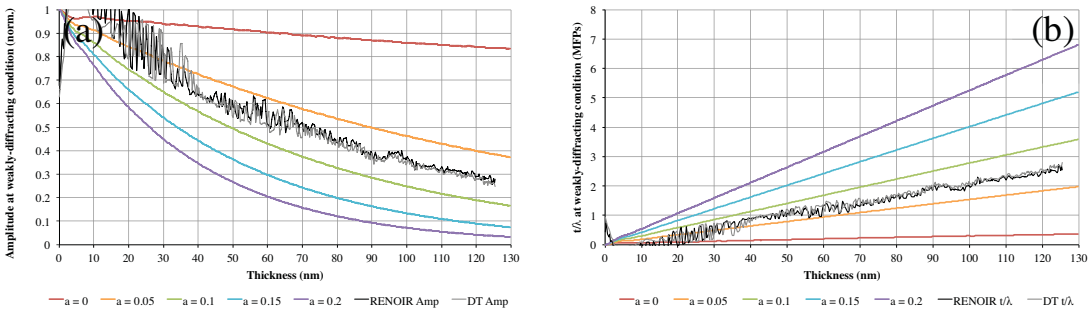


Figure 4.40: Weakly diffracting conditions: Experimental and simulated amplitude and amplitude-derived t/λ as a function of thickness. This is taken from the weakly diffracting condition on the interval [2.1,3.0]°. The results of the RENOIR and DT thickness determination methods (both with shell compensation) agree closely with each other for both amplitude and amplitude-derived t/λ. The experimental result lies between the *a*=0.05 and *a*=0.1 absorption parameters.

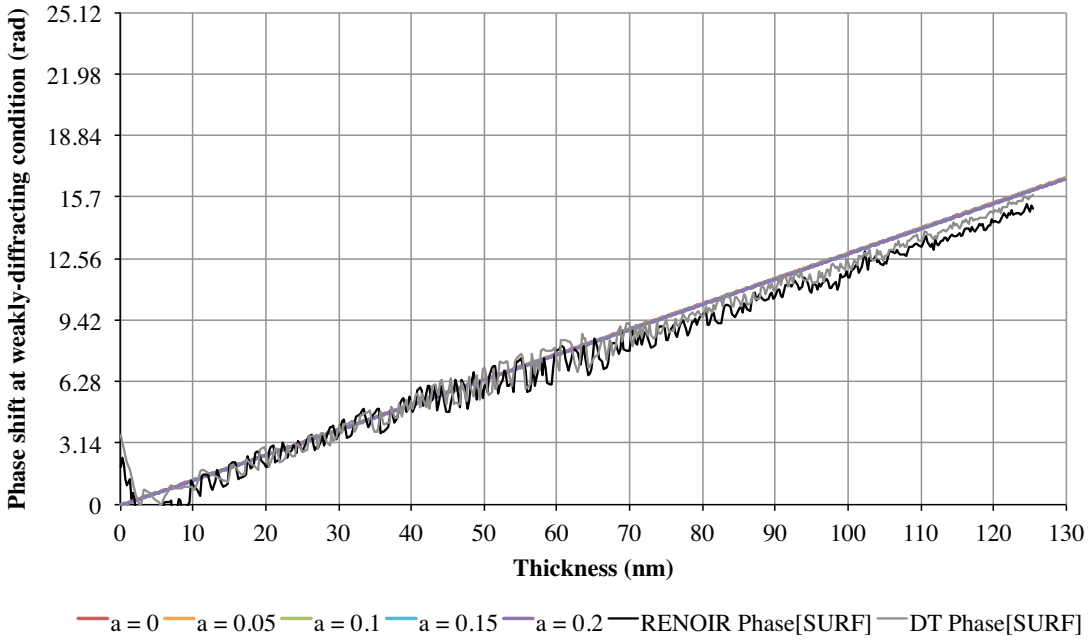


Figure 4.41: Weakly diffracting conditions: Experimental and simulated phase as a function of thickness. As Figure 4.40, but for the phase. There is no discernable change in the phase with absorption parameter, and the results are linear.

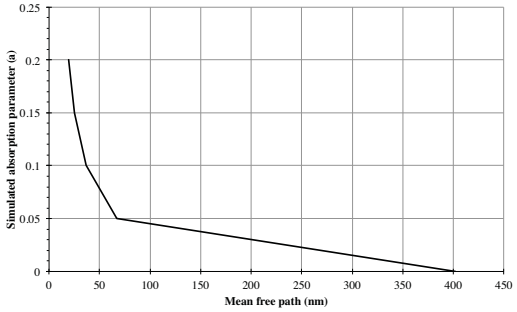


Figure 4.42: Relation between absorption parameter and mean free path. Mean free paths increase for lower absorption parameters, as expected (*e.g.* Figure 4.40).

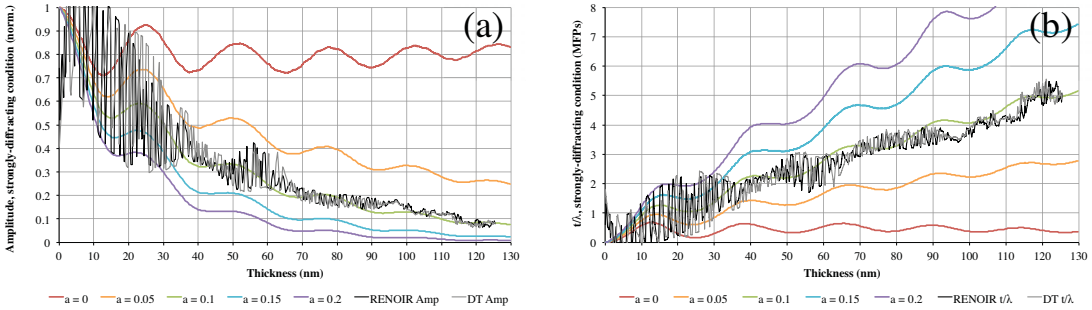


Figure 4.43: Strongly diffracting conditions: Experimental and simulated amplitude and amplitude-derived t/λ as a function of thickness. Like Figure 4.40, but averaged over 0.2° of a strongly diffracting condition (hence the increased noise). The non-monotonicity is evident with the fluctuations in the simulated data, which also are possibly present in the experimental data (but this may be noise). The results agree more closely with the $a = 0.1$ absorption parameter.

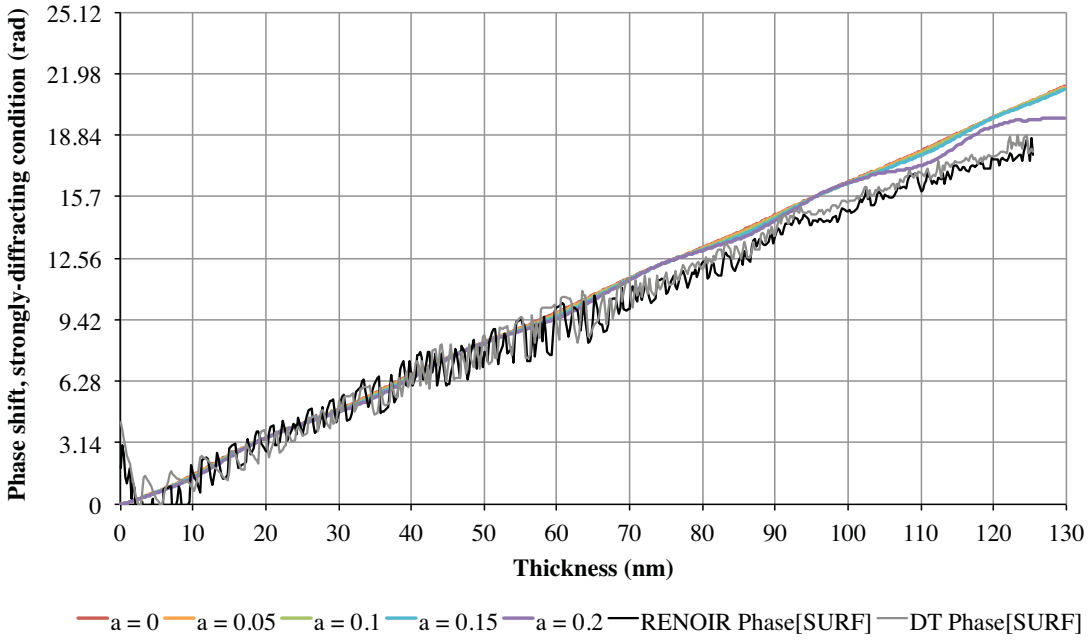


Figure 4.44: Strongly diffracting conditions: Experimental and simulated phase as a function of thickness. As Figure 4.41, but for a strongly diffracting condition. There are fluctuations with absorption parameter, and potentially fluctuations in the experimental data, but the data are noisy. There is a definite effect from the strongly diffracting condition – compare to Figure 4.41 – but it is difficult to quantify the thickness-dependent features.

weakly diffracting case: the experimental phase shift is below than the simulated phase shift, due to differences in the mean inner potential, but the strongly diffracting phase shifts are increased compared to the weakly diffracting phase shifts previously examined in Figure 4.41. This increased phase shift means an increased mean inner potential. Figure 4.41 also shows some variation in the phase with absorption parameter for a=0.2, but not for a<0.2.

In summary, comparing the simulated phase, amplitude, and amplitude-derived t/λ as a function of thickness at fixed tilt angles reinforced the conclusions from previous sections. There are diffraction effects on the phase, amplitude, and amplitude-derived t/λ. The phase data shows a change in mean inner potential between strongly and weakly diffracting conditions, and the amplitude shows a change in absorption parameter a between strongly diffracting conditions (a=0.1) and weakly diffracting conditions (a=0.08).

Conclusions from Comparison Between Simulation and Experiment

Image simulation and experiment show good agreement in this instance, within the limits of this experiment. The diffraction contrast features present in the phase, amplitude, and amplitude-derived t/λ shown in the experimental synthetic wedges were generally reproduced in the simulated wedges. The simulations had much more orientation-sensitivity than the experimental data. Therefore, resampling the simulated synthetic-wedges to the same thickness and angular scales as the experimental synthetic-wedges required averaging over simulations from ~9 different orientations, smoothing out some of the features seen in the simulations. The tilt step between images in the experimental data also does not match the simulations; angular scaling was required to compare the two datasets. This resampling and rescaling allowed for comparison between the experimental and simulated synthetic wedges.

Simulated and experimental data were compared for fixed thicknesses and fixed orientations. Amplitude comparisons were used to determine the absorption parameter a, where weakly-diffracting conditions show an a of approximately a=0.08, but strongly diffracting conditions have a=0.1. From the phase, the orientation-dependent features seen in the experimental phase measurements were confirmed with simulation. The mean inner potential was not determinable from the simulations, as they take the mean inner potential as an input value. However, there is good correspondence between phase simulations and experiments, given the uncertainty about the absorption parameter.

Generally, this section shows that simulations of electron holograms reproduce experimentally observed diffraction features, and these simulations allow for quantitative parameter determination and analysis, including analysis of observed diffraction effects. However, there is scope for improving simulations, especially regarding the absorption parameter.

Conclusions From Tilt-Controlled Electron Holography

This chapter uses a tilt-controlled series of electron holograms of a single InAs nanowire at known specimen orientations, combined with image simulations and both DT and RENOIR thickness determination from Chapter 3, to provide insight into the mean inner potential, mean free path, and simulation absorption parameter.

The Mean Inner Potential, Mean Free Path, and Absorption Parameter

Using the phase and amplitude from the tilt series of electron holograms, the mean inner potential, mean free path, and absorption parameter were examined as a function of specimen thickness, and as a function of specimen orientation. The thickness was quantified from high-angle annular dark field (HAADF) using two methods described in Chapter 3: discrete tomography (DT) and the renormalized orthogonal-images routine (RENOIR). The nanowire used in this chapter has a thin shell around an InAs core; this shell is compensated for using the DT nanowire reconstruction.

After rigorous spatial calibration of the experimental amplitude and phase line profiles, it was found that there was nonlinearity in the nominally linear holography amplitude and phase from weakly diffracting conditions when plotted as a function of thickness. This was attributed to the shell (likely *e.g.* a native-oxide) around the wire that did not show up in HAADF imaging due to the strong atomic-number dependence of HAADF imaging. Compensating for this shell fully accounted for this nonlinearity.

For wurtzite InAs, the mean inner potential is measured in two ways. The mean inner potential of wurtzite InAs determined using synthetic wedges generated from the nanowire is 14.0 ± 0.4 V using RENOIR thickness determination, and 14.5 ± 0.4 V using DT thickness determination. To compare, the mean inner potential from the maximum-thickness method is 14.0 ± 0.3 V using RENOIR thickness determination, and 14.6 ± 0.3 V using DT thickness determination, assuming a shell mean inner potential of 12 ± 4 V. The effect of nanowire charging in general is detectable in the phase images and about 0.3 V, but the difference between different charging models is not detectable and would be about 0.05 V. The measured mean free path of InAs is 48 nm using RENOIR thickness determination and 47 nm using DT thickness determination, but is quite noisy unless spatially averaged.

Both the mean inner potential or mean free path of the shell are more indeterminate. The mean inner potential is less precisely measured for the shell than the core: the mean inner potential of the shell ranges from 12 V to 21 V using RENOIR thickness determination, or 8 V to 17 V using DT thickness determination. The mean free path of the shell from the spatial average is 23 nm from RENOIR thickness determination, or 26 nm from DT thickness determination, but individual-profile measurements range between 17 and 29 nm for RENOIR thickness determination, and 19 to 35 nm for DT thickness determination. This is likely due to the small thickness of the shell.

Across the tilt series of electron holograms taken at different specimen orientations relative to the beam, the phase and amplitude both showed diffraction effects. Image simulations as a function of specimen orientation and specimen thickness were carried out and analyzed. The diffraction features visible in the experimental data were reproduced in the simulations. The simulations allowed for absorption parameter analysis. The absorption parameter appears to be orientation-dependent: approximately 0.08 for weakly diffracting conditions, and approximately 0.1 for strongly diffracting conditions. A discussion of the ramifications of these simulations for the Stobbs factor, and its relation to the results seen here, is in Chapter 7.

Requirements For Experimental Parameter Determination

This chapter shows that there are several critical requirements for experimental parameter determination (mean inner potential, mean free path, absorption parameter)

from electron holograms. There are also several ways other techniques can aid in interpretation.

The first requirement is spatial calibration of the experimental data. The spatial calibration is an absolute requirement for quantification, and errors in spatial calibration increase the errors in parameter determination. These errors can be detectable or invisible, systematic or random. The second requirement is high-fidelity data, with the best signal-to-noise ratio possible. Even holograms acquired with 64s acquisition times in this experiment have very noisy amplitude information, especially in strongly diffracting conditions. Third, the thickness of the specimen must be known as accurately and as precisely as possible – preferably through more than one technique. The thickness determination also interacts with the spatial-calibration requirement. Fourth, the orientation of the specimen relative to the electron beam must be known – there is a major difference between characteristics of strongly and weakly diffracting conditions. Fifth, if phase information is to be interpreted, then phase wraps in rapidly varying regions must be exactly unwrapped, preferably with reference to the original holograms in ambiguous instances. Sixth, the state of the specimen surface should be known – even thin shell layers can lead to changes in parameters determined from thick materials. Finally, if diffraction contrast is expected, suspected, or desired, it should be quantitatively accounted for with reference to simulation.

The results of other TEM techniques can aid in interpreting electron holograms. Knowledge of the morphology of the specimen, ideally from a technique that allows for differentiation into different components (in this case, HAADF STEM) allows for full thickness compensation, if required. Knowledge of the specimen orientation (in this case, from diffraction patterns) allows for strongly or weakly diffracting conditions to be identified. Knowledge of the state of the specimen surface (in this case, from BFTEM) allows for shells to be quantified and shell effects can thus be removed, if necessary.

Increasing Realism

The image simulations used here of electron holograms match the experimental amplitudes reasonably well given the uncertainty about absorption parameter, but they do not shed light on the phase. Returning to mean inner potential measurement, two individually sensible thickness determination techniques yield mean inner potentials that are not quite consistent with each other. The most precise values determined above are 14.0 ± 0.3 V or 14.6 ± 0.3 V measured with RENOIR and DT thickness determinations, respectively, assuming the shell mean inner potential is 12 ± 4 V. The synthetic wedge method yields slightly better agreement between the two thickness determination methods due to decreased precision: 14.0 ± 0.4 *via* RENOIR thickness determination and 14.5 ± 0.4 V *via* DT. This leads, naturally, to the question of which thickness determination method is most accurate. However, this cannot be answered from the data thus far, because the mean inner potential value calculated from isolated-atom scattering factors is 15.47 V, which is far from the experimental values. To answer this underlying question of what the true and accurate mean inner potential value is, the realism of the mean inner potential simulations must be increased.

To increase the realism of the simulations, the mean inner potential calculation must move from using isolated-atom scattering factors to accounting for realistic charge distribution in the system – including bonding, exchange, correlation, and surface

effects. The role of the shell on the surface must also be examined – does this surface shell affect the mean inner potential of the core itself? In other words, would the core mean inner potential be different without the shell? To answer these questions, the next chapter applies density functional theory to simulate mean inner potentials and explore the factors that can affect the mean inner potential of a specimen.

Chapter 5: Density Functional Theory

Outline

While Chapter 4 focuses on experimental measurement of the mean inner potential, this chapter discusses theoretical mean inner potential simulations[30] using density functional theory (DFT). DFT allows for mean inner potentials to be simulated for ideal specimens. After establishing the precision of the DFT mean inner potential simulations, comparing mean inner potential results from different specimens allows for the fundamental nature of the mean inner potential to be examined.

This chapter focuses on DFT simulations of mean inner potentials of group IV and group III-V semiconductors, as well as carbon in the diamond structure. These materials are used because InAs, studied experimentally in chapters 3 and 4, is a III-V semiconductor, and DFT simulations of mean inner potentials in these materials have been reported in the literature – a typical example is found in [8], where the objective is simply to calculate the mean inner potential as a material-dependent parameter. In contrast, DFT simulations are used in this chapter to understand the factors that affect the mean inner potential, including surface effects.

The primary approach of this chapter is to take internally self-consistent approaches to determine what factors can affect the mean inner potential, and under what conditions those factors might be relevant. The objective is to determine what general principles guide measurements of the mean inner potential, as determined using DFT simulations. While comparison with experimental data would be of theoretical interest, the goal of this chapter is not to exhaustively test experimental systems.

First, inaccurate non-DFT mean inner potential calculations are briefly carried out from isolated-atom scattering factors to establish a baseline. This is followed by three extensive DFT simulation sections, establishing first the limits of precision for DFT simulations of the mean inner potential, then establishing a set of standard DFT-simulated mean inner potentials for the materials under study (similar to the work performed in *e.g.* [8]), and thereby ensuring that the code used here does not wildly disagree with previous work. Finally, several factors are examined for their effects on the mean inner potential – the most important of these factors is the state of the specimen surface. Surface knowledge, such as that used here, is unlikely to be known for a TEM specimen – see Chapter 7 for more details.

This chapter explicitly does not discuss the effects of diffraction as explored in Chapter 4. The calculations and simulations of mean inner potential discussed here assume no diffraction effects. This chapter intends to examine factors that may affect the mean inner potential itself, instead of factors affecting the measurement using electron holography.

Non-DFT Mean Inner Potential Calculation

It is possible to calculate the mean inner potential without using DFT. Early work on computing the mean inner potential used electron isolated-atom scattering factors

[30] A terminological pedantry: this chapter refers to the DFT computations performed in this work as "simulations" instead of "calculations", because density functional theory is an approximate method for reasons detailed in Chapter 2. Ergo, for the DFT computations, the word "simulation", with its connotation of imperfection and possible refinement, was deliberately chosen over "calculation", which connotes the inevitable derivation of a single invariably correct answer.

(IASF) at zero scattering angle for in a pro-crystal geometry - see [67] and [66]. These IASF-derived mean inner potential values are only dependent on atomic number and average volume per atom. The IASF approach is quantitatively inaccurate, because the IASFs are calculated from isolated atoms, and thus ignore bonding effects and realistic electronic structure in the crystal under study [89]. Therefore, more recently, DFT has been used to produce mean inner potential values for materials using a more realistic electronic structure (*e.g.* [8], [52], [89], [90], [91], and [92]).

In this section, though, the results for isolated-atom scattering factor calculations of mean inner potentials are discussed for materials used in DFT calculation later in this chapter.

Isolated-Atom Scattering Factors

Isolated-atom scattering factor calculations of the mean inner potential use a calculation based on the volume of the unit cell and the electron scattering factors [66] [67]:

$$MIP_{IASF} = \frac{h^2}{2\pi m_e e} \frac{\sum_i^N f_{el,i}}{V_c}$$

where V_c is the volume of the unit cell and $f_{el,i}$ is the electron scattering factor for each atom, and the summation is taken over the unit cell (h, m_e, and e are the usual physical constants). This formula, when applied to the suite of materials and the relevant lattice parameters, yields the mean inner potentials seen in Figure 5.1. In Figure 5.1, aREF is the reference experimental lattice parameters from [113] and aFD being the minimum-energy lattice parameters for finite-difference mode GPAW DFT simulations, discussed in the following sections. While the IASF mean inner potentials are precise, this pro-crystal calculation does not account for inter-atom electronic interactions, such as bonding, so it should be incorrect quantitatively.

DFT Parameters and Conditions

The following sections detail the parameters and conditions used for DFT simulations of mean inner potential in this chapter. The DFT simulations in this work were carried out exclusively using the grid-based projector-augmented wavefunction (GPAW) method, implemented as the GPAW software library [114].

GPAW is also used in **Paper II: Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials, to calculate mean inner potentials**. The same approach is taken in this work as in **Paper II**, and it provides details of how the mean inner potential is extracted from GPAW output files.

In the following sections, the "standard" settings used from the GPAW code are listed, for both finite difference (FD) and linear combination of atomic orbitals (LCAO) modes[31]. The method for extracting mean inner potentials from the GPAW output is also discussed. Finally, the two different simulation geometries – "thin-film" and "nanowire" – and the two methods of measuring the mean inner potential from this data – "subtraction" and "integration" – are examined.

[31] Note that, as GPAW is a grid-based code, each unit cell of material was allocated an integer number of grid points in each direction (x, y, and z) to avoid aliasing effects.

Isolated-Atom Scattering Factor Mean Inner Potentials			
Material	**a (Å)**	**a type**	**MIP (IASF) (V)**
C	3.567	aREF	20.86
C	3.577	aFD	20.68
Si	5.431	aREF	13.76
Si	5.477	aFD	13.41
Ge	5.658	aREF	15.60
Ge	5.765	aFD	14.74
GaAs	5.653	aREF	15.38
GaAs	5.773	aFD	14.44
Material	**a (Å)**	**a type**	**MIP (IASF) (V)**
InP	5.869	aREF	15.22
InP	5.908	aFD	14.92
InAs	6.059	aREF	15.47
InAs	6.100	aFD	15.16
GaP	5.490	aREF	14.60
GaP	5.545	aFD	14.17
AlP	5.451	aREF	13.41
AlP	5.513	aFD	12.96
AlAs	5.662	aREF	13.97
AlAs	5.740	aFD	13.40

Figure 5.1: Mean inner potentials calculated using isolated-atom scattering factors (IASF). The mean inner potentials shown here are calculated using the volume of the unit cell and the isolated-atom scattering factors for each element (see text and [67]). Isolated-atom scattering factors from [67]. Reference lattice parameters from [113]. These can be compared with the results from DFT calculations elsewhere in the chapter.

FD Mode Lattice Parameters				
Material	**a_{REF} (Å)[113]**	**a_{FD} (Å)**	**Δa**	**ΔVol**
C	3.567	3.577	+0.28%	+0.84%
InP	5.869	5.908	+0.66%	+2.01%
InAs	6.059	6.100	+0.68%	+2.04%
Si	5.431	5.477	+0.85%	+2.56%
GaP	5.490	5.545	+1.00%	+3.04%
AlP	5.451	5.513	+1.14%	+3.45%
AlAs	5.662	5.740	+1.38%	+4.19%
Ge	5.658	5.765	+1.89%	+5.78%
GaAs	5.653	5.773	+2.12%	+6.50%

Figure 5.2: DFT minimum-energy lattice parameters, FD mode. Here, the experimental lattice parameters from [113] are shown alongside the minimum-energy lattice parameters simulated using FD mode DFT calculations using GPAW, ordered from least-mismatch to most-mismatch. There is a material-dependent mismatch between the a_{REF} and the a_{FD} lattice parameters. The ordering of the materials in this table is used throughout this chapter.

Technical Specification

All simulations were run using a pair of 8-core Mac Pro computers clustered together with the same software packages and input for the simulations. One computer (the master) had 16 GiB of RAM; the other (the adjunct) had 8 GiB. Both computers were running Mac OS X 10.6. Parallel clustering was done with MPICH2 v. 1.0 and execution with Python 2.6.1; 8 parallel "nodes" were instantiated on the master, and 4 parallel "nodes" on the adjunct, to yield 2 GiB of RAM per node. GPAW software library version 0.6 stable [114] was used, together with its dependency, ASE software library version 3.2.0 stable [115]. Structural optimization was performed with a quasi-Newton force minimizer as implemented in ASE. Lattice parameters were determined with the direct command-line implementation of GPAW; all other simulations were performed using Python input files that called the GPAW library.

FD and LCAO Modes

In the GPAW software, the finite-difference (FD) and the linear combination of atomic orbitals (LCAO) simulation modes are both available, and both FD mode and LCAO mode are used to simulate mean inner potentials in this chapter.

FD mode and LCAO mode differ in how the electron wavefunctions are solved. FD mode uses a grid-based finite-difference method: specifically, residual minimization by direct inversion in the iterative subspace is used, although other finite-difference options, such as conjugate gradient, are available in the GPAW code [61]. LCAO mode specifies and solves a linear combination of atomic orbitals spanning a basis set defined by zeta functions – see [116] and [117] for further details. LCAO mode and FD mode also have slightly different input parameters. LCAO mode requires all the same input parameters as FD mode, but with the addition of a parameter specifying which basis set is used. For both FD and LCAO modes, the output data is still on a grid, and an exchange-correlation functional (*e.g.* LDA or PBE) must still be specified.

Multiple LCAO basis set options are available, with a basis set of more zeta functions representing increased specificity [117]. Double zeta polarized basis sets were used when mean inner potentials were calculated, but other basis sets are possible In order of ascending specificity, the basis sets used in this chapter and their abbreviations: single zeta polarized (szp), double zeta polarized (dzp) triple zeta polarized (tzp), triple zeta double polarized (tzdp), quadruple zeta polarized (qzp), quadruple zeta double polarized (qzdp), and, finally, quadruple zeta triple polarized (qztp). Increased specificity also means increased RAM usage.

Advantages and Constraints of Both the FD and LCAO Modes

Both FD mode and LCAO mode can be used to determine mean inner potentials, but each mode has its own advantages and constraints, some of which are noted later in this chapter. As discussed in the later section on precision, FD mode is generally more precise than LCAO mode for mean inner potential determination, so FD mode was preferred in this work. However, FD mode is slower and uses more RAM than LCAO mode, limiting the specimen size that can be simulated. Therefore, LCAO mode was used as both a check against and comparison for FD mode.

LCAO mode is not only faster, but, critically, it requires less memory than FD mode. Unfortunately, the LCAO basis set for indium was not operating, meaning no In-

containing materials could be treated with LCAO mode. For larger systems than the ones simulated here, LCAO mode might be useful.

Simulation Geometry: "Thin-Film" and "Nanowire"

Two different simulation geometries were used for mean inner potential simulation in this chapter – "thin-film" and "nanowire". Both use periodic boundary conditions in all three dimensions. The difference between the "thin-film" and the "nanowire" is that the "thin-film" has interfaces between free space and material in one direction, and the "nanowire" has interfaces between free space and material in two directions. Therefore, the thin-film simulation has continuous material in two directions, and the nanowire simulation has continuous material in one direction. Mean inner potential DFT simulation in the literature is conventionally done in a thin-film geometry *e.g.* [89].

Surfaces are required in the simulation because a single bulk unit cell of the material is insufficient for mean inner potential simulation. A simulation of the mean inner potential must include a reference region of free space far from the material, and cannot use bulk alone [118].[32] However, the choice of which surface is used may make a difference to the mean inner potential. Therefore, multiple different surfaces must be examined.

Both "thin-film" and "nanowire" geometries are used in this work. The thin-film geometry is used for almost all of the DFT simulations presented in this chapter, and for all quantitative mean inner potential determination. The nanowire simulation geometry is applied in a few simulations to determine the nanoscale topology of the electric potentials surrounding the specimen.

Two Methods: Subtraction and Integration

Two methods are used in this work to determine the mean inner potential value from the DFT thin-film simulations. The mean inner potential is determined from the simulated electrostatic potential[33] on the three-dimensional grid. Two methods used in this chapter for extracting the mean inner potential from this simulated electrostatic potential are the subtraction method, which is used in the literature (*e.g.* [89]), and the integration method, which model the thin-film mean inner potential measurement directly.

With the subtraction method, the average total electrostatic potential in the center of the material (averaged over two unit cells) is subtracted from the average total electrostatic potential in the center of the free space around the material. The subtraction method is free of any direct surface contribution – no part of the surface is included in the averaged regions. The subtraction method is also used in **Paper II: Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials.**

The integration method presents a more realistic comparison for thin systems. To determine the mean inner potential using the integration method, the electrostatic

[32] Note that for a pro-crystal, isolated-atom scattering factor mean inner potential calculation, this requirement does not matter, as the isolated-atom scattering factors are calculated for single isolated atoms in free space.

[33] As noted above, the method for turning the GPAW-simulated electrostatic potential into mean inner potential is detailed in **Paper II: Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials**.

potential in vacuum is set to 0, then the total electrostatic potential over the volume is averaged and divided by the specimen thickness. The integration method includes the surface contribution directly in the average. The integration method is analogous to the experimental mean inner potential measurement, where the total phase shift is divided by the total thickness. Comparing the integration method mean inner potentials with the subtraction method mean inner potentials gives an idea of how much of a discrepancy may exist between thin specimens and thick specimens.

Essentially, the subtraction method is the mean inner potential measured from a thick specimen, while the integration method simulates the mean inner potential for the exact specimen that is simulated. The subtraction method is generally used in this chapter; the integration method is an alternative used in the following sections occasionally on selected simulations to compare thin and thick specimens.

Validity and Precision Testing

This section tests the validity, precision, and accuracy of DFT simulations of the mean inner potential by varying the simulation input parameters that should not influence the mean inner potential. There is scope for improving precision determination of DFT-determined mean inner potentials. In the literature, *e.g.* [8], error analysis of DFT mean inner potential determination usually only tests a small number of additional scenarios. In addition, since this work uses GPAW, which is a different DFT program than [8] (which uses Wien2k), the GPAW mean inner potentials should be verified as similar to those reported in the literature.

In the following sections, precision determination was performed on both the mean inner potentials and the lattice parameters, using a standardized test suite of DFT simulations.

Test Suite Properties

A standardized test suite of DFT simulations for precision determination of mean inner potentials was developed for this work, and is detailed in this section. The test suite is designed to determine the possible operating envelope of DFT simulation parameters for mean inner potential simulations using the GPAW DFT program. In this instance, the operating envelope is defined as the set of simulation parameters that yielded stable, internally consistent mean inner potential values. Mean inner potentials from two simulations with simulation parameters both from the operating envelope should be the same, within the margin of error.

The test suite had three components: Minimum-energy lattice parameter determination for both FD and LCAO modes, mean inner potential tests in FD mode for a single material as a function of variation in simulation parameters from a baseline "standard" simulation, and mean inner potential tests in LCAO mode. The lattice parameter tests covered all materials used in this. The mean inner potential tests focused on only C, Si, Ge, and GaAs. Multiple materials are necessary because, in GPAW, the atomic setup is generated for each individual element, and therefore multiple elements must be tested to ensure stability and precision.

The "standard" simulation parameters for the mean inner potential tests in the test suite, determined from preliminary simulations to be within the operating envelope, were the PBE exchange-correlation functional [63], 6×6×1 k-points, a real-space grid spacing no larger than 0.200 Å, 7 unit cells of material (meaning 15 monolayers), and a width of vacuum equivalent to 5 unit cells. These parameters were applied to a thin-

film simulation in FD mode with the relevant cubic structure (diamond structure for unary systems, zincblende structure for binary systems)[119].

The mean inner potential test suites examine the effect of adjusting the number of k-points ($k \times k \times 1$), the number of unit cells of material n_{mat}, the width of the vacuum surrounding the material n_{vac}, and the spacing between grid points h (in Å). The width of the vacuum n_{vac} is chosen to be the same width as an integer number of unit cells of material (*i.e.* if $n_{mat}=n_{vac}$, then the thickness of material will equal the width of the vacuum). Each parameter is varied singly, and all of the other "standard" simulation parameters are held constant. The exchange-correlation functional is not included as part of the test suite for error determination because different exchange-correlation functionals would simulate different underlying electronic systems, but different exchange-correlation functionals are briefly examined.

First, however, minimum-energy lattice parameters must be determined from the GPAW program.

Why Test Lattice Parameters?

The choice of lattice parameter is important, as explained in this section. The experimental lattice parameters are perhaps the most accurate for experimental comparison, but they are not the lattice parameters that yield the minimum-energy configuration in the DFT simulations.

The experimental lattice parameters of the bulk system are reported in the literature (the values used here are from [113]), and a DFT simulation using that experimental lattice parameter should yield a mean inner potential value that can be compared with experiment. However, in common with other DFT programs and as reported previously [91] [120] [121], the minimum-energy lattice parameter calculated in the GPAW program will be different from the experimental lattice parameter (*cf.* [113] and [117]).

Minimum-energy lattice parameters are most important for simulations performing structure optimization. Optimizing the structure of the material means determining the best positions for the atoms, but the simulation volume is not re-sized during structure optimization. Structure optimization with a non-minimum-energy lattice parameter creates built-in strain in the DFT simulation, because a non-minimum-energy lattice parameter leads to a simulation volume of the wrong size for the minimum-energy atomic configuration. Using a non-minimum-energy lattice parameter would thus lead to a non-optimized structure, defeating the purpose of the structure optimization. LCAO mode and FD mode will each determine slightly different minimum-energy lattice parameters as well [117].

The lattice parameter has a strong effect on the mean inner potential. As noted above in the isolated-atom scattering factor section, the mean inner potential is inversely dependent on the volume of the unit cell, to a first approximation. Therefore, for a small change in the lattice parameter Δa, the unit cell volume should change by a factor $\sim 3\Delta a$.

Therefore, the objective of the lattice parameter determination section is to determine the minimum-energy lattice parameters for all material systems in FD and LCAO modes. These minimum-energy lattice parameters will be used both during the mean inner potential tests, and for selected other simulations (especially structure-optimized

simulations, or simulations that must be compared with structure-optimized simulations).

Lattice Parameter FD and LCAO Test Results

The lattice parameters are determined for both FD and LCAO mode separately, under the standard conditions listed above.

The minimum-energy lattice parameter test results for FD mode are presented in Figure 5.2, ordered from least difference between experiment and simulation (C at Δa=+0.28%) to most (GaAs, Δa=+2.12%). All of the FD minimum-energy lattice parameters determined here are larger than their experimental counterparts.

This leads to the choice of material for the mean inner potential test suites. Four materials are chosen: C, Si, Ge, and GaAs. C has both the smallest lattice parameter in absolute numbers, and the smallest lattice parameter difference between experiment and minimum-energy. C provides the best "pathfinder" material, as not only is the discrepancy between experimental and minimum-energy lattice parameters the smallest, but C is computationally easier to simulate in GPAW than any other material considered here due to the small lattice parameter of diamond C (a=3.567Å). This small lattice parameter leads to fewer grid points per unit cell necessary to maintain a good sampling density). Si has extensive experimental mean inner potential measurements in the literature [50]. For lattice parameters with a large discrepancy between experimental and minimum-energy values, Ge provides a unary test, and GaAs provides a binary test. All the materials listed in Figure 5.2 will have mean inner potentials simulated under the "standard" simulation parameters.

The minimum-energy lattice parameter test results for LCAO mode are presented in Figure 5.3, with the elements in the order from Figure 5.2. Indium-containing materials are not listed, as the In basis sets do not function in LCAO mode. Figure 5.3 shows that, as the LCAO basis set is extended and improved with more zeta functions, the lattice parameter from LCAO mode moves closer to the FD mode lattice parameter. The single zeta polarized (szp) basis is the most approximate basis set, but the szp basis is also the least accurate when minimum-energy lattice parameters compared with experimental lattice parameters. The double zeta polarized (dzp) basis is significantly more accurate, and further basis-set refinements do not dramatically improve the minimum-energy lattice parameter, but do become more computationally intensive. Therefore, the dzp basis set would the best balance of approximation and accuracy for LCAO basis sets in this work. All of this is fully consistent with previous results [117].

Now, the experimental lattice parameters and the FD minimum-energy lattice parameters are applied to the mean inner potential test suite. The experimental lattice parameter (for Si, 5.431 Å) and the FD minimum-energy lattice parameter (for Si, 5.477 Å) are used for both FD and LCAO mode test suites. The experimental lattice parameter is experimentally relevant. The FD minimum-energy lattice parameter is used for FD mode because it is the minimum-energy lattice parameter, and for LCAO mode to provide comparison between FD mode and LCAO mode.

The LCAO dzp-basis lattice parameter is not used for the test suite. It would be used for LCAO mode relaxation simulations because it is the minimum-energy lattice parameter in LCAO mode. LCAO mode relaxation simulations are not performed in

LCAO Mode Lattice Parameters (Å)

	a_{REF}	a_{szp}	a_{dzp}	a_{tzp}	a_{tzdp}	a_{qzp}	a_{qzdp}	a_{qztp}
C	*3.567*	3.602	3.601	3.598	3.598	3.596	3.596	3.594
Si	*5.431*	5.517	5.496	5.492	5.486	5.491	5.485	5.481
GaP	*5.490*	5.595	5.574	5.575	5.566	5.568	5.565	5.556
AlP	*5.451*	5.557	5.536	5.531	5.522	5.530	5.521	5.519
AlAs	*5.662*	5.786	5.759	5.755	5.750	5.755	5.749	5.746
Ge	*5.658*	5.850	5.813	5.789	5.790	5.788	5.789	5.781
GaAs	*5.653*	5.833	5.804	5.799	5.794	5.794	5.793	5.788

Difference between LCAO Mode Lattice Parameter and Reference (%)

	Δa_{REF}	Δa_{szp}	Δa_{dzp}	Δa_{tzp}	Δa_{tzdp}	Δa_{qzp}	Δa_{qzdp}	Δa_{qztp}
C	0.00%	0.98%	0.95%	0.87%	0.87%	0.81%	0.81%	0.76%
Si	0.00%	1.58%	1.20%	1.12%	1.01%	1.10%	0.99%	0.92%
GaP	0.00%	1.91%	1.53%	1.55%	1.38%	1.42%	1.37%	1.20%
AlP	0.00%	1.94%	1.56%	1.47%	1.30%	1.45%	1.28%	1.25%
AlAs	0.00%	2.19%	1.71%	1.64%	1.55%	1.64%	1.54%	1.48%
Ge	0.00%	3.39%	2.74%	2.32%	2.33%	2.30%	2.32%	2.17%
GaAs	0.00%	3.18%	2.67%	2.58%	2.49%	2.49%	2.48%	2.39%

Figure 5.3: DFT minimum-energy lattice parameters, LCAO mode. As with Figure 5.2, the minimum-energy lattice parameters simulated using GPAW in LCAO mode are different from the reference lattice parameter (from [113]), excepting the indium compounds (not calculated in LCAO mode). Additionally, LCAO mode simulations require choosing a zeta-function basis set, and the effect is also shown here (from single-zeta-polarized (szp) up to quadruple-zeta-triple-polarized (qztp)). With increasing zeta-function specificity, LCAO mode minimum-energy lattice parameters are closer to the reference lattice parameter. The double-zeta-polarized (dzp) basis set was used for LCAO mode testing.

this chapter. The following sections use LCAO mode only to compare mean inner potential test suite precision between FD mode and LCAO mode.

In the next two sections, the mean inner potential test suite is applied to FD mode and LCAO mode calculations. The results from the mean inner potential test suite are both the operating envelope (where the mean inner potential is relatively invariant to simulation parameters), and precision determination of DFT simulations of mean inner potentials.

FD Mean Inner Potential Test Suite Results

In this section, the mean inner potential test suite is carried out in FD mode on both the experimental lattice parameter and the FD minimum-energy lattice parameter for three diamond-structure materials (C, Si, Ge), and one zincblende-structure material (GaAs). The test suite, as detailed above, varies k, the number of k-points; n_{vac}, the width of vacuum; n_{mat}, the width of material; and h, the grid spacing. Only the C results are shown here; the Si, Ge, and GaAs results are similar, and seen in Appendix D.

C behaves well under the FD test suite for both subtraction and integration methods, and is used to establish some of the operating envelope. Using the standard subtraction method for mean inner potential calculation, Figure 5.4 shows the mean inner potential test suite results using the experimental lattice parameter, and Figure 5.5 shows the mean inner potential results using the FD minimum-energy lattice parameter. Figures 5.4 and 5.5 shows that, while there are potentially some deviations at the least-precise regions, for k-point sampling in the range $4 \leq k \leq 16$ and the same with grid-point sampling $h \leq 0.2$ Å, there are no systematic variations, so these ranges define the operating envelope. The thickness of material and vacuum leads to strong deviations for small numbers of unit cells, which vanish as more are added. The change in thickness or width between each n_{mat} or n_{vac} data point is two monolayers (ML) of material (or equivalent width) up to 29 ML of material and/or vacuum. The deviation at low thicknesses means that the region 11 ML $\leq n_{mat} \leq$ 29 ML and 5 ML $\leq n_{vac} \leq$ 29 ML is taken to be within the operating envelope[34]. There is no substantial difference in the precision between using the experimental lattice parameter and the FD minimum-energy lattice parameter, but the two techniques lead to different determinations of the mean inner potential. The integration method mean inner potentials for C are seen in Figures 5.6 and 5.7 for the reference lattice parameter and FD minimum-energy lattice parameter, respectively. Figures 5.6 and 5.7 show that, for the integration method, the variation with n_{mat} and n_{vac} are larger than the variations with k and h. This is because the integration method includes the entire thin film, so, as the proportion of near-surface volume to bulk-like volume changes, the integration method mean inner potentials will change. The exact value of the mean inner potential from these methods will be examined in the subsequent section on material structure.

Mean inner potentials calculated using both the integration method and the subtraction method show deviations at low thicknesses, but the integration method also shows deviations for thicker thin-films. The integration method sums over the entire thickness, so the direct effect of the surface is included, and the proportion of

[34] Values of n_{mat} and n_{vac} used here are those corresponding to (2m+1) ML, where m is an integer, to yield identical surfaces.

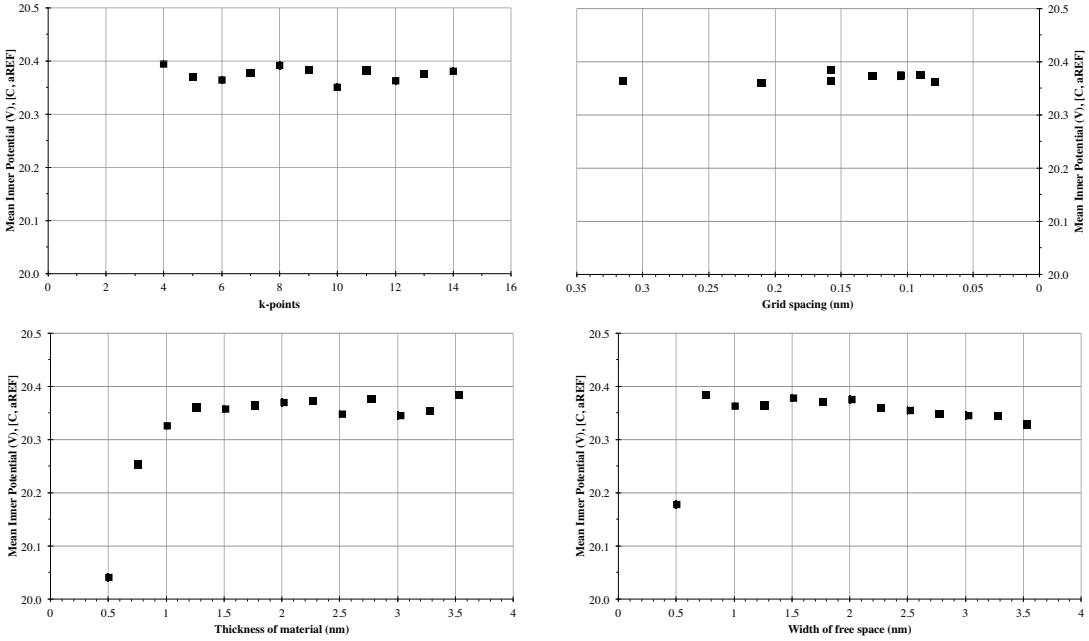


Figure 5.4: DFT mean inner potential precision determination for diamond C: aREF lattice parameter, FD mode, subtraction method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. Some parameters have larger effects than others. Some systematic effects are visible – the thinnest films have lower mean inner potentials than the films with thickness > 1.5 nm. The data in this figure is the source for the C aREF row in Figure 5.12, a precision-determination figure.

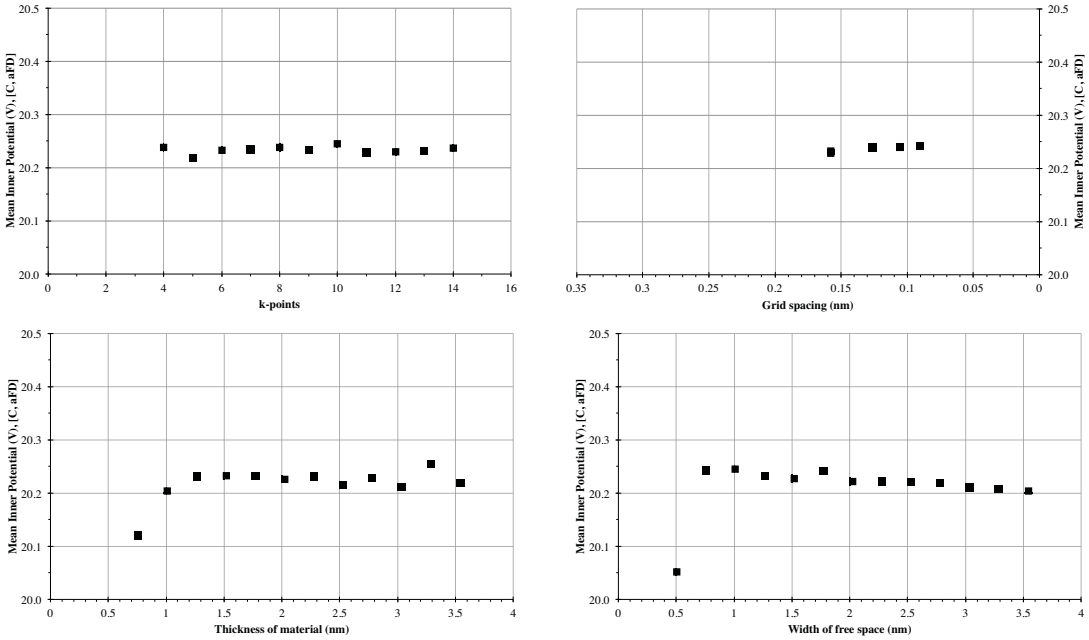


Figure 5.5: DFT mean inner potential precision determination for diamond C: aFD lattice parameter, FD mode, subtraction method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. The data in this figure is the source for the C aFD row in Figure 5.12, a precision-determination figure.

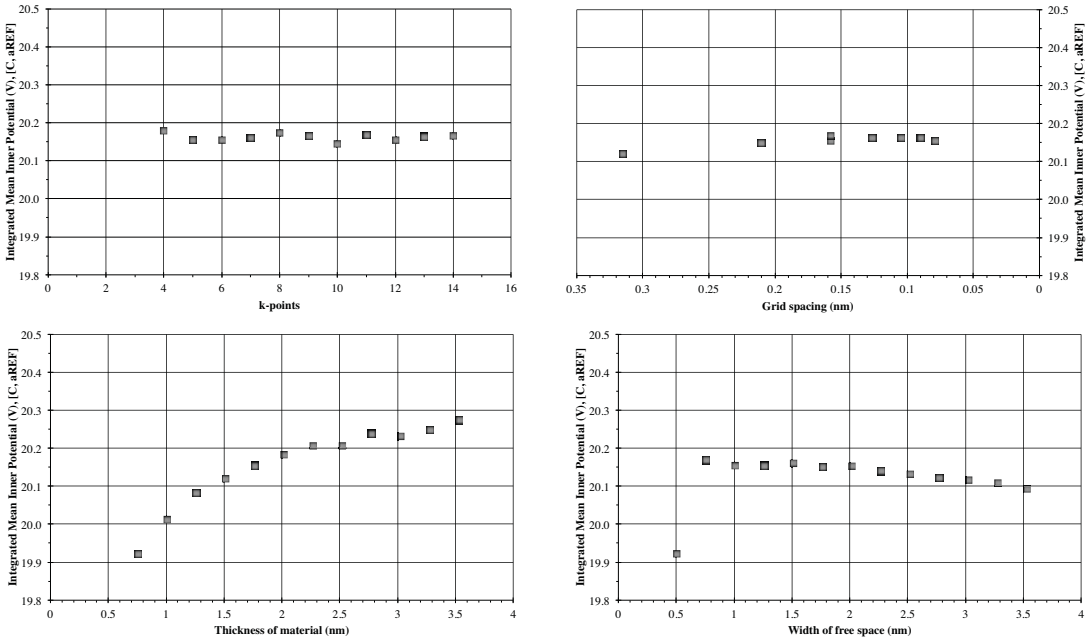


Figure 5.6: DFT mean inner potential precision determination for diamond C: aREF lattice parameter, FD mode, integration method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. Note upward slope on thickness graph not present in subtraction-method Figure 5.4. The data in this figure is the source for the C aREF row in Figure 5.13, a precision-determination figure.

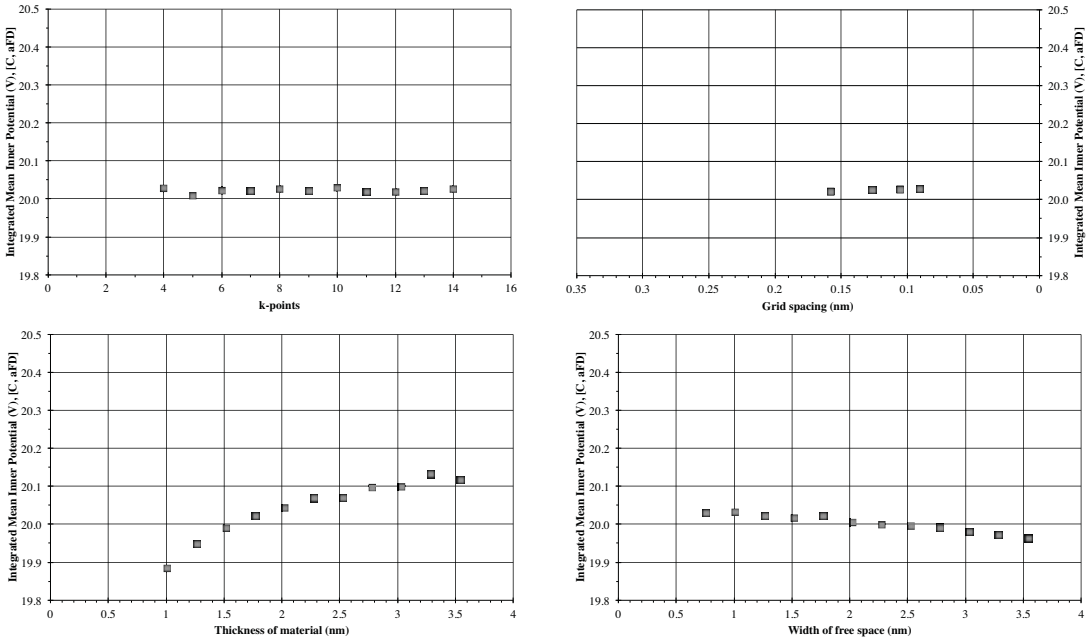


Figure 5.7: DFT mean inner potential precision determination for diamond C: aFD lattice parameter, FD mode, integration method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. Note upward slope on thickness graph not present in subtraction-method Figure 5.5. The data in this figure is the source for the C aFD row in Figure 5.13, a precision-determination figure.

surface relative to bulk-like material is decreased for thicker thin-films. However, the deviation at low thicknesses for the subtraction method arises because the thinner thin-films are thin enough that the middle of the thin-film is not bulk-like. The integration method is using the exact same DFT output data as the subtraction method, so the differences between them are systematic differences in mean inner potential determination. The integration method still uses the same operating envelope as the subtraction method, because the systematic differences between the integration method and the subtraction method are not random error in the DFT data, but inherent to the two methods.

For Si, Ge, and GaAs, the analogous test suite results are seen in Appendix D in FD mode. These simulations show that the same general trends hold for each individual element, and that the operating envelopes for n_{mat}, n_{vac}, k, and h determined from C are useful for all elements in the test suite. The same difference between the integration method and the subtraction method is noted. The collective numerical precision from these simulation parameters is addressed in the following Precision section.

LCAO Mean Inner Potential Test Suite Results

The mean inner potential test suite for LCAO mode involves the same multiple-parameter testing on the same materials as the FD mode mean inner potential test suite analyzed in the previous section. The results of the test suite for LCAO mode are functionally similar, yet different to the test suite results from FD mode.

As noted above, the LCAO mode dzp-basis minimum-energy lattice parameter is not tested here. The dzp-basis minimum-energy lattice parameter would be used for simulations of relaxed atom positions. However, none of these are carried out in this work. Instead, LCAO mode is used to compare and contrast with the FD mode error analysis. This provides a comparison with FD mode, which helps determine which is more accurate and contrast the margins of error of FD mode and LCAO mode.

In contrast to the FD mode results, the LCAO mode results are less precise. Diamond C is taken as representative of the mean inner potential test suite; the results for Si, Ge, and GaAs are visible in Appendix D. The subtraction method mean inner potential test suite results from diamond C in LCAO mode are in Figure 5.8 for the reference lattice parameter and Figure 5.9 for the FD minimum-energy lattice parameter. The integration method applied to the LCAO mode test suite for the mean inner potential in diamond C is seen in Figure 5.10 for the reference lattice parameter, and Figure 5.11 for the FD minimum-energy lattice parameter. Figures 5.8, 5.9, 5.10 and 5.11 show the same general behavior as their counterpart graphs from FD mode simulations, with some critical differences. Notably present in the LCAO mode diamond C graphs is dependence on grid spacing h (not as notable in the Si, Ge, or GaAs graphs in Appendix D), where decreased h yields increased mean inner potential. This h-dependent effect does not converge, and can be considered example of imprecision. Otherwise, the FD mode and LCAO mode test suite results are qualitatively similar. Therefore, the same operating envelope is used for LCAO mode as FD mode. The results from the simulations for Si, Ge, and GaAs are seen in Appendix D.

The test suite results for LCAO mode and FD mode are compiled to determine the overall precision of the simulations, which is examined in the next subsection.

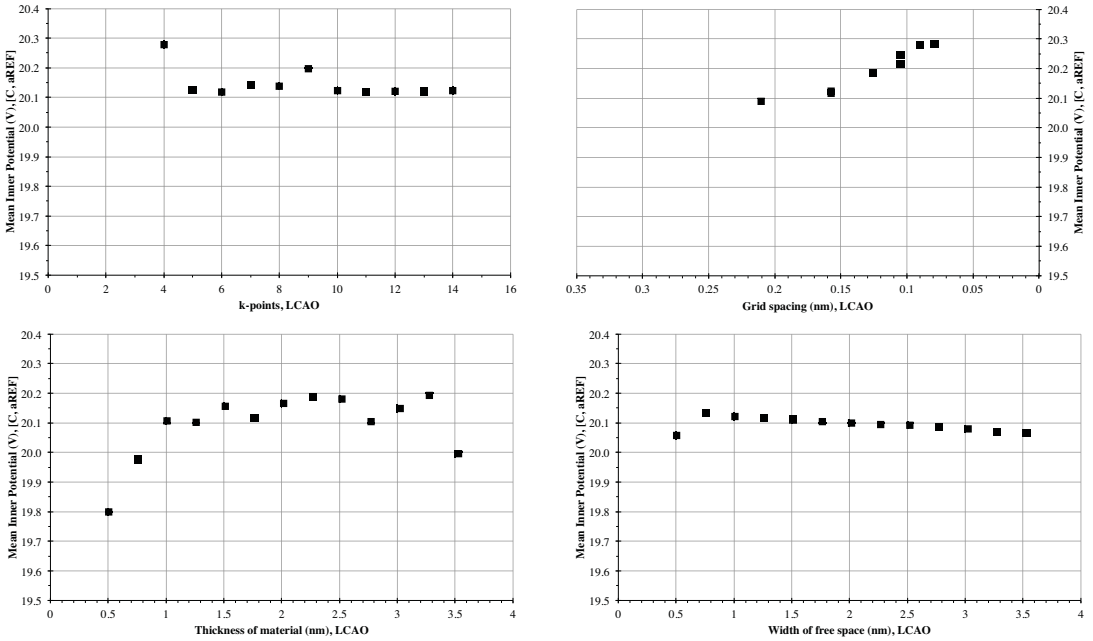


Figure 5.8: DFT mean inner potential precision determination for diamond C: aREF lattice parameter, LCAO mode, subtraction method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. Note decreased precision as compared to FD mode (Figure 5.4). The data in this figure is the source for the C aREF row in Figure 5.14, a precision-determination figure.

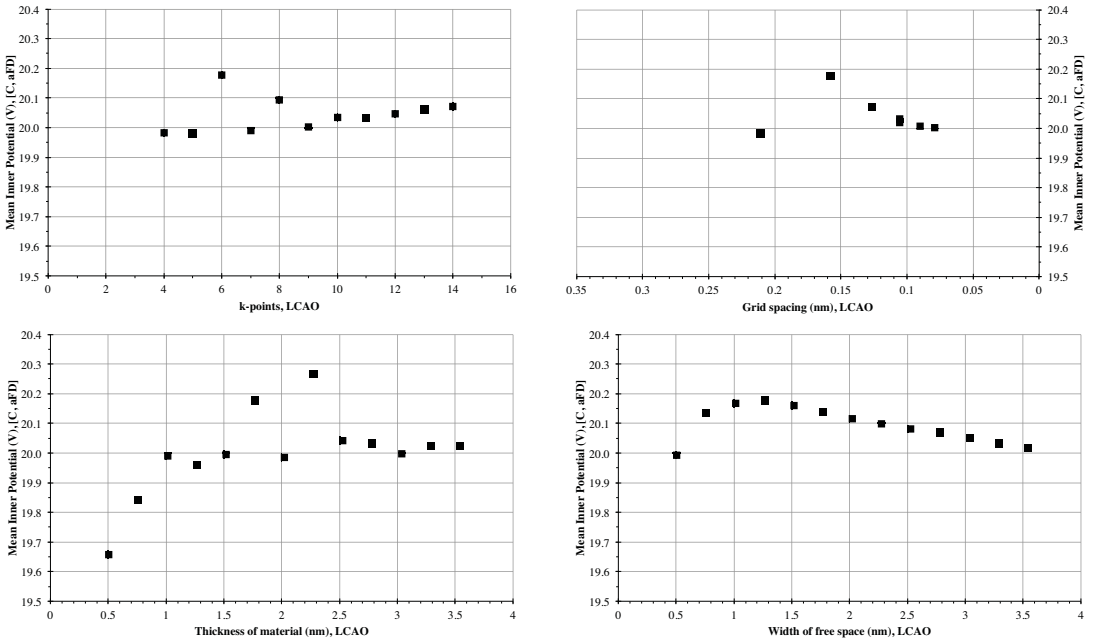


Figure 5.9: DFT mean inner potential precision determination for diamond C: aFD lattice parameter, LCAO mode, subtraction method. By comprehensively testing the effect on the mean inner potential by varying different simulation parameters, the operating envelope of the GPAW code can be determined. Note decreased precision as compared to FD mode (Figure 5.5). The data in this figure is the source for the C aFD row in Figure 5.14, a precision-determination figure.

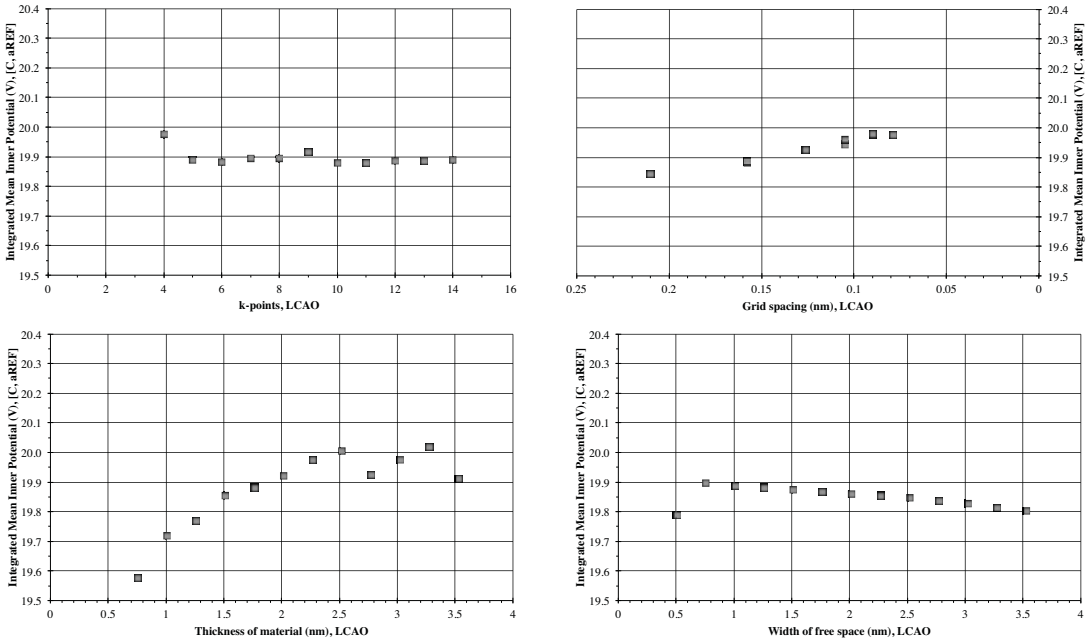


Figure 5.10: DFT mean inner potential precision determination for diamond C: aREF lattice parameter, LCAO mode, integration method. Compare to Figure 5.3 and Figure 5.6. Note decreased precision as compared to FD mode (Figure 5.6), and integration-method-caused thickness-related imprecision. The data in this figure is the source for the C aREF row in Figure 5.15, a precision-determination figure.

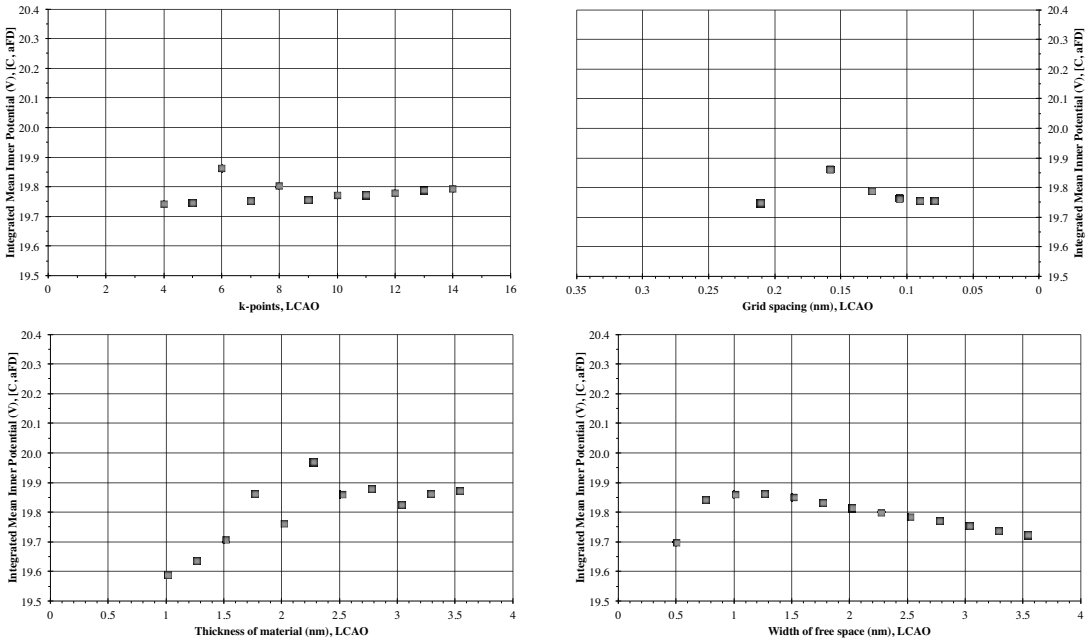


Figure 5.11: DFT mean inner potential precision determination for diamond C: aFD lattice parameter, LCAO mode, integration method. Compare to Figure 5.4 and Figure 5.7. Note decreased precision as compared to FD mode (Figure 5.7), and integration-method-caused thickness-related imprecision (*cf*. Figure 5.4). The data in this figure is the source for the C aFD row in Figure 5.15, a precision-determination figure.

Precision

The precision of the DFT simulations of the mean inner potential can be determined from the mean inner potential test suites above. The mean inner potential results from simulations inside the operating envelope are taken as an ensemble, and the standard deviation of this ensemble allows for quantitative precision determination. First, FD mode is analyzed, followed by LCAO mode. Precision values for unknown systems are then derived from the average of the precisions of the systems examined here. The mean inner potential test suite results shown in Appendix D are included in this section.

From the FD test suite mean inner potentials inside the operating envelope, Figure 5.12 shows the subtraction-method results and Figure 5.13 shows the integration-method results. The n_{mat}, n_{vac}, k and h columns are derived from the operating envelopes detailed above, and are 1 standard deviation (σ), assuming a normal distribution. Color has been added in two ways, independently for each figure: first, the n_{mat}, n_{vac}, k and h columns in each table are grouped and colored together, such that dark green is the lowest (best) value from that set, yellow the middle value, and dark red the highest (worst). The total FD 2σ margins of error – adding the n_{mat}, n_{vac}, k and h errors in quadrature – are grouped and colored separately, using green for the lowest (best) value from the set and red for the highest (worst).[35]

This identifies the relative sources of error. For the subtraction-method confidence intervals in Figure 5.12, the thickness of material and width of vacuum provide the most imprecision, while the h-spacing and k-points provide the least. In the integration-method confidence intervals in Figure 5.13, the thickness of material becomes a much larger source of imprecision error than for the subtraction-method, and the h-spacing and k-points remain the lowest sources of imprecision. This increase in error due to the thickness is expected, given the integration-mode direct inclusion of surface effects.

In Figures 5.12 and 5.13, the 2σ overall margin of error corresponds to the 95% confidence interval of the mean inner potential in the operating envelope, in V, calculated from n_{vac}, n_{mat}, k, and h. For the subtraction-method mean inner potentials in FD mode, the confidence intervals range from ±0.040 V for C with a_{FD} to ±0.082 V for Ge with a_{REF}. The materials with the greatest differences in a_{REF} and a_{FD} lattice parameters also show the greatest imprecisions with subtraction method mean inner potential determination. In contrast, not only is this relationship broken when the integration method is used, but the margins of error are much higher: from ±0.066 V for GaAs with a_{FD} to ±0.131 V for C with a_{REF}. The GaAs-a_{FD} 2σ margin of error is actually a slight decrease in mean inner potential imprecision using integration-method relative to subtraction-method, but the C-a_{REF} 2σ margin of error is 160% more imprecise relative to the subtraction method. This increase for the integration method error is due to the general increase in both n_{mat} and n_{vac} imprecision error using the integration method relative to the subtraction method. The n_{mat} imprecision is unsurprising, given the change in the surface-to-bulk ratios using the integration method, but the n_{vac} imprecision is unexpected, as the vacuum region is a flat-potential region. The latter requires more investigation to determine the exact cause,

[35] These exclude error derived from choosing the PBE exchange-correlation functional, as simulating using another exchange-correlation functional would be a fundamentally different simulation, and also relatively few exchange-correlation functionals are implemented in the GPAW program, making statistical analysis difficult.

FD Mode Subtraction-Method MIP Confidence Intervals

Material	a (Å)	a type	n_{mat} (V)	n_{vac} (V)	h (V)	k (V)	Total 2σ
C	3.567	aREF	0.013	0.016	0.008	0.013	**0.051**
C	3.577	aFD	0.012	0.014	0.005	0.007	**0.040**
Si	5.431	aREF	0.013	0.017	0.006	0.008	**0.047**
Si	5.477	aFD	0.017	0.017	0.003	0.007	**0.050**
Ge	5.658	aREF	0.035	0.017	0.005	0.012	**0.082**
Ge	5.765	aFD	0.025	0.025	0.010	0.009	**0.075**
GaAs	5.653	aREF	0.019	0.023	0.008	0.017	**0.071**
GaAs	5.773	aFD	0.031	0.017	0.015	0.011	**0.080**

Figure 5.12: Mean inner potential confidence intervals, FD mode and subtraction-method. For the data above and in Appendix D, the standard deviation of the mean is calculated for each parameter, then the individual parameters added in quadrature and doubled to yield the 2σ confidence interval for the mean inner potential (MIP). The elements with smaller lattice parameter mismatch (Figure 5.2) show better confidence intervals. Green means most precise, red means least precise. Middle four columns use one color range, "Total 2σ" column uses different color range.

FD Mode Integration-Method MIP Confidence Intervals

Material	a (Å)	a type	n_{mat} (V)	n_{vac} (V)	h (V)	k (V)	Total 2σ
C	3.567	aREF	0.060	0.023	0.005	0.010	**0.131**
C	3.577	aFD	0.058	0.023	0.003	0.006	**0.125**
Si	5.431	aREF	0.031	0.028	0.004	0.006	**0.086**
Si	5.477	aFD	0.031	0.029	0.002	0.005	**0.086**
Ge	5.658	aREF	0.049	0.029	0.003	0.009	**0.115**
Ge	5.765	aFD	0.035	0.034	0.006	0.007	**0.098**
GaAs	5.653	aREF	0.023	0.031	0.005	0.011	**0.081**
GaAs	5.773	aFD	0.017	0.026	0.009	0.008	**0.066**

Figure 5.13: Mean inner potential confidence intervals, FD mode and integration-method. As with Figure 5.12, the data above and in Appendix D is used to calculate the standard deviation of the mean for each parameter, then the individual parameters added in quadrature and doubled to yield the 2σ confidence interval for the mean inner potential (MIP). The elements with smaller lattice parameter mismatch (Figure 5.2) show worse confidence intervals for integration-method mean inner potentials. Green means most precise, red means least precise. Middle four columns use one color range, "Total 2σ" column uses different color range. Compared with Figure 5.12, the integration-method mean inner potentials generally show higher variation due to thin-film thickness and vacuum width, leading to less precise mean inner potentials.

but Figure 5.7 shows a systematic decline in potential for larger vacuum widths. In short, for FD mode, while subtraction-method precision is ±0.040-±0.082 V, integration-method precision is ±0.066-±0.131 V due to systematic effects. Therefore, the subtraction method is more precise for FD mode mean inner potential determination, and will be used in the following sections.

Comparing LCAO mode precision to FD mode precision shows that LCAO mode simulations are generally less precise than their FD mode counterparts, and that, in LCAO mode, using the integration method for mean inner potential determination is generally slightly more precise than the subtraction method. Figure 5.14 shows the 95% confidence intervals for the subtraction method of mean inner potential determination from LCAO mode simulations. From Figure 5.14, the most precise LCAO mode mean inner potential is the Si mean inner potential using the a_{FD} lattice parameter (±0.102 V), and the least precise is the C value using the a_{FD} lattice parameter (±0.291 V). Comparing Figure 5.14 with the FD mode counterpart in Figure 5.12, the LCAO mode margins of error are much larger than the FD mode margins of error. This indicates that LCAO mode of mean inner potential simulation is much less precise than FD mode for subtraction-method mean inner potential simulations. For LCAO mode integration-method mean inner potential margins of error, Figure 5.15 shows that the precision in LCAO mode improves relative to the subtraction method margins of error (unlike FD mode), yielding a narrowest-margin-of-error that is barely changed from the subtraction-method calculations, at ±0.105 V (but for GaAs with a_{REF}, which had a precision of ±0.152 V with the subtraction-method), and a worst precision of ±0.247 V (as compared to ±0.291 V) for C using the a_{FD} lattice parameter. This is perhaps due to the summation over the entire thin-film yielding better statistics than the summation over just the middle of the thin-film. However, the precision of integration-method LCAO mode is still far worse than the precision of the integration-method FD mode. In summary, LCAO mode is less precise than FD mode, for both the integration-method determination and the subtraction-method determination.

For materials that are not part of the test suite, an average precision value must be determined. For known systems, the precision is not clearly correlated with the overall mean inner potential values (see the following section). Therefore, it is reasonable to assume that an average of the 2σ values for each mean inner potential method and simulation mode is valid for an unknown similar system, such as the ones shown here. This yields a precision in mean inner potential determination for an unknown system of:

- ± 0.06 V for subtraction-method FD mode simulations,
- ± 0.10 V for integration-method FD mode simulations,
- ± 0.19 V for subtraction-method LCAO mode simulations, and
- ± 0.16 V for integration-method LCAO mode simulations.

Therefore, the most precise FD mode DFT simulation with subtraction-method mean inner potential determination is used in this work, but precision will be accounted for in all DFT simulations. This also reinforces the conclusion that the choice of DFT program and parameters affects the simulated mean inner potential precision, and the DFT simulation parameters involved should be investigated thoroughly to determine precision.

LCAO Mode Subtraction-Method MIP Confidence Intervals

Material	a (Å)	a type	n_{mat} (V)	n_{vac} (V)	h (V)	k (V)	Total 2σ
C	3.567	aREF	0.059	0.021	0.069	0.050	**0.210**
C	3.577	aFD	0.095	0.054	0.077	0.058	**0.291**
Si	5.431	aREF	0.033	0.029	0.031	0.031	**0.124**
Si	5.477	aFD	0.034	0.030	0.019	0.014	**0.102**
Ge	5.658	aREF	0.048	0.031	0.033	0.051	**0.166**
Ge	5.765	aFD	0.104	0.036	0.036	0.036	**0.242**
GaAs	5.653	aREF	0.069	0.028	0.001	0.016	**0.152**
GaAs	5.773	aFD	0.101	0.051	0.001	0.025	**0.231**

Figure 5.14: Mean inner potential confidence intervals, LCAO mode and subtraction-method. As with Figure 5.12, the data above and in Appendix D is used to calculate the standard deviation of the mean for each parameter, then the individual parameters added in quadrature and doubled to yield the 2σ confidence interval for the mean inner potential (MIP). Green means most precise, red means least precise. Middle four columns use one color range, "Total 2σ" column uses different color range. Compared with Figure 5.12, the LCAO mode mean inner potentials generally show higher variation due to thin-film thickness and vacuum width, leading to less precise mean inner potentials.

LCAO-Mode Integration-Method MIP Confidence Intervals

Material	a (Å)	a type	n_{mat} (V)	n_{vac} (V)	h (V)	k (V)	Total 2σ
C	3.567	aREF	0.076	0.030	0.040	0.028	**0.191**
C	3.577	aFD	0.096	0.049	0.049	0.034	**0.247**
Si	5.431	aREF	0.045	0.027	0.021	0.019	**0.119**
Si	5.477	aFD	0.037	0.035	0.011	0.008	**0.106**
Ge	5.658	aREF	0.050	0.036	0.020	0.034	**0.146**
Ge	5.765	aFD	0.074	0.035	0.021	0.025	**0.177**
GaAs	5.653	aREF	0.038	0.035	0.001	0.010	**0.105**
GaAs	5.773	aFD	0.064	0.045	0.000	0.016	**0.159**

Figure 5.15: Mean inner potential confidence intervals, LCAO mode and subtraction-method. As with Figure 5.12, the data above and in Appendix D is used to calculate the standard deviation of the mean for each parameter, then the individual parameters added in quadrature and doubled to yield the 2σ confidence interval for the mean inner potential (MIP). Green means most precise, red means least precise. Middle four columns use one color range, "Total 2σ" column uses different color range. Compared with Figure 5.14, the integration-method mean inner potentials appear about as imprecise as the subtraction-method ones in LCAO mode.

The precision yields acceptance/rejection thresholds for comparing mean inner potential simulations. If two simulations of different specimens (*e.g.* with and without a surface layer) differ by more than the margin of error, it is likely that the mean inner potentials of the two systems are different. Thus, with the precision of the DFT simulations determined, this precision is applied to evaluate simulations of the mean inner potential itself.

Baseline Mean Inner Potential Simulations: (110) Surfaces

In this section, a set of baseline DFT mean inner potential simulations is established. DFT simulations in FD mode are processed to yield mean inner potentials using both the subtraction method (to establish the baseline) and the integration method (for comparison). First, the effect on the mean inner potential of material and crystal structure is investigated. For different crystal structures, diamond and zincblende structures for the unary and binary compounds, respectively, are examined, and then the wurtzite structure is compared with the zincblende results.

All simulations in this section are thin-films using a [110] surface normal (in the cubic system); for the III-V material systems, this surface contains equal amounts of both materials, thereby avoiding built-in slopes in vacuum. FD mode calculations are used to provide the most precise mean inner potential calculations.

Materials Using Diamond and Zincblende Structures

The mean inner potentials for the full set of materials under study are now simulated in the diamond and zincblende structures, establishing the baseline for the DFT mean inner potential simulations in this chapter. The materials listed in Figure 5.2 are divided into two groups, depending on whether or not the material was one of the four in the mean inner potential test suite: the materials in the test suite (C, Si, Ge, and GaAs) are used for both mean inner potential determination and conclusions about the DFT method, and the other materials (InP, InAs, GaP, AlP, and AlAs) are used only for mean inner potential determination. The materials that were not part of the test suite use the average error for the mean inner potential determination method and simulation mode determined in the previous section on precision from the full test suite; the materials that were part of the test suite use the predetermined errors for the method and mode determined for the material and lattice parameter.

The subtraction-method results, in Figure 5.16, exhibit several traits about the mean inner potential. First, the mean inner potential is material-dependent, as expected from the literature and isolated-atom scattering factors. As a result, the mean inner potential is critical to determine for a given material. However, the mean inner potentials in Figure 5.16 are all similar: apart from diamond C (which is, for a_{FD}, 20.23 ± 0.04 V and, for a_{REF}, 20.36 ± 0.05 V), the mean inner potentials in Figure 5.16 range from 11.11 ± 0.06 V for AlP (a_{FD}) to 14.73 ± 0.08 V for Ge (a_{REF}). This reinforces the conclusion drawn initially from the isolated-atom scattering-factor data that the mean inner potential should be known for each material.

Second, Figure 5.16 shows that the mean inner potential is lattice-parameter dependent, and larger lattice parameters yield lower mean inner potentials. As noted earlier, mean inner potentials calculated from isolated-atom scattering factors scale as the volume (*i.e.* the cubed lattice parameter, a^3). The lattice-parameter dependence of the DFT simulations is shown in the last two columns in Figure 5.16: MIP×a^3 is the mean inner potential multiplied by the cubed lattice parameter. If changing the lattice

FD Mode Subtraction-Method Mean Inner Potentials (D/ZB Structure)						
Material	**a (Å)**	**a type**	**MIP (Sub) (V)**	**MIP % Error**	**MIP × a^3**	**Δ**
C	3.567	aREF	20.36 ± 0.05	0.25%	924.192	
C	3.577	aFD	20.23 ± 0.04	0.20%	926.009	0.2%
Si	5.431	aREF	12.59 ± 0.05	0.37%	2017.566	
Si	5.477	aFD	12.32 ± 0.05	0.41%	2023.788	0.3%
Ge	5.658	aREF	14.73 ± 0.08	0.55%	2667.130	
Ge	5.765	aFD	13.88 ± 0.07	0.54%	2659.568	-0.3%
GaAs	5.653	aREF	14.18 ± 0.07	0.50%	2562.151	
GaAs	5.773	aFD	13.27 ± 0.08	0.60%	2553.737	-0.3%

Material	**a (Å)**	**a type**	**MIP (Sub) (V)**
InP	5.869	aREF	13.87 *± 0.06*
InP	5.908	aFD	13.58 *± 0.06*
InAs	6.059	aREF	14.18 *± 0.06*
InAs	6.100	aFD	13.91 *± 0.06*
GaP	5.490	aREF	13.28 *± 0.06*
GaP	5.545	aFD	12.91 *± 0.06*
AlP	5.451	aREF	11.46 *± 0.06*
AlP	5.513	aFD	11.11 *± 0.06*
AlAs	5.662	aREF	12.35 *± 0.06*
AlAs	5.740	aFD	11.85 *± 0.06*

Figure 5.16: FD mode subtraction-method baseline mean inner potentials (MIP) for the diamond or zincblende structure of unary or binary materials, respectively. These are calculated for the "baseline" scenario discussed in the text (110 surface). Italicized error values use the average error from the test suite for the method. The **Δ** column compared to the "MIP % Error" column shows that the bonding effects on the mean inner potential of using aFD *vs*. aREF are indistinguishable from the effect of the change in volume of the unit cell.

parameter over the range studied is the sole source of the mean inner potential change in the DFT simulations in Figure 5.16, MIP×a^3 should be constant for each material. The difference column, in the far right of Figure 5.16, shows the difference between the a_{FD} and a_{REF} MIP×a^3 values. This is at or under the margin of error (also seen as a percentage in Figure 5.16), meaning that charge-distribution effects from changing the lattice parameter up to 2% (for GaAs) do not bleed through into the mean inner potential for this simulation. Larger changes are not tested, and, in this range (of ±2%), it is sensible to assume that the DFT-simulated mean inner potential scales inversely with unit cell volume.

Third, in Figure 5.16, the margin of error in the mean inner potential ranges from 0.20%-0.60%. As noted above, FD mode with subtraction-method determination is the most precise of the DFT mean inner potential determination methods studied in this work, and is comparable to the margin of error reported in previous DFT simulations [50]. The 0.20%-0.60% margin of error in Figure 5.16 is quite favorable when compared to the experimental mean inner potential determination in Chapter 4, which had a margin of error of ~2%. Thus, DFT simulation can yield more precise mean inner potentials than experiment.

Fourth, the DFT-simulated mean inner potential values in Figure 5.16 are lower than the isolated-atom scattering factor mean inner potentials in Figure 5.1. This is due to the charge redistribution and bonding effects in the crystal, accounted for in the mean inner potential simulated using DFT but not the one simulated from isolated-atom scattering factors, as noted previously in [50].

The FD mode integration-method mean inner potentials are seen in Figure 5.17. These follow the same trends seen in the FD mode subtraction-method analysis, if less precisely. The difference between integration and subtraction modes, given as the "Sub – Int (V)" column, is higher for the group IV elements compared to the III-V compounds, but is generally less than the margin of error in the integration-method mean inner potential. This discrepancy between the subtraction and integration methods arises from the surface inclusion in the integration method, perhaps indicating that thicker specimens are preferred for experimental mean inner potential measurement to minimize the direct surface contribution to the measured phase shift, especially for group-IV materials.

In general, DFT simulations of zincblende and diamond structures for group-IV and group III-V materials yield high-precision mean inner potential values on known systems. The next section examines the effect of using a wurtzite structure instead of zincblende, to see what effect the crystal structure has on simulations using the same materials.

Materials Using Wurtzite Structures

The mean inner potentials of wurtzite structure analogs of the III-V zincblende structures detailed above are analyzed in this section. Only FD mode DFT simulations with subtraction-method mean inner potentials are used to investigate if there is an effect on the mean inner potential for these materials from using a different crystal structure.

The results of the wurtzite-structure simulations can be seen in Figure 5.18. The effect of crystal structure on the mean inner potential is minimal. For almost all of the III-V materials studied here, the differences between zincblende and wurtzite structures are

FD Mode Integration-Method Mean Inner Potentials (D/ZB Structure)

Material	a (Å)	a type	MIP (Int) (V)	Sub - Int (V)	MIP % Error	MIP*a^3	Δ
C	3.567	aREF	20.15 ± 0.13	0.210	0.65%	914.670	
C	3.577	aFD	20.02 ± 0.12	0.212	0.62%	916.324	0.2%
Si	5.431	aREF	12.50 ± 0.09	0.092	0.69%	2002.858	
Si	5.477	aFD	12.23 ± 0.09	0.092	0.70%	2008.719	0.3%
Ge	5.658	aREF	14.60 ± 0.12	0.124	0.79%	2644.679	
Ge	5.765	aFD	13.78 ± 0.10	0.105	0.71%	2639.390	-0.2%
GaAs	5.653	aREF	14.12 ± 0.08	0.059	0.57%	2551.468	
GaAs	5.773	aFD	13.24 ± 0.07	0.036	0.50%	2546.874	-0.2%

Material	a (Å)	a type	MIP (Int) (V)	Sub – Int (V)
InP	5.869	aREF	13.82 *± 0.10*	0.051
InP	5.908	aFD	13.54 *± 0.10*	0.045
InAs	6.059	aREF	14.15 *± 0.10*	0.038
InAs	6.100	aFD	13.87 *± 0.10*	0.042
GaP	5.490	aREF	13.23 *± 0.10*	0.052
GaP	5.545	aFD	12.86 *± 0.10*	0.055
AlP	5.451	aREF	11.45 *± 0.10*	0.008
AlP	5.513	aFD	11.11 *± 0.10*	0.004
AlAs	5.662	aREF	12.31 *± 0.10*	0.033
AlAs	5.740	aFD	11.83 *± 0.10*	0.017

Figure 5.17: FD mode integration-method baseline mean inner potentials (MIP) for the diamond or zincblende structure of unary or binary materials, respectively. These are calculated for the "baseline" scenario discussed in the text (110 surface). Italicized error values use the average error from the test suite for the method. The "Sub – Int (V)" column shows that the difference between the integration- and subtraction- methods (Figure 5.16) for the mean inner potential of the baseline simulation are less than or approximately the margin of error (except for C). The **Δ** column compared to the "MIP % Error" column shows that the bonding effects on the mean inner potential of using aFD *vs*. aREF are indistinguishable from the effect of the change in volume of the unit cell, the same as for the subtraction-method shown in Figure 5.16.

FD Mode Subtraction-Method Mean Inner Potentials (WZ Structure)				
Material	**a (Å)**	**a type**	**MIP (Sub) (V)**	**ZB – WZ (V)**
GaAs	5.653	aREF	14.13 ± 0.07	+0.05
GaAs	5.773	aFD	13.32 ± 0.08	+0.05
Material	**a (Å)**	**a type**	**MIP (Sub) (V)**	**ZB – WZ (V)**
InP*	5.869	aREF	13.88 ± *0.06*	+0.01
InP*	5.908	aFD	13.61 ± *0.06*	+0.03
InAs*	6.059	aREF	14.24 ± *0.06*	+0.06
InAs*	6.100	aFD	13.94 ± *0.06*	+0.03
GaP	5.490	aREF	13.28 ± *0.06*	+0.00
GaP	5.545	aFD	12.88 ± *0.06*	-0.03
AlP	5.451	aREF	11.46 ± *0.06*	-0.01
AlP	5.513	aFD	11.12 ± *0.06*	+0.01
AlAs	5.662	aREF	12.32 ± *0.06*	-0.01
AlAs	5.740	aFD	11.87 ± *0.06*	+0.02

*: Calculation used n_{mat}=11 and n_{vac}=5 (staying within the operating envelope) due to memory constraints

Figure 5.18: Mean inner potentials from wurtzite structures compared to zincblende structures. Using FD mode and the subtraction method, mean inner potentials can be compared between wurtzite and zincblende structures (*cf*. Figure 5.17) of the same material. Non-polar orientations are used. When this is done, the differences are within the margin of error of the simulations, showing that the effect of wurtzite vs. zincblende crystal structure is negligible for GPAW and these materials.

within the margin of error, and no example is far outside it. Therefore, it can be concluded that the effect on the mean inner potential of changing from a zincblende to a wurtzite structure is minimal for the set of materials studied in Figure 5.18.

The Effect of Structure and Material

The DFT simulations of mean inner potential provide the baseline for the following sections. The effect of material is clear from Figure 5.16: Different materials have different mean inner potentials, as determined from DFT simulations, and as expected. Equally clearly, from Figure 5.18, the effect on the mean inner potential due to wurtzite or zincblende structure for the materials studied here is within the margin of error of the DFT simulation.

DFT simulations of the mean inner potential with the margin of error seen in Figure 5.16 of 0.20%-0.60% are much more precise than the experimental measurement in the previous chapter. For comparison, the experimental mean inner potential determination in Chapter 4 had an error of ±~2%. Chapter 4 also showed that the effect on the mean inner potential of using RENOIR or DT thickness determinations was ~4%, and the analysis in chapters 3 and 4 was not able to determine which thickness determination was more accurate. Thus, in principle, DFT simulations of mean inner potentials should distinguish between different thickness determinations.[36]

In addition, surface effects are a clear source of possible inaccuracy in mean inner potential determination. Basic surface effects are visible in the difference between the integration-method and subtraction-method mean inner potentials, where including the surface leads to mean inner potential measurement effects, and also in the thinnest slabs in the test-suite section, where the middle of the specimen is not sufficiently bulk-like. Therefore, the next section investigates possible surface effects on the mean inner potential: from the crystal facet forming the surface, from surface reconstruction or relaxation, or from surface adsorbates.

Surface Effects

How To Investigate Surface Effects?

The following sections aim to determine how the surface influences the mean inner potential. Using the subtraction method for mean inner potential determination, the mean inner potential is measured in the bulk of the material relative to vacuum outside the material. Thus, are there effects from the surface between material and vacuum? If so, which surface configuration effects are the largest? This requires changing the surface in defined ways as follows:

- **Surface Facet:** Changing from a (110) surface facet to a (001), (111) or (211) facet.
- **Relaxation:** Changing from a bulk-terminated surface to a clean surface with structurally optimized atomic positions (relaxing all atomic positions), possibly with higher symmetry (*e.g.* (2×1)).
- **Adsorbates:** Changing from a bulk-like surface to a surface with adsorbate atoms in structurally optimized positions (relaxing all atomic positions).

[36] However, a difference between the specimen and reference lattice parameters is a clear potential source of inaccuracy in the mean inner potential determination. This point is revisited later in this work.

- **Reconstruction:** Changing from a bulk-like surface to a fixed, experimentally derived surface configuration from the literature, possibly with higher symmetry (*e.g.* (2×1)).

These basic components are also combined (any of these can be combined with the Surface Facet case listed above):

- **Relaxed Reconstruction:** Starting from a Reconstruction surface configuration with experimentally determined atomic positions, and relaxing all atomic positions to obtain a structurally optimized end state.
- **Reconstruction with Adsorbates:** As Reconstruction, with a different surface material and all atom positions determined from the literature.
- **Relaxed Reconstruction with Adsorbates:** Starting from a Reconstruction with Adsorbates surface configuration, and proceeding to a relaxed end state, including relaxing all atoms.

Simulations are compared to the (110) baseline simulations from the previous section, which provide a standard for comparison.

What Constrains Surface Effect Investigations In This Work?

Several restrictions are in place for these simulations. Binary compounds are only examined in the (110) non-polar[37] orientation to avoid sloping potentials in free space. Systems involving relaxation generally only use the minimum-energy lattice parameter – a_{FD} for FD mode – to avoid building strain into the simulated system. However, reference lattice parameters are generally used for comparisons with the literature positions. For diamond C, both a_{REF} and a_{FD} lattice parameters are used, due to the previously noted small mismatch between them. In all cases, lattice parameters used are clearly indicated. FD mode simulations are used because of their increased precision. Structure optimization, as noted above, are performed by using a quasi-Newton force minimizer and minimizing the force on each atom to <0.05 eV/Ångström.

These questions of surface effects are explored through two different types of mean inner potential simulation: systematic *ab initio* structurally optimized simulations, and selected experimentally determined surface configurations. Systematic *ab initio* structurally optimized simulations provide the fundamental testing of whether the targeted surface parameters have an effect by testing them one at a time relative to established baseline values using minimum-energy atomic positions. Selected surface configurations from the literature – either full atomic positions or only atomic stoichiometry – provide the additional knowledge that surface conditions related to physically possible configurations are tested.

Mean inner potentials determined using surface-sensitive systematic *ab initio* structurally optimized DFT simulations are more important for this work than experimentally determined surface configurations, but both types are important. The DFT structural optimization used in this chapter is an identical procedure for all systems, which provides a basis for comparison that is internally self-consistent. The DFT structural optimization is self-consistent because each surface configuration is determined in the same fashion. However, the optimized structure may not be the

[37] The non-polar orientations for the III-V compounds are where the stoichiometry of the surface is balanced equally between the group III- and group V-components.

ideal structure. In contrast, the literature reports of experimentally determined surface conditions are dependent on experimental conditions and preparation, which will likely be different for every system in the literature, precluding a direct comparison between two different experimental surface configurations. In addition, the error in experimental atom position determination will likely be different for each specimen presented in the literature.

Another reason to use structural optimization instead of solely experimental atom positions is the effect of the TEM electron beam on the surface. The surface state of a TEM specimen is vulnerable to dynamic modification from beam-induced atomic displacement or contamination depositing during beam exposure. Therefore, the exact surface state of the TEM specimen will likely be difficult to determine, and may not be the same as the one determined using other techniques. Overreliance on fixed surface atom positions determined using other techniques would pose problems for the analysis. This is less of a concern for *ab initio* structurally optimized systems in the following sections, because the objective is to examine systematically how surface effects modify the mean inner potential.

The objective of the following sections is to examine how systematic changes in the surface affect the mean inner potential. From these effects, it is possible to posit what factors may affect the mean inner potential measured experimentally.

Systematic *ab initio* Surface Effects

Different Surface Facets, Unrelaxed

Surface orientation is the logical first parameter to be examined, because it is the clearest theoretically: does changing the surface facet but retaining the same bulk-like surface termination change the mean inner potential in a statistically significant way? Because the surface facets used in this section is bulk-terminated, the mean inner potentials determined in this section do not rely on any structural optimization or surface atom position determination.

Due to the use of non-(110) surface orientations, only C, Si, and Ge are used to investigate the effect on the mean inner potential of surface facet. Only bulk-terminated (non-relaxed) surfaces are examined in this section; relaxed surfaces are examined in the following section.

Figure 5.19 shows mean inner potentials from (001), (111) and (211) surfaces, calculated using the subtraction method and FD mode. For the shift from a (110) orientation to an (001) orientation, C, Si, and Ge all show increases in the mean inner potential, up to +1.05V in the case of C with a_{REF}, but also +0.30 V for Ge for both lattice parameters and +0.31 or +0.54 for Si. These shifts are well outside the margin of error, and indicate that the mean inner potential is surface facet dependent.

However, Figure 5.19 also shows that the change in mean inner potential with surface facet is material-dependent. While C, Si, and Ge all show changes in (001) orientations relative to (110), only C shows a change in (111) or (211) orientations relative to (110). The small mean inner potential shift in these simulations of Si and Ge mean inner potentials is not statistically significant relative to the (110) orientation baseline. Therefore, a change in surface facet does not always lead to a change in mean inner potential.

Surface-Facet Modification Effect on Mean Inner Potentials
FD-Mode Subtraction-Method, D/ZB Structure

Material	a (Å)	a type	Surface Facet	Relaxed?	MIP (V)	Δ from (110) baseline
C	3.567	aREF	(001)	No	21.41 ± 0.05	**+1.05 V**
C	3.577	aFD	(001)	No	21.26 ± 0.04	**+1.04 V**
Si	5.431	aREF	(001)	No	12.90 ± 0.05	**+0.31 V**
Si	5.477	aFD	(001)	No	12.85 ± 0.05	**+0.54 V**
Ge	5.658	aREF	(001)	No	15.03 ± 0.08	**+0.30 V**
Ge	5.765	aFD	(001)	No	14.17 ± 0.07	**+0.30 V**
C	3.567	aREF	(001)	Yes	21.42 ± 0.05	**+1.06 V**
C	3.577	aFD	(001)	Yes	21.27 ± 0.04	**+1.05 V**
Si	5.477	aFD	(001)	Yes	12.90 ± 0.05	**+0.59 V**
C	3.567	aREF	(111)	No	21.18 ± 0.05	**+0.82 V**
C	3.577	aFD	(111)	No	21.04 ± 0.04	**+0.82 V**
Si	5.431	aREF	(111)	No	12.65 ± 0.05	+0.06 V
Si	5.477	aFD	(111)	No	12.37 ± 0.05	+0.06 V
Ge	5.658	aREF	(111)	No	14.77 ± 0.08	+0.04 V
Ge	5.765	aFD	(111)	No	13.81 ± 0.07	-0.07 V
C	3.567	aREF	(211)	No	20.75 ± 0.05	**+0.39 V**
C	3.577	aFD	(211)	No	20.62 ± 0.04	**+0.39 V**
Si	5.431	aREF	(211)	No	12.57 ± 0.05	-0.02 V
Si	5.477	aFD	(211)	No	12.29 ± 0.05	-0.03 V
Ge	5.658	aREF	(211)	No	14.67 ± 0.08	-0.06 V
Ge	5.765	aFD	(211)	No	13.83 ± 0.07	-0.05 V

Figure 5.19: Effect on mean inner potentials from modifying bulk-terminated surface facet, compared to baseline. Bolded entries in the "**Δ from (110) baseline**" column differ from the baseline by more than the 95% confidence interval - thus, the mean inner potential is clearly modified from the baseline in that instance. Comparing this table to the baseline mean inner potentials seen in Figure 5.17, the surface facet – even when bulk-terminated – can clearly have an effect on the mean inner potential. This effect does not disappear when the thin films in question are relaxed – relaxation does not appear in this instance to have a strong effect on the mean inner potential compared to bulk-terminated surfaces.

Notably, while [50] only shows DFT-simulated mean inner potentials that are lower than the IASF values, Figure 5.19 shows that this is not always the case. The shift for C(001) yields a MIP of 21.41 ± 0.05 V from DFT, which is higher than the isolated-atom scattering factor calculation of 20.86 V at a_{REF}. The results for C(111), 21.18 ± 0.05 V, are also above the 20.86 V threshold; C(211) is below the threshold, at 20.75 ± 0.05 V. When combined with the surface facet dependence, there is a clear necessity for a surface-inclusive theoretical explanation of the mean inner potential, which will be presented in a later section.

Next, the *ab initio* relaxation simulations provide a comprehensive, internally consistent examination of the effect on the mean inner potential of structure optimization.

Relaxation *ab initio*

In this section, structure optimization effects on the mean inner potential are tested by relaxing clean bulk-like surfaces to minimum-force states. As structure optimization is involved, a_{FD} lattice parameters are used instead of a_{REF} for all materials except for C, where both a_{FD} and a_{REF} are used. This section only examines (1×1) "reconstructions" – the later section on reconstructions addresses larger surfaces.

Mean inner potentials from structurally optimized (110) surfaces are shown in Figure 5.20 for both unary and binary compounds. The III-V compounds, with the exception of GaAs, show increases in their mean inner potentials due to structural optimization, while, from the group-IV elements, Si shows a slight decrease, Ge shows no change, and C shows a large decrease. Therefore, even slight atomic rearrangements into relaxed states can lead to changes in the mean inner potential outside the simulation margin of error.

In addition, mean inner potentials determined from structural optimization of (001)-surface thin-films are also shown in Figure 5.19, for group-IV elements only. These show different behavior than the structurally optimized (110)-surface thin films: instead of the mean inner potential of C increasing relative to the unrelaxed case, it does not change (within the margin of error), remaining above both the isolated-atom scattering factor mean inner potential and the (110) mean inner potential. This supports the conclusion that the choice of surface atom configuration has an effect on the mean inner potential that is dependent on the exact configuration.

The difference in mean inner potential between the relaxed (110) and (001) facets widens for C with either lattice parameter, indicating that there is a chance that, in the relaxed state, the two surfaces would be experimentally distinguishable. The gap between the bulk-like C a_{REF} (110) and (001) facet mean inner potentials is 1.05V (or 5.2% of the (110) value); the gap between the relaxed C a_{REF} (110) and (001) facet mean inner potentials is 1.63V (or 8.0% of the (110) bulk-like value). This 8.0% difference in mean inner potential is large enough to have been experimentally distinguishable, if this had been measured in Chapter 4 at the ~2% margin of error. The two-dimensional implications of this are examined in a subsequent section, addressing what happens to the mean inner potential if a specimen has both (001) and (110) facets.

In summary, Figures 5.20 and 5.19 show that differences in mean inner potential between different crystallographic facets are seen for DFT simulations on both bulk-

Structural Optimization Effects on Mean Inner Potentials
FD-Mode Subtraction-Method, D/ZB Structure

Material	a (Å)	a type	Surface Facet	Relaxed?	MIP (V)	Δ from (110) baseline
C	3.567	aREF	(110)	Yes	19.79 ± 0.05	**-0.57 V**
C	3.577	aFD	(110)	Yes	19.72 ± 0.04	**-0.50 V**
Si	5.477	aFD	(110)	Yes	12.16 ± 0.05	**-0.15 V**
Ge	5.658	aFD	(110)	Yes	13.90 ± 0.07	+0.03 V
GaAs	5.773	aFD	(110)	Yes	13.28 ± 0.08	+0.01 V
InP	5.908	aFD	(110)	Yes	13.92 *± 0.06*	**+0.34 V**
InAs	6.100	aFD	(110)	Yes	14.09 *± 0.06*	**+0.18 V**
GaP	5.545	aFD	(110)	Yes	13.08 *± 0.06*	**+0.17 V**
AlP	5.513	aFD	(110)	Yes	11.58 *± 0.06*	**+0.47 V**
AlAs	5.740	aFD	(110)	Yes	12.10 *± 0.06*	**+0.25 V**

Figure 5.20: Effect on mean inner potentials from structure optimization, compared to baseline. Letting the atomic structure become optimized leads to the mean inner potential changing sometimes, but not always – and the group IV elements see decreased mean inner potential (if changed), and the III-V materials see increased mean inner potential (if changed). Therefore, the effect of structural optimization must be examined on a case-by-case basis. Compare to Figure 5.17. Bolded entries in the "**Δ from (110) baseline**" column are at or outside the 95% confidence interval for the calculation. Italicized confidence intervals are derived from the average of the test suite confidence intervals for the condition examined.

Surface-Adsorbate Modification Effect on Mean Inner Potentials
FD-Mode Subtraction-Method, D/ZB Structure

Material	a (Å)	a type	Surface Facet	Adsorbate	Relaxed?	MIP (V)	Δ from (110) baseline
C	3.567	aREF	(110)	H	Yes	16.20 ± 0.05	**-4.16 V**
C	3.577	aFD	(110)	H	Yes	16.12 ± 0.04	**-4.01 V**
Si	5.477	aFD	(110)	H	Yes	11.33 ± 0.05	**-0.99 V**
Ge	5.765	aFD	(110)	H	Yes	12.38 ± 0.07	**-1.50 V**
C	3.567	aREF	(110)	O	Yes	22.29 ± 0.05	**+1.93 V**
C	3.577	aFD	(110)	O	Yes	22.19 ± 0.04	**+1.96 V**
Si	5.477	aFD	(110)	O	Yes	15.73 ± 0.05	**+3.41 V**
C	3.577	aFD	(110)	Si	Yes	19.42 ± 0.04	**-0.81 V**
Si	5.477	aFD	(110)	C	Yes	11.86 ± 0.05	**-0.46 V**
Si	5.477	aFD	(110)	Ge	Yes	12.83 ± 0.05	**+0.51 V**
Ge	5.658	aFD	(110)	C	Yes	13.62 ± 0.07	**+0.26 V**
InAs	6.100	aFD	(110)	H	Yes	12.97 *± 0.06*	**-0.94 V**
InAs	6.100	aFD	(110)	O	Yes	15.90 *± 0.06*	**+1.99 V**

Figure 5.21: Effect on mean inner potentials from adding surface adsorbates, compared to baseline. Emplacing and structurally optimizing surface adsorbates has a very strong effect on the mean inner potential (*cf.* Figure 5.17, Figure 5.20). This effect appears for all materials studied, but with different magnitudes, depending on the adsorbate and the thin-film material. Most dramatically, C shows either a ~ -4 V decrease with a H adsorbate, and a ~ +2 V increase with an O adsorbate.

like and structurally optimized specimens. Therefore, the changes in mean inner potential with surface state are studied further in the following sections.

While the specimen surfaces examined in this section are optimized structurally, a realistic surface state might also include a native oxide or other adsorbate surface layers. The effect on the mean inner potential of adsorbate layers is examined in the following section.

Adsorbate Layers *ab initio*

In this section, adsorbate layers are applied to the thin-film DFT-simulated specimen, and the effect on the mean inner potential is examined. Adsorbate layers on the specimen surface are possible experimentally: a surface layer of some material (possibly a native oxide) is seen on the nanowire in Chapters 3 and 4. Additionally, the TEM electron beam may induce more specimen surface contamination. While the specimen with DFT-optimized structure, including surface adsorbates, might not be the same specimen that is measured in the TEM, DFT-optimized structures with adsorbate layers can provide a guide to which materials are more or less influenced by surface effects. By relaxing the atom positions in the simulation, surface-dependent mean inner potentials are self-consistently testable. As with the previous section, only (1×1) reconstructions are addressed here – larger ones are addressed in the later section on reconstructions. The adsorbate atoms are generated on the surface by changing the atomic species of the top layer of thin-film atoms, but this is only the starting position; all atoms in the simulation are relaxed to their structurally optimized positions.

Figure 5.21 shows mean inner potentials for (110) surfaces with adsorbate layers. In Figure 5.21, the mean inner potential is substantially modified by the addition of adsorbate layers. Comparing Figure 5.21 with Figures 5.20 and 5.19, surface adsorbate effects on the mean inner potential are larger than the relaxation effects and often larger than the surface facet effects. Continuing with diamond C and a_{REF} as the archetypical example, the MIP deviates from the clean, bulk-terminated-surface value by -4.16 V for a H-terminated (110) surface and +1.93 V for an O-terminated (110) surface. Relative to the mean inner potential of the clean bulk-terminated (110) surface, the H-terminated surface yields a change in mean inner potential of -20.4%, the O-terminated surface yields a change of +9.5%. While the atom positions of the adsorbates might not be the same as for experimental specimens, the DFT mean inner potential simulations show large surface adsorbate effects. The difference predicted by DFT mean inner potential simulations between H-terminated and O-terminated diamond surfaces (~30%) would be large enough to be experimentally detectable.

In summary, adsorbates have a strong effect on the mean inner potential. The addition of adsorbates onto surfaces leads to large changes in the DFT-simulated mean inner potential for multiple material systems and orientations. While the nature of the specimen surface might not be experimentally determinable in some cases, DFT predicts that the nature and atomic arrangement on surfaces matters for the mean inner potential. Now, the following sections examine the DFT-simulated mean inner potential from selected experimental surfaces for comparison.

Selected Experimental Surfaces

Reconstruction Scenario Choices

While the previous sections examined surface effects on the mean inner potential through self-consistent structural optimization, the following sections use experimental data to test the surface dependence of the mean inner potential. The DFT simulations in this section calculate mean inner potentials from individual surface atomic configurations related to experimental data, including surface reconstructions and adsorbate layers.

The surface structures reported in the literature can be divided into three categories depending on how much information is given. In descending order of desirability for DFT simulation of mean inner potentials, these three categories are: structures with full experimentally determined atomic coordinates (first case), structures with simulation-determined atomic coordinates (second case), and structures with incomplete atomic coordinate data but specifying some form of reconstruction symmetry (third case). In the first case, the full experimentally determined atomic coordinates are used as input for the DFT simulation of the mean inner potential; this situation is examined first in the structure-as-specified and then after structural optimization, and both pre- and post-relaxation mean inner potential values are given. In the second and third cases, it is suboptimal to use atomic coordinates determined *in silico* in the literature, as they are determined using different algorithms, but are not experimental data. Surfaces where only the symmetry is given (*i.e.* without atomic coordinates) necessitate structure determination. Therefore, for DFT simulation of the second and third cases, the surface reconstruction symmetry and surface atomic species are specified, the structure is optimized, and the only the mean inner potential after structure optimization is quoted.

As in the previous section, the DFT-structural-optimization atomic positions are likely not the same as experimental atomic positions. However, the results of structural optimization are self-consistent, even if they may be inaccurate when compared with experimental atomic positions. Simulations using structurally optimized atomic positions and experimental atomic positions may also show differences, providing additional insight into how the mean inner potential might depend on structure.

In the following sections, DFT-simulated mean inner potentials from experimentally relevant specimen surfaces are examined. First, clean surfaces (including reconstructions larger than (1×1) unit cells) are simulated, followed by surfaces with adsorbate layers.

Mean Inner Potential From Reconstructions

Given the strong surface dependence of the mean inner potential discussed in the previous sections, reconstructions larger than (1×1) unit cells of clean surfaces are expected to lead to changes in the mean inner potential relative to (1×1) surfaces. In this section, reconstructed clean surfaces larger than (1×1) are examined, with both experimentally determined atomic positions, and atomic positions determined through structural optimization, as discussed in the previous section. Due to computational constraints, only (2×1) reconstructions are discussed in this section.

One literature reference – [122] – for Si (001)(2×1) had experimental atomic positions to use directly as simulation input. This scenario, as shown in Figure 5.22, yielded a mean inner potential of 12.02 ± 0.05 V with experimental atomic positions, and 12.14

Surface-Structure Configurations from Literature: Effects on Mean Inner Potentials
FD-Mode Subtraction-Method, D/ZB Structure

Material	a (Å)	a type	Refs	Surface	Adsorbate	Relaxed	MIP (V)	Δ from (110) baseline
C	3.567	aREF	[120] [123]†	(001),(2×1)	Ø	Yes	21.40 ± 0.05	**+1.04 V**
Si	5.477	aFD	[122]*	(001),(2×1)	Ø	No	12.02 ± 0.05	**-0.30 V**
						Yes	12.14 ± 0.05	**-0.18 V**
Si	5.477	aFD	[124]*	(001),(2×1)	H	No	11.52 ± 0.05	**-0.80 V**
						Yes	11.52 ± 0.05	**-0.80 V**

*: Used full atomic positions; for relaxation simulations as starting positions. Coordinates from literature are rescaled from experimental lattice parameters.
†: Used surface symmetry only.

Figure 5.22: Effect on mean inner potentials from using realistic surface structure. When experimentally determined surface structures are applied, the mean inner potential changes relative to the baseline. This confirms the effects discussed in Figure 5.21: adding surface adsorbates, and/or letting the thin-film achieve an optimized structure, can lead to changes in the mean inner potential.

Surface-Adsorbate Modification Effect on Mean Inner Potentials
FD-Mode Subtraction-Method, D/ZB Structure

Material	a (Å)	a type	Surface Facet	Adsorbate	Relaxed?	MIP (V)	Δ from (110) baseline
C	3.577	aFD	(110)	H/O	Yes	18.96 ± 0.04	**-1.27 V**

Figure 5.23: Effect on the mean inner potential from using different top/bottom adsorbates. Only one case is seen in this table – but it shows that the mean inner potential is modified by the addition of different adsorbates on the top and bottom facets of the thin film roughly as expected – as an average between the H- and O-terminated cases (*cf.* Figure 5.21).

± 0.05 V when structurally optimized atomic positions. Structure optimization in this case slightly increased the mean inner potential from the experimental configuration. The (2×1) reconstruction has an effect on the mean inner potential. Both experimental-atomic-positions and structurally-optimized-atomic-position mean inner potential values are notably lower for Si(001)(2×1) than Si(001)(1×1), which is given in Figure 5.19 as 12.85 ± 0.05 V when bulk-terminated, 12.90 ± 0.05 with structural optimization. Comparing to the (110) surface shown in Figures 5.16 and 5.20, however, the bulk-terminated (110) surface has a higher mean inner potential, at 12.32 ± 0.05 V, than the specimen with (001)(2×1) experimental atom positions, but the relaxed (110) surface mean inner potential is within the margin of error of the relaxed (001)(2×1) surface.

The C(001)(2×1) surface was also examined. This surface did not have literature-derived atomic positions, but instead used the surface knowledge from literature reports including DFT simulations [120] backed by experimental surface confirmation of the symmetry [123]. Thus specified, the mean inner potential of structurally optimized C(001)(2×1) is given in Figure 5.22 as 21.40 ± 0.05 V. This is within the margin of error of both the bulk-like and relaxed mean inner potentials for C(001)(1×1) (given in Figure 5.19). Therefore, comparing C and Si, the usage of a (2×1) clean-surface-reconstruction instead of a (1×1) clean-surface-reconstruction does not always change the mean inner potential.

Thus, a general conclusion cannot be drawn about surface sensitivity of the mean inner potential for reconstructions of these clean surfaces, given these two examples – it can have an effect (as seen for Si), but it does not need to (as seen from C). Only having two data points is insufficient for broad conclusions, but the effect of surface reconstruction should be further studied. In the next section, the effect on the mean inner potential of adsorbate layers from experimental information is briefly examined.

Adsorbate Layers From Experimental Configurations

The logical next step in DFT mean inner potential simulation from clean reconstructed surfaces is to include adsorbate layers. These are perhaps the most experimentally relevant systems that can be examined *in silico*; however, they come with the accumulated challenges of all of the above simulations, and only one example is shown in this section. This section simulates the mean inner potential of Si(001)(2×1)+H, using both experimentally determined atomic positions and structural optimization, and compares with baseline values.

Figure 5.22 shows the mean inner potential of Si(001)(2×1) with adsorbed H in a literature-derived atomic configuration from [124]. This yields a mean inner potential of 11.52 ± 0.05 V using atomic positions from the literature. The mean inner potential does not change when the atomic positions are structurally optimized. This is a change of -0.80 V from the baseline Si(110) mean inner potential. For comparison, the mean inner potential of Si(001) is +0.54 V above the baseline Si(110) mean inner potential. Compared to Si(110)(1×1)+H, which has a change of -0.99 V from the baseline Si MIP, Si(001)(2×1)+H shows a larger change relative to the clean surface facet. Therefore, the mean inner potential is modified by surface adsorbates persists even when the surface adsorbates are in experimentally determined configurations, and the mean inner potential derived from structurally optimized atomic positions can be the same as the mean inner potential derived from experimentally determined atomic positions.

This section details only one data point – Si(001)(2×1)+H. However, this section uses experimental atom positions and DFT simulation to reinforce two key points about the mean inner potential. First, the mean inner potential can change with surface adsorbate, even in a realistic surface adsorbate configuration. Second, DFT simulation using the atomic positions derived using structural optimization can give the same mean inner potential as atomic positions from experimental systems, meaning that structurally optimized DFT-simulated mean inner potentials may be useful for comparison with experiment.

Multiple-Surface Effects

The above sections clearly indicate that surface effects can influence the mean inner potential. However, each of the previous sections has run simulations with all surfaces being identical, and these simulations have shown strong mean inner potential modification for a range of different surface configurations.

This raises a question: If the mean inner potential is a surface-dependent scalar value measured between the middle of the material and vacuum, how does the mean inner potential vary when the specimen has more than one type of surface?

In this section, two different kinds of systems with multiple different surfaces are examined using DFT: first "thin film"-type simulation geometries with different top and bottom surface adsorbates, and then "nanowire"-type simulation geometries with two different exposed crystallographic facets.

Heterosurfaces

While all of the previous simulations have relied on the top and bottom surfaces of the simulated thin film being identical, it is possible to run DFT simulations of thin films where the top and bottom surfaces are different – "heterosurfaces". In the thin-film case, they must retain the same crystallographic facets, but different adsorbates can be placed on the top and bottom surfaces. The results are then compared to the single-species adsorbates examined in the previous sections.

Using hydrogen on one surface and oxygen on the other allows for good comparison with the previous set of simulations. H and O are used because, in Figure 5.21, the H-termination decreased the mean inner potential, and the O-termination increased it, relative to the bulk-terminated mean inner potential. Therefore, using one of each kind of surface – an O_{top}/H_{bottom}-termination - should, in principle, provide a number between the H-terminated and O-terminated values.

Figure 5.23 shows the mean inner potentials from a heterosurface specimen using a structurally optimized H-termination on the top surface and O-termination on the bottom surface. The expected mean inner potential of a material where one surface is terminated with H and the other with O might be between the value from a DFT simulation where both surfaces are solely H-terminated and the value from a DFT simulation where both surfaces are solely O-terminated. Figure 5.23 shows this to be the case. In Figure 5.23, the mean inner potential from C(110) with a_{FD} is 18.96 ± 0.04 V for O_{top}/H_{bottom}-termination. For comparison, Figure 5.21 shows the C(110) solely H-terminated mean inner potential to be 16.12 ± 0.04 V and the C(110) solely O-terminated mean inner potential to be 22.19 ± 0.04 V. A simple average of these two would indicate a value of 19.15 V, which is near, but not quite, the O_{top}/H_{bottom}-terminated mean inner potential. Therefore, the effect of terminating the surfaces

differently may not be a simple influence, but it does have an influence somewhat as expected.

While this section quickly examines mean inner potentials from a thin-film simulation using two different surface adsorbates, the next section examines the effects of terminating with two different surfaces in a single simulation – the "nanowire" geometry.

Two-Dimensional Simulations: The Effect of Corners

All of the thin-film simulations in the previous sections have only one top-and-bottom set of interfaces between material and vacuum. However, what if there are two different vacuum-material interfaces in the same simulation – either of the same crystallographic type, or of different ones? In this subsection, this question is addressed.

This question arises because there are several implications of the thin-film case, deriving from the scalar nature of the electrostatic potential and the infinite nature of the thin-films studied above. Consider two separate thin-films, both with bulk-like terminations and infinite in the in-plane direction, but thin-film A has a (110) terminating facet, and thin-film B has a (001) terminating facet. According to Figures 5.16 and 5.19, if this thin-film is made out of diamond carbon with the reference lattice parameter, then thin-film A has a mean inner potential of 20.36 ± 0.05 V and thin-film B has a mean inner potential of 21.41 ± 0.05 V, when the difference is taken between the middle of the material and free space a sufficient distance outside it. Now, consider the effect of making thin-film A finite in the (001) direction, and thin-film B finite in the (110) direction, such that A and B are now identical (but still finite in the third direction), yielding infinitely long "nanowires". The two nanowires made from A and B should now have the same mean inner potential, because they are identical objects, but they were generated from two objects with different mean inner potentials. What exactly happens to make this occur?

This is tested in this section by examining two different cases: One where a "nanowire" simulation has (001) and (110) surface facets, and the other where a "nanowire" simulation has (110) and $(1\,\bar{1}0)$ surface facets. Due to computational constraints, C was examined in both cases, while Si was only examined for the simulation using (110) and $(1\,\bar{1}0)$ surface facets. To display them as images, the three-dimensional electrostatic potential output is scaled down to two dimensions, where the "infinite" dimension without a surface is averaged [$(1\,\bar{1}0)$ for the (001)/(110) simulation, (001) for the (110)/$(1\,\bar{1}0)$ simulations.

The electrostatic potential results from DFT simulations of a C "nanowire" combining (110) and (001) bulk-terminated surface facets are seen in Figure 5.24, depicting electrostatic potential[38] contours every 0.1 V. The left image in Figure 5.24 shows the full range of the electrostatic potential in the simulation, where the blue dots are the location of the atoms (low potential), and the red contours show fringing fields in vacuum (high potential). The image on the right on Figure 5.24, by contrast, shows the same volume as the left image, but with different contours plotted – every 0.1 V

[38] Technically, the figures referenced in this section show contour plots of the Hartree potential, labeled as V_{Htg} in the GPAW code. This will be quantitatively incorrect near the atomic cores, but this section considers the fringing fields outside the material, which are accurate for this potential.

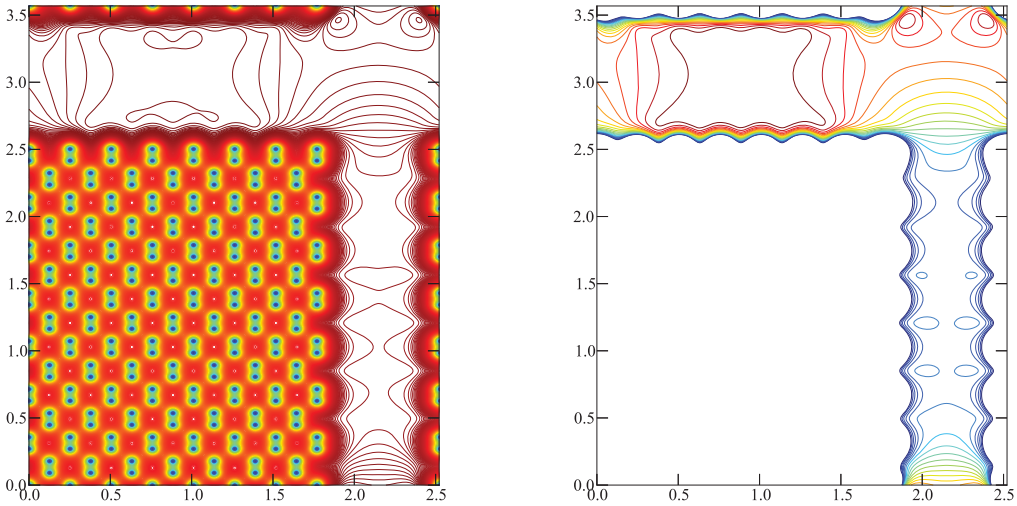


Figure 5.24: Diamond C two-surface bulk-like "nanowire" (001)/(110) electrostatic potential simulations. The spatial x/y scale is in nm. The left image shows the full range of the electrostatic potential in the simulation, where the blue dots are the location of the atoms (low potential), and the red contours show fringing fields in vacuum (high potential). The image on the right, by contrast, shows the same volume as the left image, but with different contours plotted – every 0.1 V across only the maximum 2.0 V of the potential. This yields an interpretable map of the fringing fields in vacuum. The vacuum potential level is different between the (001) (top/bottom) and (110) (left/right) facets, with fringing fields between them See text for more discussion.

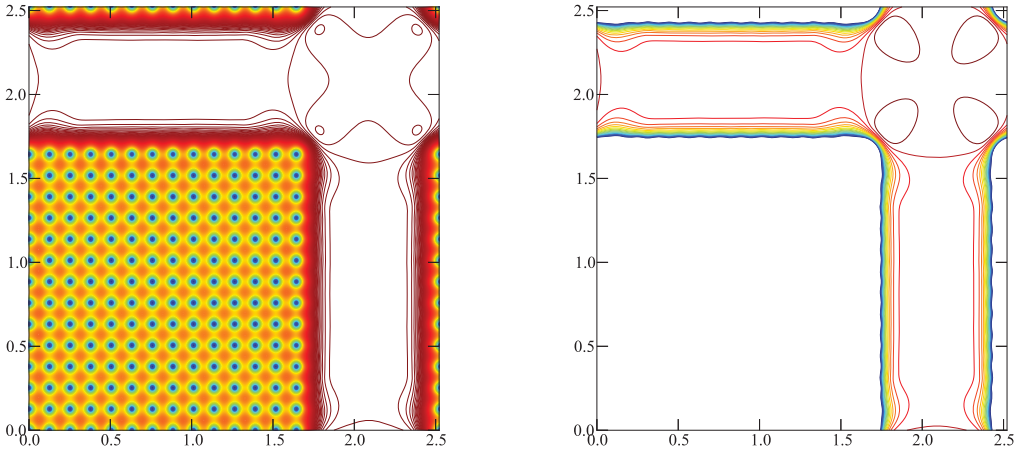


Figure 5.25: Diamond C two-surface bulk-like "nanowire" (110)/ ($1\bar{1}0$) electrostatic potential simulations. The spatial x/y scale is in nm. For interpretation of left and right images, see Figure 5.24. This yields an interpretable map of the fringing fields in vacuum. Compared with Figure 5.24, there are many fewer fringing fields when both the top/bottom and left/right facets are the same, but there are still fringing fields at the corners.

across only the maximum 2.0 V of the electrostatic potential. The right image yields an interpretable map of the fringing fields around the wire. Several features are interpretable from Figure 5.24. First, there is a difference in the vacuum level outside the different facets – from counting contours, the vacuum outside the C(001) facet has an electrostatic potential that is ~1.5 V higher than the vacuum outside the C(110) facet. The thin-film mean inner potential of C(001), according to Figure 5.19, should be +1.05 V compared to C(110). Thus, the facet-dependent difference in the two-dimensional simulation is higher than expected. Second, the cut plane through the unit cell of the bulk-like surface termination seem to matter only a little – the two C(001) surfaces in Figure 5.24 have atoms in the same locations, meaning they are cut through different parts of the unit cell. Despite this, in Figure 5.24, there are some differences in electrostatic potential outside the top and bottom C(001) facets, but these differences are relatively minor when compared with the electrostatic potential between the C(001) and C(110) facets. The conclusion from the two-dimensional C(001)/(110) simulation in Figure 5.24 is that a nanowire simulation with two different surface facets leads to fringing fields.

Now, the effect on the nanowire-geometry simulations of fringing fields and structural optimization is examined. The $C(110)/(1\bar{1}0)$ simulation, shown in Figure 5.25 shows weaker fringing fields than the C(001)/(110) simulation, and these fringing fields are largely suppressed by relaxation. Figure 5.25 shows fringing fields in vacuum around the corners of the "nanowire", but fewer fringing fields than the C(001)/(110) simulation in Figure 5.24, and the two facets have the same vacuum electrostatic potential. In contrast, in Figure 5.26, the $C(110)/(1\bar{1}0)$ electrostatic potential simulation after relaxing atomic positions from bulk-terminated starting points is seen. The fringing fields seen in Figure 5.25 are largely suppressed in this relaxed case, seen in Figure 5.26. Some atomic rearrangement is visible, especially near the corners. Therefore, the atomic arrangement at the surface has a large effect on the fringing fields in vacuum, especially near corners. This may help solve the two-dimensional problem of different surface facets yielding different mean inner potentials: if the surface of a specimen is not bulk-like, but rearranges itself to suppress fringing fields, then the mean inner potential would be the same no matter where the reference point in vacuum was taken. However, this rearranged, optimized surface would yield a mean inner potential that is likely not the same as a value calculated from "thin-films" as examined earlier in this chapter.

To appreciate how difficult this might be, a $Si(110)/(1\bar{1}0)$ electrostatic potential simulation with bulk-like and relaxed structure is now examined. This is a smaller structure than the C structures, due to computational constraints. The bulk-like Si structure is shown in Figure 5.27, yielding more fringing fields than the C equivalent in Figure 5.24, but the mean inner potential outside the facets is the same for both facets seen in Figure 5.27. When this Si structure is relaxed, the result is shown in Figure 5.28. This result is quite different from the C result. While the C structure (Figure 5.26) only showed strong effects near the corners, the Si structure (Figure 5.28) shows large-scale rearrangement along the facets as well. This might be due to the smaller computation size of the Si simulation, but might be more likely attributable to Figure 5.28 being a material-dependent minimum-energy atomic arrangement for Si. Further work is necessary on this point. Larger simulations might help systematize this two-dimensional atomic position optimization.

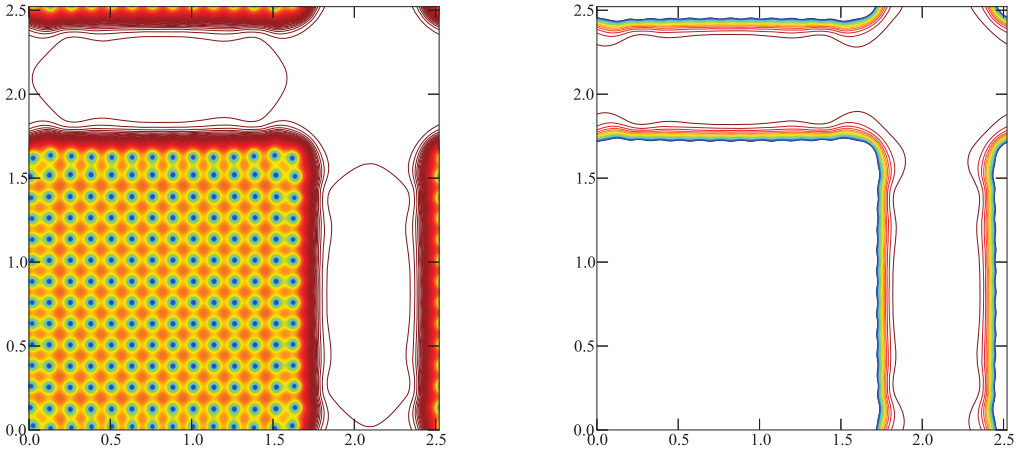


Figure 5.26: Structural optimization of diamond C two-surface bulk-like "nanowire" (110)/(1$\bar{1}$0) electrostatic potential simulations. The spatial x/y scale is in nm. For interpretation of left and right images, see Figure 5.24. This yields an interpretable map of the fringing fields in vacuum. Compared with Figure 5.25, the structural optimization almost completely removed the fringing fields at the nanowire corners, and the atom positions (left image, blue dots) have changed. Structurally optimized surfaces are probably more realistic than bulk-terminated ones; *ergo*, DFT simulations including the effects of corners may be better than "thin film" simulations for comparison with experimental mean inner potential measurements.

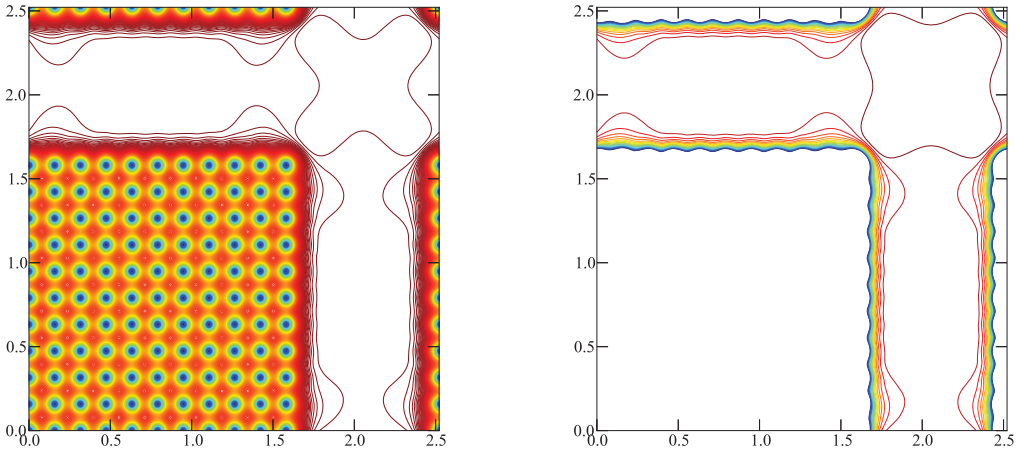


Figure 5.27: Si two-surface bulk-like "nanowire" (110)/ (1$\bar{1}$0) electrostatic potential simulations. The spatial x/y scale is in nm. For interpretation of left and right images, see Figure 5.24. Compared to the C counterpart (*cf*. Figure 5.25), there are the same features – same vacuum levels outside the two surfaces, and corner effects, including fringing fields.

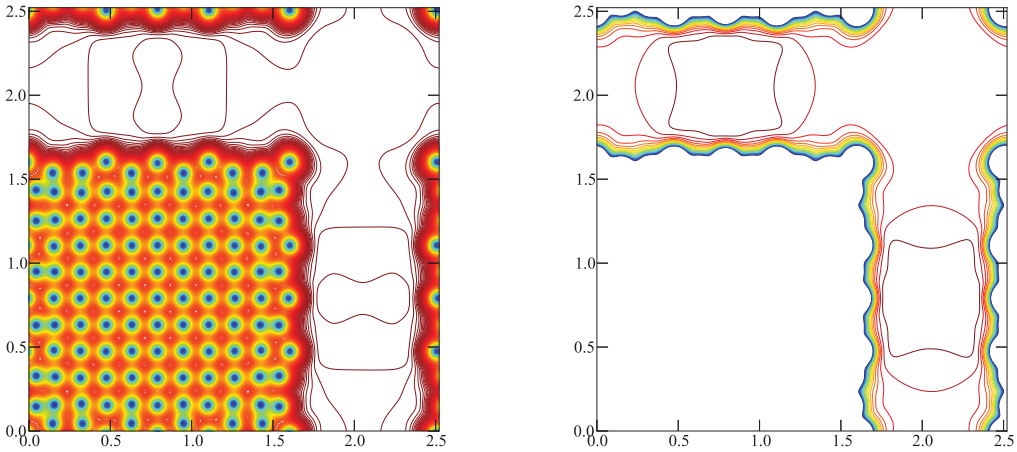


Figure 5.28: Structural optimization of Si two-surface "nanowire" (110)/(1$\bar{1}$0) electrostatic potential simulations. The spatial x/y scale is in nm. For interpretation of left and right images, see Figure 5.24. Compared to the C counterpart (*cf*. Figure 5.26), the fringing field around the nanowire is larger after structural optimization compared to with bulk-like surface termination (Figure 5.27). The effect of structural optimization, therefore, may be difficult to predict *a priori*.

In conclusion, the two-dimensional simulations yield important insight into the surface facet dependence of the mean inner potential. These simulations resolve the facet dependence of the mean inner potential for "thin-films", seen in previous sections, by simulating nanowires with two different surface facets, which shows fringing fields develop, and the effect is not exactly that predicted from thin-film simulations. Relaxing atomic positions yields changes in the fringing fields, but these changes are material-dependent, and likely also surface-facet-dependent. Further work in this line – for the same materials and different surfaces, and for different materials – might yield additional insight into the fundamental sources of the mean inner potential.

Summary of Surface-Dependent Effects

The sections above show that mean inner potential is strongly surface-dependent. The mean inner potential depends on the configuration of atoms at the surfaces of materials under study, and it depends on the atomic species. Some surface modifications change the mean inner potential more than others. Of the three mean inner potential surface effects examined here, the surface adsorbate effect appears largest, followed by the effect from using a different surface facet, and the smallest, but still nontrivial, effect is from surface optimization.

These three effects act on the mean inner potential in unpredictable ways. Some adsorbates increase the mean inner potential, others decrease it. For bulk-like terminations, the (110) orientation appears to yield the minimum mean inner potential for the materials studied here, but the magnitude of the other orientations' mean inner potential shifts relative to the (110) case is different for each orientation. The effect of relaxation on the mean inner potential, as seen in Figure 5.20, can be either positive or negative.

What all these scenarios have in common is that changes in surface configuration yield mean inner potential shifts. For an explanation of why this might be, the section that follows analyzes the theoretical sources of the mean inner potential.

What is the Mean Inner Potential?

To understand the theoretical basis of the mean inner potential, and why the specimen surface could affect it, there is limited previous work. However, it is instructive to consider the formulation of the mean inner potential in the high-energy approximation relevant for transmission electron microscopy put forward by Saldin and Spence [65]:

$$|e|V_0 + |e|V_{xc}(\varepsilon_F) = \varepsilon_F + \Phi$$

In this equation, the mean inner potential V_0 depends on the Fermi energy ε_F, work function Φ, and the exchange-correlation potential $V_{xc}(\varepsilon_F)$ at the Fermi level. This equation includes accelerating-voltage dependence: in lower-energy cases, $V_{xc}(\varepsilon_F)$ decreases, so the Fermi energy and work function, if considered together, provide an upper limit to the mean inner potential.

In the Saldin and Spence mean inner potential formulation [65], the mean inner potential is explicitly surface-dependent. This is because the mean inner potential as encountered by the high-energy electron as it propagates through the specimen is the entire difference in electrostatic potential between vacuum and that point in the crystal, which inherently includes the work function.

This helps explain the dependence of the mean inner potential on surface properties seen in the previous sections, because the work function will change with different surface atom configurations and facets, as seen in [125]. While [125] provides only one direct work function comparison for the materials studied in this chapter, it is between p-doped Si in (001) and (111) orientations. The former is listed as 4.91 eV, and the latter 4.60 eV; the difference between these compares well with their mean inner potentials: 12.90 V and 12.65 V, respectively, in Figure 5.19.

A formulation of the mean inner potential from a pro-crystal derived from isolated-atom scattering factors (at zero scattering angle) will not only neglect bonding effects, but is also strictly insensitive to surface effects such as those shown here to have great influence. Therefore, isolated-atom scattering factors are deeply non-ideal for comparisons of V_0 with experimental data. However, this relation between electron scattering factors and the mean inner potential also may allow for insight into the fundamental nature of the scattering factor through deriving the electron scattering factors at zero scattering angle from DFT simulations of mean inner potential.

The results seen here are also qualitatively consistent with the literature on work functions, where *e.g.* thin-film [126] and nanowire geometries with atom-position relaxation in [127] and [128] are used with DFT to investigate work functions of Al. In those reports, crystallographic facet dependence and relaxation are both shown to affect the work function, which is consistent with the results in this chapter (*e.g.* Figure 5.19). While the work function is not the sole component of the mean inner potential in the Saldin and Spence mean inner potential formulation, investigations of factors affecting the work function may be useful proxies for investigations of the mean inner potential surface dependence, both experimentally and theoretically.

However, the work function itself is not a fundamental quantity. The work function can be decomposed into band bending and electron affinity components [129]. Changes in both can be seen when surfaces are modified [129], and consideration of this may provide some insight into the effects of surface layers on the mean inner potential. Examining factors affecting the electron affinity may yield more insight into how the mean inner potential is dependent on material parameters [130].

However, the Saldin and Spence formulation may be incomplete: for high energies, the beam electron will not undergo exchange and correlation effects with the electrons in the material [65]. This exchange-correlation effect removal is considered for the Fermi level, but not for the work function. The work function definition in [129] discusses removing an electron from the bulk of a material to a distance far outside the material; this electron undergoes exchange and correlation effects with the electrons in the material. Therefore, the Saldin and Spence mean inner potential formulation, above, can be extended:

$$|e|V_0 = \Phi + \varepsilon_F - |e|V_{xc}(\varepsilon_F) - |e|V_f$$

where V_f comprises all the exchange and correlation components of the work function that do not contribute to it as a result of the velocity of the beam electron.

A final minor, yet significant, point about the DFT simulation of work functions is the PBE exchange-correlation functional used here will yield different results for work functions as compared to the LDA exchange-correlation functional [126]. In this chapter, the effects of exchange-correlation functional are not considered in the error

estimates, because changing exchange-correlation functionals would be a fundamental difference in the system under simulation. However, the most appropriate choice should be systematically determined.

The surface effects on the mean inner potential discussed in this section are the result of rearrangement of atomic structure or adsorbate layer addition. The surface effect would be subtler if dopants are used to modify the Fermi level in the specimen due to Fermi level pinning from space-charge effects [129] – under those conditions, the Fermi level is shifted, but Fermi level pinning at the surface would keep the work function constant; however, Fermi level pinning can be modified through adsorbed layers *e.g.* oxygen on GaAs [129]. Thus, in the Saldin and Spence formulation of the mean inner potential, the specimen-dopant effects seen in Chapter 2 could affect the mean inner potential in complex ways.

Is the Mean Inner Potential Measurable?

The strong surface-dependence of the mean inner potential shown in this work raises the natural question of whether the mean inner potential is measurable at all. A change to surface conditions appears to be able to change the mean inner potential, and, since specimen surface conditions are difficult to measure in the TEM, is it still possible to measure the mean inner potential? In short, yes – if care is taken.

The following sections consider the consequences for mean inner potential measurement if the DFT simulations are correct, and the mean inner potential is shown to be surface-dependent. The next section considers the effects of the specimen surface on mean inner potential accuracy and precision, and the section following discusses the ways surface dependence would change experiments measuring the mean inner potential.

Effect of Surfaces on Mean Inner Potential Precision and Accuracy

Precision and accuracy in mean inner potential measurements are both important, but both have different effects. Precision means how repeatable the surface structure preparation is, and accuracy means how well this surface structure is known.

For mean inner potential determination, precision and accuracy in surface knowledge have different requirements and different applications. Precision allows for comparison between specimens. If the specimen surfaces were always prepared by the same procedure, then, while the specimen surface geometry might not be accurately known, mean inner potential measurements would be relatively precise and repeatable. Two specimens whose surface preparations are the same can be compared to determine if there is a difference in *e.g.* electronic structure or dopant between them. The surface geometry of the specimen itself might be determinable with precise mean inner potential measurements coupled to a reference library comprising a suite of DFT mean inner potential simulations of different possible surface structures.

The effect of specimen surfaces on mean inner potential accuracy is straightforward. The precision to which the surface is known affects the accuracy of the mean inner potential measurement. If the surface is unknown (*e.g.* the InAs nanowire in Chapter 4), then the surface effects considered here cannot be completely known. The more surface information that is known, the closer a DFT simulation geometry can be matched to experiment and used as a reference point.

The effect of specimen surfaces on mean inner potential precision is subtler. The more precise the surface preparation is, the more precise the mean inner potential measurement will be. If the surface of the specimen is prepared in the same way for each mean inner potential measurement, then the precision of the mean inner potential measurement should be unaffected, as the underlying specimen would be the same. While surface-structure precision requires undergoing the same surface-preparation procedures for each specimen, surface-structure accuracy would active determination of the surface structure. An additional surface-sensitive technique (like surface x-ray diffraction [124]) would be necessary to measure the atom positions. If the surface state is known, then a single high-fidelity measurement could be compared with DFT simulation to determine the mean inner potential.

However, one tautologically unavoidable effect of transmission electron microscopy is electron beam irradiation, which may affect either electronic structure or atom positions. Irradiation may significantly modify surface layer electronic structure, for example, [131] reports that Cl chemisorbed onto Si(111) yields a two-dimensional hole gas that is destroyed under electron irradiation due to hole trap generation. In addition, the surface adsorbate atomic positions themselves may react under high-energy electron beam irradiation differently than the substrate material. The ideal minimum-energy surface configuration may be unstable under electron irradiation. Therefore, even if the surface structure is accurately known before or after the TEM experiment, it may change under irradiation. Thus, effects on the mean inner potential due to electron beam irradiation may be present and, if they are present, should be studied carefully.

Electron-beam-induced charging may also influence the mean inner potential. If electronic structure at the surface is a critical factor in mean inner potential determination, then specimen charging in the beam may cause non-ground-state charge distributions throughout the specimen, including at surfaces, rendering the DFT simulation non-representative of the real charge distribution in the specimen. This effect on the mean inner potential may also be considered in future work.

The long-range fringing electrostatic potentials seen in the nanowire-geometry simulations may be either undetected or improperly compensated for. These long-range potentials may be mistaken for charging in experimental images and may be inaccurately compensated for. If the long-range electric potential effects influence the reference wave, then the path-difference electron phase measurement taken using electron holography would not accurately map out the electric fields in the specimen [37].

All of the surface-sensitive considerations raised in this section would have consequences for mean inner potential measurements. Experimentally, the specimen surface structure should be stable or known. The range of mean inner potentials possible with different plausible surface structures should be compared with DFT simulation. The effect of fringing electrostatic potentials should be compensated for. However, these effects do not preclude future mean inner potential measurements.

Opportunities for Future Mean Inner Potential Measurements

The mean inner potential's surface dependence in this work poses both opportunities and challenges for electron holography. The previous section detailed the challenges this poses; this section details the opportunities. This section discusses the

ramifications of, and ways to measure, the DFT simulations' surface-dependent mean inner potentials, assuming they are a real effect.

It is important to note that experimental mean inner potential measurements so far attempted have not addressed the issue of surface effects, and simulations have only treated surface effects briefly (as detailed in Chapter 2). This chapter discusses how and why the mean inner potential could be surface-dependent, which, if true, emplaces certain signposts for future mean inner potential measurements.

Uncontrolled specimen surfaces may be why the literature values for the mean inner potential have a high dispersion, as discussed in Chapter 2. In addition to determining the thickness to sufficient accuracy, the surface state must also be controlled or known. If the measurements of the mean inner potential reported in the literature came from specimens with different surface conditions, then the specimens would possibly have different mean inner potential values – thus, measurements of the mean inner potential from different specimens would not agree, no matter how precise the measurement.

Given the dependence of the mean inner potential on surface conditions, electron holography may be a useful technique for surface-sensitive studies in the TEM, if the surface state of the specimen can be both controlled experimentally and simulated using DFT. This surface control may prove to be a challenge in the non-ultrahigh vacuum TEM, but careful selection of specimen and preparation may allow for this level of control to be realized experimentally.

Some specific mean inner potential experimental measurement can be supposed to test this surface dependence. If a material can be inserted into the microscope with the surface in one state, and the surface modified *in situ* without changing the specimen thickness, then the surface-dependent shift in mean inner potential might be measurable directly. Specifically, taking a reconstructed Si, C, or InAs surface that is covered in O and replacing the O with H should yield a dramatic increase in the mean inner potential. Because this would be a shift in the total mean inner potential of the specimen, it would be more detectable for thicker specimens. To take one example, neglecting diffraction effects, according to the DFT simulations, for a 100nm specimen irradiated by 120kV electrons, the O-terminated Si(110)(1×1) surface listed above would measure a 13.6 radian phase shift, and the same specimen H-terminated would measure a 9.8 radian phase shift. If this modification could be induced in-situ, then the progress of the surface modification could be measured using electron holography.

Another possible way to access surface-dependent mean inner potentials experimentally would be to prepare a surface and to measure its decay. As a thought experiment, if cumulative beam damage destroys a hydrogen-terminated surface, then starting with a hydrogen-terminated Si, C, or InAs (110)(1×1) surface and desorbing the hydrogen would likely increase the mean inner potential over time, and the magnitude of the shift would be reflective of the progress of the desorption.

However, it is also instructive to compare these results and considerations with the experimental measurements carried out in the previous chapters of this thesis.

Simulations Compared With Experiments In This Thesis

While this chapter explains the importance of surface knowledge – including InAs – surface knowledge is absent from the experimental chapters. The TEM investigations of the InAs nanowire in chapters 3 and 4 do not determine the surface state – only that there is a shell surrounding the nanowire, likely a native oxide, but they do yield highly limited information on the surface.

At this point, comparison between DFT-simulated mean inner potentials and the experimentally measured mean inner potential of InAs in Chapter 4 may be instructive, if not necessarily quantitatively valid. The mean inner potential of clean (110) InAs matches with the experimental data. From Chapter 4, the mean inner potential of InAs with RENOIR thickness determination is 14.0 ± 0.4 V from the synthetic-wedge method and 14.0 ± 0.3 V from the maximum-thickness method and with DT thickness determination, the mean inner potential is 14.5 ± 0.4 V from the synthetic-wedge method and 14.6 ± 0.3 V from the maximum-thickness method[39]. From Figure 5.16, the mean inner potential of clean (110) InAs is 14.18 ± 0.06 V. *In the absence of surface effects, and assuming the experimental lattice parameter is the same as the reference lattice parameter*, this would slightly prefer the RENOIR thickness measurement to the DT one.

The specimen has an unknown surface state, so the picture becomes resolutely ambiguous when surface effects are included. InAs (110) clean-surface relaxation boosts the mean inner potential by +0.18 V for the a_{FD} lattice parameter; it is likely to have the same effect for the a_{REF} lattice parameter, subject to volume scaling. This would yield a mean inner potential of 14.36 ± 0.06 V, which is barely inside the margin of error of the DT thickness determination. If the surface is fully hydrogen-terminated (and using the a_{FD} lattice parameter), then the mean inner potential should decrease by -0.94 V, and it should increase for a fully oxygen-terminated surface by +1.99 V (from Figure 5.21). The fully oxygen-terminated InAs surface would yield a mean inner potential in this case above 16 V, which is well outside the margins of error for the experimental measurement using either RENOIR or DT. This would not be a problem if the nanowire surface were known.

Unluckily, the InAs nanowire has a shell. As discussed in chapters 3 and 4, this shell is present in the HRTEM images and the holograms, but is not present in the HAADF images. The HRTEM investigation shows a non-smooth surface, and the shell appears to be different from the core. Given that this InAs nanowire spent years of time in air between growth and TEM characterization, it is likely, although by no means certain, that this is a native oxide, which is generally consistent with the above data.

Many unknowns are present. The exact composition and atomic arrangement of the oxide is unknown. It is somewhat unclear whether the above mean inner potential measurements are fully faithful measurements of the InAs nanowire shell. It is also definitely unclear whether the above simulations are accurate representations of the InAs surface.

There is one significant countervailing influence, however: the relative stability of mean inner potential measurements when compared to the simulation. The other experimental citation for the mean inner potential in InAs used cleaved wedges (and

[39] In both cases, as noted in Chapter 4, the maximum-thickness method assumes a shell mean inner potential of 12 ± 4 V.

showed signs of charging in vacuum), but yielded a mean inner potential of 14.5 V with a wide statistical distribution [50], possibly due to the surface effects discussed here in addition to the factors considered in that work. This value agrees closely with the values determined experimentally here. The experimental data is imprecise, but, according to the DFT simulations, if the surface were in a random state, there might be larger differences between experimental measurements. Possibly, this agreement between experimental measurements is due to a factor not considered quantitatively here, such as charging effects on surface charge distributions, or inadvertently similar surface preparations.

However, ideally, exact surface conditions would be either known or prepared for the nanowire TEM specimen. Therefore, the best method to investigate surface effects on the mean inner potential may be to take a specimen and modify the surface *in situ*.

The next chapter, Chapter 6, documents an attempt to perform an *in situ* hydrogen-termination of an InAs nanowire surface, following a literature protocol as closely as possible, and the end result is not surface modification, but nanowire dissolution.

Chapter 6: Nanowire Dissolution

Outline

Chapters 4 and 5 lead logically to Chapter 6. In Chapter 4, the mean inner potential of an InAs nanowire is measured, and, in Chapter 5, the surface state is shown to have a large effect on the mean inner potential. However, the exact surface state of the nanowire specimen examined in chapters 3 and 4 is unknown. Therefore, surface control in the TEM on nanowire specimens would allow to compare DFT simulations with experimental measurements, leading both to improved accuracy of mean inner potential measurements, and to experimental analysis of the surface-dependence of the mean inner potential. Therefore, the surface control objective is to remove the surface layer from the nanowire and replace it with a hydrogen termination, using a recipe reported previously [132].

This chapter discusses a set of experiments aimed originally at surface control in the TEM. However, instead of modifying the surface, the nanowire dissolves, providing information about how the nanowire system behaves at elevated temperature. The original experimental attempt at nanowire surface control on InAs/InP nanowires is described, and then a set of repeated experiments on GaAs nanowires are detailed whose goal is to characterize the dissolution behavior. These experiments show the fundamental and unavoidable challenges in analyzing nanowire dissolution behavior using *in situ* TEM. This is not a comprehensive examination of the dissolution behavior, but could serve as the foundation for future work.

InAs Experiment: Proof of Concept

Motivation and Experiment

The InAs experiment and results, focusing on the dissolution behavior, are reported in **Paper III: Atomic resolution imaging of *in situ* InAs nanowire dissolution at elevated temperature.** In this section, the experiment in **Paper III** is briefly discussed: an InAs nanowire specimen was heated in the environmental TEM (ETEM) at DTU Cen with hydrogen gas, and the nanowires dissolved. However, for a full summary, the reader is directed to **Paper III**.

This work was motivated by a previous nanowire surface cleaning report, achieved under different conditions than those achievable in the ETEM. This previous work reported nanowire surface cleaning for scanning tunneling microscopy on InAs nanowires using hydrogen and elevated temperature [132]. In [132], the nanowire surface is cleaned using atomic hydrogen at a pressure around 10^{-6} mbar, and nanowire temperatures of 350-400°C, and the scanning tunneling microscope verified the surface termination.

This appeared feasible to attempt in the transmission electron microscope as an *in situ* experiment. The ETEM allows for gases to be injected directly into the specimen-containing region between the objective lens polepieces, and specimen holders (Gatan model 652) that can heat the specimen *in situ*. This approach has been used fruitfully both for general *in situ* TEM experiments, with several examples in the review paper [133], and specific Si nanowire growth examples. The Si growth experiments have examined on ledge-like growth [134] in an instrument similar to the ETEM, and also under a lower-pressure regime where both catalyst dynamics [135] and sidewall morphology [136] have been monitored during growth. While the temperature range

used in [132] was feasible (350-400°C), only molecular hydrogen at millibar pressures was available in the ETEM. InAs nanowires similar to the nanowires used in [132], also including short InP segments, were also available.

When this experiment was attempted *in situ* to the best possible extent, unlike in the previous work, surface cleaning was not observed. When no reaction was seen at lower temperatures[40], the temperature was increased, and nanowire dissolution resulted, primarily taking place at the interface between the catalyst particle and the nanowire. The full movie of this behavior is included as Video V6.A[41], and **Paper III** details the exact features seen in Video V6.A, including nanowire dissolution and indium oxide particle formation; the particle formation was not seen on subsequent experiments, and will not be discussed further here.

This difference in effect between the UHV and TEM experiments is possibly due to beam effects, environmental differences, or non-uniform heating. Some of these effects are discussed later in this chapter for GaAs nanowires. The high-energy electron beam may be disrupting the surface termination, which is relatively unlikely to be completely explanatory given that the beam was turned off or exposing only some of the nanowires on the grid during the attempted surface modification, and other nanowires on the grid also showed signs of dissolution (see GaAs nanowire dissolution discussion later). The difference between atomic and molecular hydrogen is likely a critical factor; while the beam or the heating filament in the specimen holder may be cracking some of the H_2, it is not clear how much of that cracked H^+ would reach any individual nanowire on the grid. The pressure differential (6 orders of magnitude between the millibar-range TEM work and 10^{-6} millibar-range experiment in [132]) may also play a role, but the higher ETEM pressure should, in principle, allow for more surface cleaning interaction. The carbon film on the specimen grid might also be interfering with clean surface formation. The specimen holder heating filament may give non-uniform heating across the specimen, and means the read-out temperature from the holder should be close to, but is likely not exactly, the temperature at the observed point on the specimen. In addition, the surface layer on the InAs nanowire is unknown, and may be somewhat different from the surface layer on the nanowire reported in the literature. Any or all of these factors may explain the lack of surface cleaning observed in the TEM experiments.

Following those experiments, the dissolution behavior itself became the focus of characterization. The nanowire dissolution behavior could yield useful information about the nanowire growth dynamics and processes due to the elevated temperature at which dissolution takes place. Also, GaAs nanowires were used for the follow-up experiments, due to a limited supply of InAs nanowires, the inherent destructive nature of the dissolution experiments, and previous literature on GaAs nanowire heating in the TEM, which showed the particle changing shape and consuming part of the nanowire at elevated temperature [137].

[40] As also noted later in this chapter, the temperature control readout is based on the filament temperature, and will likely be different at the specimen location itself.
[41] Acquired with the assistance of Jörg Jinschek.

GaAs Experiment: Dissolution Analysis

Objective of GaAs Nanowire Dissolution Experiments

The objective of the GaAs nanowire dissolution experiments was to attempt to quantitatively characterize the nanowire dissolution behavior during the dissolution process, and figure out what challenges are faced. A previous report in the literature [137] examined GaAs nanowires at elevated temperature in the TEM, which showed the particle changing shape and absorbing part of the nanowire, but [137] did not study this behavior dynamically, instead performing a before-and-after study. However, the before-and-after effects reported in [137], while consistent with dissolution, are relatively minimal when compared with *in situ* observation of InAs nanowire dissolution in the previous section.[42]

Multiple experimental factors may influence GaAs nanowire dissolution. However, only two members of the large set of possible parameters are examined here in the interests of brevity and exploration: heating temperature and beam dose rate. More parameters (as noted at the end of this chapter) could be examined, but these two parameters provide insight into how to characterize the dissolution behavior.

General Results of GaAs Nanowire Dissolution Experiments

In this section, the general results from GaAs nanowire dissolution experiments are detailed. The nanowire dissolution behavior is described, where the catalyst particle moves down the nanowire, with some parts of the nanowire spontaneously decomposing. In the following section, two specific factors – specimen heating temperature and electron dose rate – are discussed.

The experiments in this section were carried out on nanowires all harvested from the same nanowire substrate specimen (grown at 450°C[43]). The nanowires were mechanically harvested onto Au grids with holey carbon films, similar to the procedure undertaken in chapters 3 and 4. The grid containing the nanowires were then inserted into the ETEM, which was operated in high-vacuum mode (without additional gases introduced to the specimen region)[44], and the specimen heated *in situ*, with the output acquired in bright-field TEM mode[45] and recorded digitally on a CCD camera. Representative GaAs nanowires were selected for study, where the catalyst particle on top of the nanowire was also attached to the carbon film. The carbon film allowed for features on the film to be used as landmarks for tracking the dissolution behavior.

First, a representative dissolution experiment performed on GaAs highlights the routine features of dissolution experiments, shown in Figure 6.1 and in Video V6.B. In the TEM, the specimen grid with nanowires was heated to 480°C, and the nanowires were observed to dissolve. The dissolution behavior primarily takes place at the interface between the Au particle and the GaAs nanowire. The Au particle propagates along the nanowire length, usually absorbing the nanowire in discrete

[42] Some of the GaAs dissolution results have been previously presented in [138], a non-peer-reviewed conference poster on which the author of this dissertation was the primary author.
[43] GaAs nanowire specimens provided by Jesper Wallentin at Lund University.
[44] The ETEM was used to ensure that, in the case that further *in situ* gas experiments were to be of interest, the same microscope would be used for all dissolution experiments.
[45] Without the use of an objective aperture, due to the desire to retain interpretable dose rates in the images if necessary.

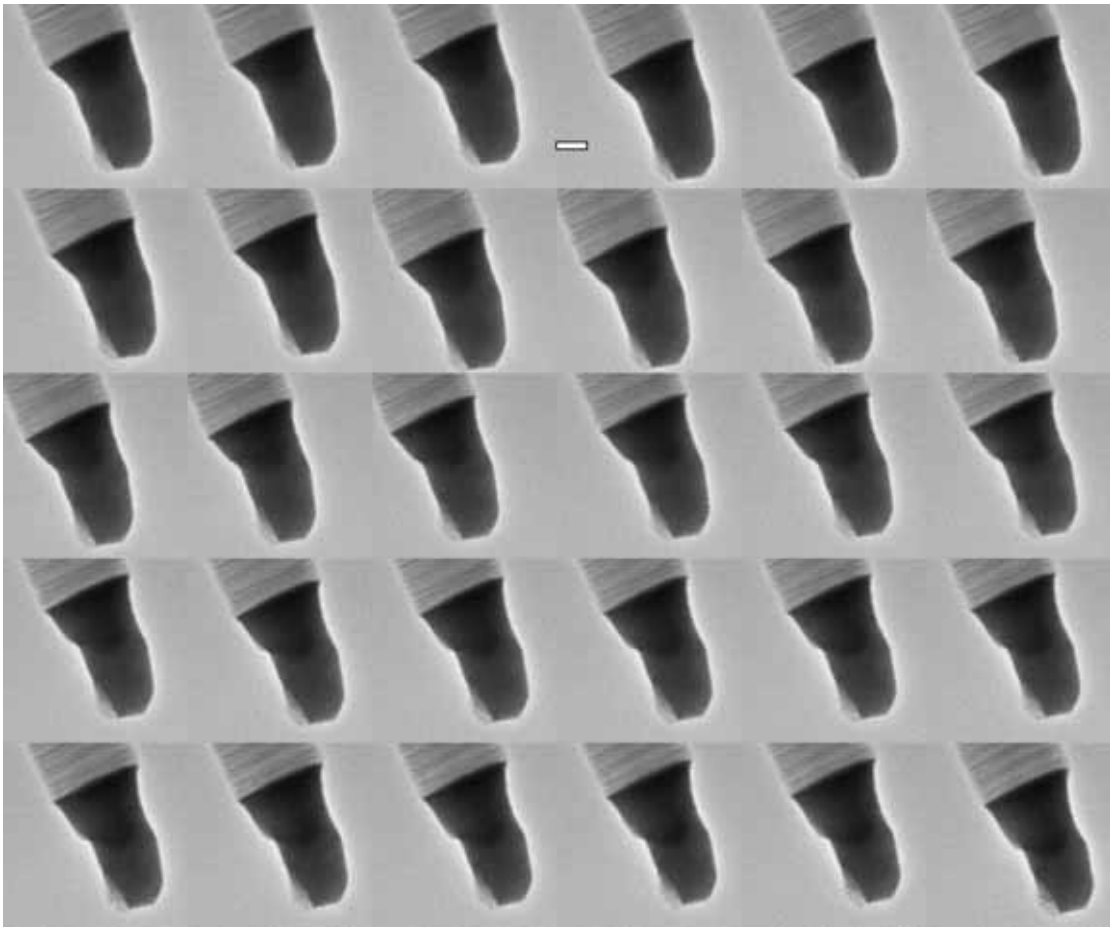


Figure 6.1: GaAs *in situ* nanowire dissolution, typical TEM image series. GaAs nanowire dissolution at 480°C; 1 frame per minute, grid of 30 images (6×5, chronologically across, then down). The scale bar is 25nm. See Video V6.B. The nanowire is light grey (upper left), the Au particle dark grey (middle). In this sequence of images, the particle is dissolving the nanowire in relatively discrete increments (the crystallographic defects in the nanowire can be followed). The particle reshapes as it dissolves the nanowire.

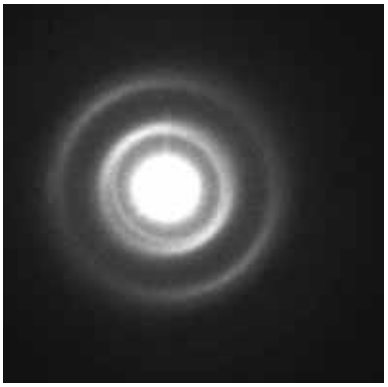


Figure 6.2: Uncalibrated TEM diffraction pattern taken from Au particle during dissolution. Misalignments in the imaging system likely led to the split rings seen. However, the presence of rings indicates polycrystallinity in the particle.

intervals that are possibly related to atomic planes, similar to a reversal of the ledge-like growth process seen for Si in [134]. The particle appears to probably remain crystalline, as seen in Figure 6.2 – not conclusively so, however. This front-like behavior allows for dissolution rates to be measured. The nanowire material appears to enter the particle, likely creating a eutectic of Au-Ga-As as is normally seen in Au catalyst particles [95], and the nanowire material exits the particle through an unknown mechanism. The particle adheres to the C film and dynamically reshapes over time, making exact knowledge of the particle geometry difficult. Additionally, parts of the nanowire far from the Au particle decompose during exposure to the 480°C environment.

The next section analyses the quantitative effects on the dissolution rate from modifying the heating holder temperature and the dose rate of the electron beam.

Temperature and Beam Effects

From this baseline study of dissolution, two factors affecting the dissolution rate were investigated: temperature, and beam dose rate. Dissolution rates were analyzed by tracking the dissolution front along as many distinguishable points as possible: usually four points were tracked, but at least two tracking points were used. This approach might lead to a mis-estimation because the projected rate of dissolution is tracked instead of the real linear dissolution rate along the nanowire, but, while the nanowire/particle interface was not edge-on, it was close, meaning that the effects are minimal in this study (but would be something to consider in follow-up work).

Temperature poses a specific challenge: the read-out temperature from the specimen holder is not likely to be the same as the actual temperature of the specimen region being imaged. Therefore, the temperatures quoted here are not to be taken as the local temperatures at the region of interest. Even with this quantitative caveat in mind, the temperature can be selected on the heating holder, and the resulting effect on the dissolution rate can be observed.

Investigating temperature dependence leads to four conclusions. First, nanowires heated at 400°C, seen in Figure 6.3, show clear signs of dissolution, indicating that dissolution occurs at a range of temperatures. Second, comparing Figure 6.1, heated at 480°C to Figure 6.4, heated at 560°C, shows that the specimen exhibits increased dissolution rate with increased temperature. This is unsurprising, due to the increase in thermal energy with temperature. Third, as seen in Figures 6.5 and 6.6, the dissolution behavior occurs at elevated temperature even when the nanowires have not been exposed to the beam.

The fourth conclusion from the temperature dependence is that temperature can change how the dissolution front propagates along the nanowire. This can be seen in Figure 6.4, taken at 560°C, where the dissolution does not appear to move discretely, or "layer-by-layer". At lower temperatures (*e.g.* 480°C, discussed in the previous section), the dissolution tends to propagate in discrete intervals, absorbing each new "layer" quickly before moving to the next one. As the temperature increases, the dissolution front has a higher probability to propagate further down the nanowire rather than finish absorbing its most recent "layer", so the dissolution front becomes non-planar. At 560°C, as seen in Figure 6.4, this non-planar behavior makes the dissolution front difficult to track. This limits the analyzable temperature range,

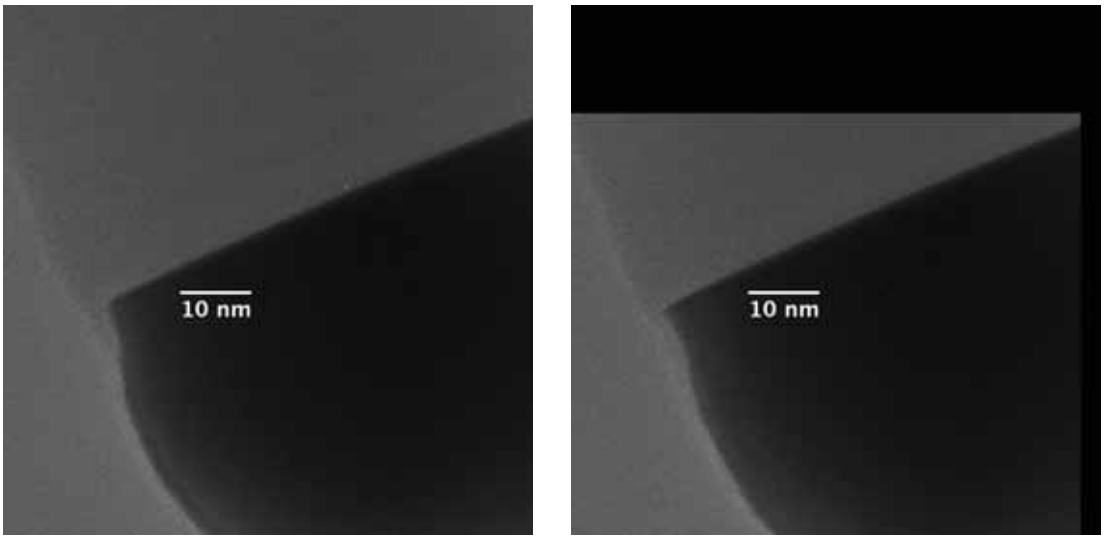


Figure 6.3: GaAs *in situ* TEM nanowire dissolution behavior at *t* (left) and *t*+10 (right) minutes at 400°C. The black frame on the right TEM image is the result of drift correction. The corner of the Au particle can be seen to migrate into the wire, showing that dissolution takes place even at this temperature. The dissolution rate is slower than that seen at 480°C (*cf.* Figure 6.1).

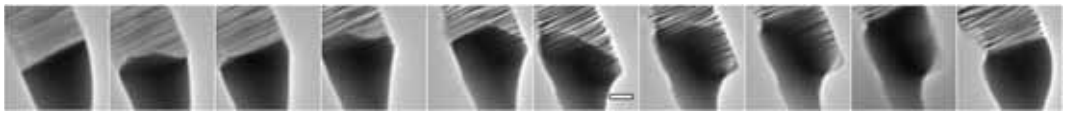


Figure 6.4. Nanowire dissolution behavior *in situ* at 560°C. The scale bar is 25nm. 1 minute interval between TEM images. Dissolution at elevated temperature is not as planar as at 480°C (*cf.* Figure 6.1). The Au particle consumes the wire much faster than at 480°C, and in a way that is difficult to characterize – the last four images show the nanowire being consumed in the beam direction. Originally presented in [138].

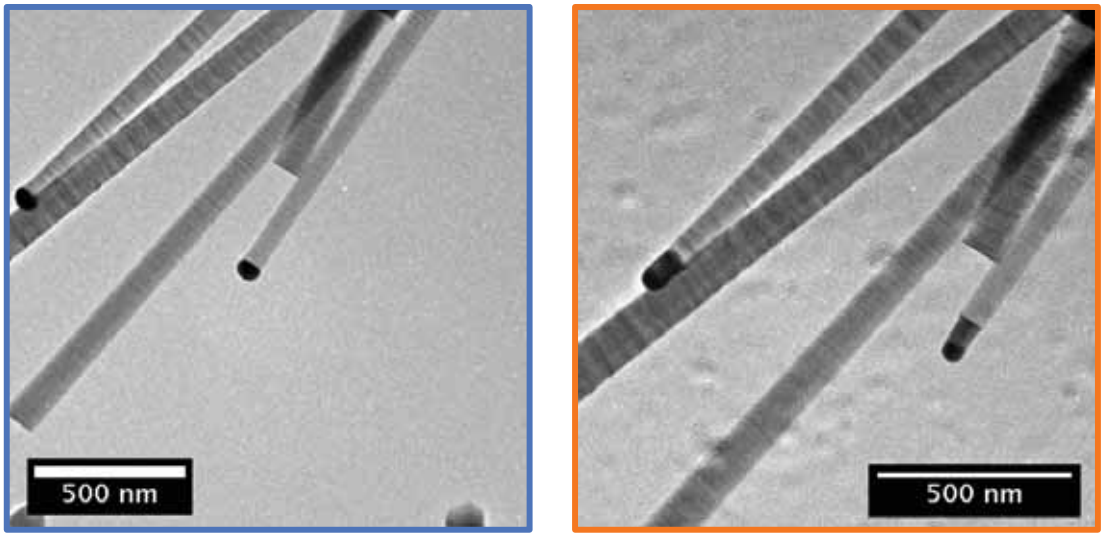


Figure 6.5: GaAs nanowire dissolution without electron beam exposure. TEM images before heating (blue) and after heating (orange) acquired from nanowires not exposed to the electron beam during heating. The specimen went from room temperature to 560°C and back to room temperature before imaging took place. This pair of images shows nanowire dissolution without beam exposure while at elevated temperature.

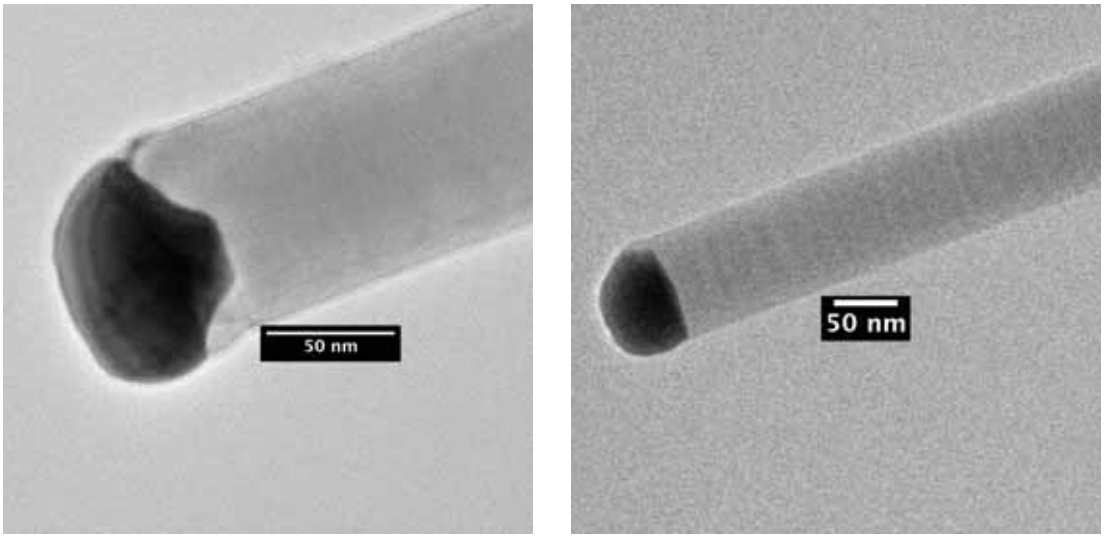


Figure 6.6: GaAs nanowire before *in situ* TEM heating (left) and while at 480°C without prior beam exposure at elevated temperature (right). The Au particle appears to dissolve the nanowire radially before dissolving axially. As with Figure 6.5, dissolution occurs even without electron beam exposure; as with Figure 6.4, it would be difficult to characterize the dissolution change between the pair of images quantitatively. Technically, the right image was the second one taken at elevated temperature, but it immediately followed the first.

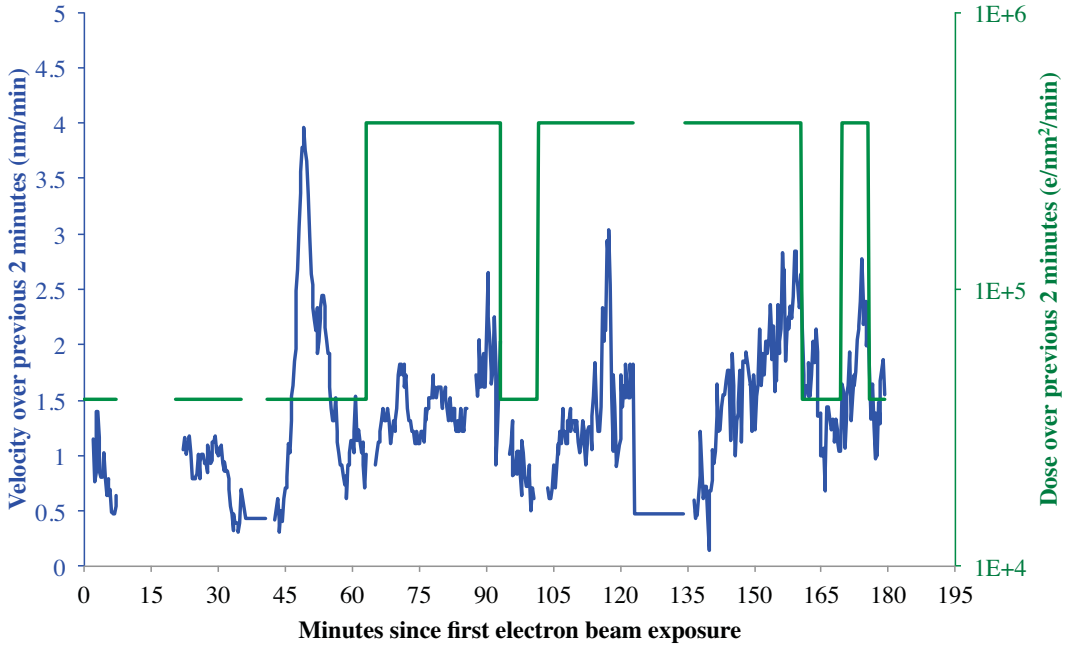


Figure 6.7: Quantification of the nanowire dissolution front propagation rate and *in situ* TEM electron beam dose at 480°C. The blue line shows the dissolution rate, as measured from the nanowire dissolution front. The green line shows the electron beam dose (breaks in the line mean the beam was not exposing the specimen). The dissolution rate appears to speed up with increased electron dose. However, the dissolution rate is noisy. The spike around 45 minutes was at the same time as the Au particle reshaping. Originally presented in [138].

because front-like motion is easier to quantify and analyze than non-planar movements of the Au particle into the nanowire.

Electron beam effects pose a different sort of challenge. The dissolution behavior still occurs on nanowires that have not been exposed to the electron beam, detectable in before-and-after room-temperature TEM images in Figure 6.5 and in images taken immediately after heating, shown in Figure 6.6. However, does the beam itself have an effect on the dissolution behavior? This is difficult to answer *in situ*, because the beam is required for TEM characterization. However, to investigate this effect, the electron beam was alternated between three electron dose rates at 480°C, and the rate of dissolution observed.

The effect of the electron beam dose rate is apparently to catalyze the dissolution reaction. The three electron dose rates used were 0 (beam off), 4×10^4 and 4×10^5 electrons per nm^2 per second, using calibrations provided by the CCD manufacturer (Gatan). The results can be seen in Figure 6.7, where the dissolution rate was taken as the rate of change in the dissolution front over the previous two minutes, to smooth the fluctuations and increase the reproducibility of front-tracking measurements. The rate at the 4×10^4 dose rate was about 0.5-1 nm/minute, and the rate at the 4×10^5 dose rate was about 2-3 nm/minute. However, as can be seen from this graph in Figure 6.7, the dissolution rate does not stay constant, even at a single dose rate: it fluctuates significantly partly linked to dynamic particle reshaping, and perhaps also due to the stochastic nature of the dissolution absorbing the nanowire plane-by-plane. These trends may be linked to simple beam heating – where the beam interaction imparts kinetic energy to the specimen, or cumulative damage effects might build up or dissipate at appropriate dose rates. While the cause of the dose rate dependence is not known, there is a clear electron beam effect on the nanowire dissolution rate that must be considered in future work.

Comparing these two factors illustrates the complications with quantitative nanowire dissolution analysis. Temperature is under full control – the nanowires can be examined at room temperature, where dissolution does not take place, or at elevated temperature, where it does. Different temperature-dependent dissolution characteristics can be mapped out, such as the dissolution rate or the front-like versus non-planar dissolution-propagation behavior. In contrast, the beam effects are more troubling. Because the electron beam must be used to characterize the specimen, the effect of the beam should be carefully disentangled from the effects from the other factors in the experiment. Constant dose rate must be used if other factors are being varied, but the effects of that particular dose rate should be known. Therefore, while temperature is a primary driving factor behind the nanowire dissolution, beam effects appear unavoidably catalytic, in an unspecified manner.

Complications with Quantitative Dissolution Analysis

However, the dissolution behavior itself resists quantitative *in situ* TEM analysis for three major reasons. These reasons make some analysis not possible, but other types of quantitative analysis can still be done through measuring *e.g.* nanowire length as a function of time. First, the projection nature of TEM images creates a fundamental limit on the interpretability of the dissolution data. Second, the products of the dissolution have not been detected. Third, the beam effects deeply complicate the ability to characterize the dissolution behavior, noted in the previous section.

One factor seen in these nanowire dissolution experiments limits quantitative analysis: the electron microscope only obtains two-dimensional images of three-dimensional objects. Because the catalyst particle is dynamically reshaping as it dissolves the nanowire, it is not possible to know the particle's exact geometry or thickness, and characterization techniques in the TEM generally do not reveal full three-dimensional information without reconstruction. During nanowire dissolution, the interface between the nanowire and the particle does not stay edge-on with respect to the beam, as can be seen in Video V6.A, where it tilts the nanowire from laying on the substrate to fully edge-on in the beam. This change is difficult to compensate for, as a non-edge-on interface must be guessed at. Because the particle appears to be crystalline, energy-filtered imaging to determine the projected thickness of the catalyst particle would be subject to diffraction contrast, and would also be subject to variations in the specimen grid's carbon film thickness, but is probably the best option. Energy-filtered imaging was not used here due to a desire to avoid adjusting the energy filter during use. Additionally, the particle/wire interface would still be slightly misoriented (*i.e.* not edge-on with respect to the beam), which might pose some challenges for quantitative interpretation of the full projected mass of the particle. HAADF STEM, discussed in a previous chapter, is a serial-recording imaging technique, not parallel, limiting its ability to characterize dynamic behavior – and the focused probe beam may disrupt or influence the dissolution behavior. It would not be impossible to use HAADF STEM – but probably unfavorable. However, any image, even if it yields the thickness directly, will have several additional challenges. The image intensity might not be easily related to nanoparticle thickness, because the contours of the underlying carbon film are unknown. The shell around the nanowire will likely play an additional role; it appears to be left behind after dissolution (as can be seen in Video V6.A), and what role the shell plays in dissolution rate is unclear. Taking all of this into account, the properties of the particle itself would be a major objective of future work.

Additionally, it is unknown what happens to the nanowire dissolution products. The nanowire dissolution products are not being produced in large amounts, because the total volume of nanowires on each TEM grid is small. The nanowire material is likely entering the particle, because the image intensity appears consistent throughout the particle, indicating a likely uniform composition. The nanowire is providing a source of additional material to the particle, but the particle does not appear to grow continuously. Therefore, there must be some sink for the Ga and As in the particle that probably maintains a steady-state Au/Ga/As concentration – but the nature of that sink is unclear. Secondary crystallite formation, like that seen in the InAs experiment detailed above, is not observed.

Conclusions and Future Dissolution Work

While the initial goal was to modify the surface of the InAs nanowires, nanowire dissolution occurred, and this dissolution was characterized quantitatively. In this brief report of dissolution behavior in nanowires, as observed using *in situ* TEM at elevated temperature, several characteristics are important. When InAs or GaAs nanowires with as-grown Au growth catalysts are heated in the TEM, they dissolve along the Au-semiconductor interface in a layer-by-layer fashion. Additionally, the catalyst particle on the nanowire adheres to the carbon grid, and parts of the nanowire sidewalls decompose. No surface modification effects are noted in the ~4nm thick nanowire surface layer, probably due to differences in the pressure regimes: the reported surface preparation [132] was at pressures around 10^{-6} mbar, while this work used higher pressures (~1 mbar) due to *in situ* TEM limitations. On GaAs nanowires,

temperature has a primary effect on the dissolution – it governs whether dissolution occurs or not – but the electron beam used for observation also has a dose-dependent measurable rate-increasing effect as well.

This is only a preliminary study; there is clear scope for future work in this area. A better method for dissolution characterization is necessary, to attempt to capture the dissolution rate and account for the dissolution products. The dissolution characteristics are influenced by temperature and electron beam dose rate; the dissolution might also be influenced by many experimental parameters. Some of the parameters to be investigated in *in situ* TEM follow-up work are gas species, gas pressure, electron beam accelerating voltage, the total electron dose, carbon film composition, nanowire growth conditions, among others. Some of the nanowire parameters that might be relevant for dissolution behavior include nanowire composition, the presence/nonpresence of Au catalyst particles, growth temperature, and nanowire diameter. Nanowire dissolution is a complex experiment with many active parts, but might, in time, either lead to TEM surface control on a semiconductor specimen or useful technical information about nanowire growth through high-temperature dynamics.

Chapter 7: Conclusions and Future Work

Conclusions

In this dissertation, both electron holography and density functional theory (DFT) are used to explore the factors affecting the mean inner potential. First, an InAs nanowire is selected, and this nanowire's cross-sectional thickness profile is determined using multiple techniques. Then, a specimen-tilt-controlled series of holograms, spanning both strongly and weakly diffracting conditions, is compared with image simulations, ultimately determining mean inner potentials, mean free paths, and absorption parameters. DFT simulations are then performed for group-IV and III-V materials in "thin-film" and "nanowire" geometries, demonstrating strong surface effects on the mean inner potential. When surface modification on InAs and GaAs nanowires is attempted, nanowire dissolution occurs and is characterized.

In this chapter, these results are briefly summarized, and the scope for future work involving mean inner potentials measured using electron holography is outlined.

Thickness Determination

In Chapter 3, different methods of thickness determination suitable for mean inner potential measurements on nanowires were considered and applied to a single InAs wurtzite-structure nanowire. The methods chosen needed to be quantitative, with the ability to estimate both random and systematic error, with the output being the cross-sectional thickness profile of the nanowire. A suite of tomographic techniques for cross-section determination was applied – algebraic tomography (AT) algorithms, including weighted back-projection, the algebraic reconstruction technique, and the simultaneous iterative reconstruction technique, geometric tomography (GT) and discrete tomography (DT) – as well as a non-tomographic technique that was proposed: the renormalized orthogonal-images routine (RENOIR), which calibrates the intensity by using intensities and widths from an ensemble of orthogonal pairs of images in the tilt series. AT algorithms performed poorly for cross-section determination due to necessary post-reconstruction thresholding. The GT algorithm used here performed poorly for cross-section determination due to missing-wedge information; this might be alleviated with a different algorithm. The two thickness determination techniques that yielded a usable thickness profile for the entire nanowire were the DT algorithm and RENOIR. The maximum thickness, along the plateau of the nanowire, measured using the DT method was 126 ± 2 nm and with the RENOIR method was 130 ± 2 nm. These do not agree. However, there is not a clear method to resolve this discrepancy, and both DT and RENOIR are applied to the problem of thickness determination for mean inner potential measurements.

Measuring the Mean Inner Potential

In Chapter 4, the mean inner potential is measured experimentally, considering diffraction contrast systematically, using electron holography on the same InAs nanowire that was used for thickness-determination in Chapter 3. Holograms, taken with long (64s) exposures to improve the signal-to-noise ratio, were acquired and reconstructed into phase and amplitude images. The effect of diffraction contrast was then considered by taking a tilt series of holograms acquired every 0.1° over a span of 11°. After reconstruction, line profiles of the phase, amplitude and amplitude-derived t/λ were measured. All three showed strong diffraction effects across the tilt series.

At a weakly diffracting condition, the phase and the amplitude-derived t/λ did not appear to be linearly dependent on thickness in the original data. This nonlinear effect was resolved when the thin shell surrounding the nanowire was considered as not being measured with HAADF STEM. Thus, the thin shell was not included in the thickness determination but included in the measured electron holography phase and amplitude. After compensating for this shell effect by scaling the DT reconstruction of the nanowire cross-section, the phase and amplitude-derived t/λ were very close to linear in a weakly diffracting condition, and this approach allowed for determination of the mean inner potential and mean free path (λ) of the core and the shell separately. The core data were consistent, but the shell data were too imprecise to allow for quantitative analysis.

The phase, after unwrapping manually, was the focus of analysis. The strong diffraction contrast features seen in the holographic reconstruction were compared with simulations, which reproduced many of the features seen. This allowed for a "weakly diffracting" section of the tilt range, 1° wide, to be isolated for analysis. This weakly diffracting segment of the tilt series was used for quantitative determination of the mean inner potential of the core. The shell was found to have less precise mean inner potentials than the core, and the shell data were generally too noisy to analyze. The inconsistency in mean inner potentials generated from the two thickness profiles (RENOIR and DT) is due to the inconsistency in those thickness profiles; however, both profiles yield mean inner potential data, even though the two thickness profiles are inconsistent.

From the maximum thickness approach, when using a shell mean inner potential of 12 ± 4 V, for the DT thickness, the mean inner potential of wurtzite InAs is 14.6 ± 0.3 V, and, for the RENOIR thickness, the mean inner potential of wurtzite InAs is 14.0 ± 0.3 V. From the synthetic wedge approach, the mean inner potential of wurtzite InAs is 14.5 ± 0.4 *via* DT, and 14.0 ± 0.4 *via* RENOIR. For comparison, isolated-atom scattering factors, as calculated in Chapter 5, would indicate a mean inner potential of 15.47 V. The strongly diffracting conditions were also analyzed by comparing with image simulations, and found to be broadly consistent with simulations of that diffraction contrast. While not all features in the simulation were reproduced, the large-scale features were, and quantitative comparison was hindered by uncertainty over the absorption parameter.

The above analysis was also continued for the amplitude-derived mean free paths as measured using electron holography for both the core and the shell; again, the shell data was too noisy for further analysis. The amplitude information, when compared with image simulation, yielded an absorption parameter a of approximately 0.08 for a weakly diffracting condition, but a=0.1 near the strongly diffracting conditions, and was consistent across both thickness determination methods. This discrepancy indicates that perhaps an improved model for the absorption would aid comparison with simulation.

A separate assessment using DFT simulations of the mean inner potential was performed to understand factors that are not intuitively obvious from the measurement itself.

Simulating the Mean Inner Potential

Mean inner potential simulations were carried out using DFT in both thin-film and nanowire geometries. DFT was used because it provides electronic structure effects not present with the conventional isolated-atom scattering factor approximations. However, the DFT simulations also show that the surface of the material has a strong effect on the mean inner potential, even as measured between the middle of the thin-film and the region outside the thin-film. This is not due to imprecisions in the simulations; the simulations have a margin of error due to simulation parameter choices in the mean inner potential of an average of ±0.06V in the best-case scenario, as compared to a margin of error of ±0.4V in the experimental measurements. While the DFT simulations are more precise than the experimental measurements, the DFT simulations and experimental data have different factors affecting their precision.

The minimum-energy lattice parameters calculated using the DFT software package called GPAW are also considered. These are the lattice parameters that are the minimum-energy configuration for the DFT calculation, but they differ from the experimental ones for all cases studied here. The magnitude of this difference varied for different materials; diamond C had the lowest discrepancy, and GaAs the highest. This discrepancy yields a minimum-energy lattice parameter that is +0.28% larger than the experimental one for C and +2.12% for GaAs. The minimum-energy lattice parameters must be used for the structure-optimization routines, to avoid building in strain into the material.

Modifying the surface changes the mean inner potential. If the surface atom configuration is changed, or surface adsorbates are added, the mean inner potential will change. The following examples, all from diamond C using the minimum-energy lattice parameter, illustrate the effect. Changing the bulk-terminated surface facet yields mean inner potential changes - for a C{110} surface used as the baseline: 20.23 ± 0.04 V; for a C{001} surface: 21.26 ± 0.04 V, for a C{111} surface 21.04 ± 0.04 V. Optimizing the C atom positions for the entire thin film yields 19.72 ± 0.04 V for a C{110} surface, and 21.07 ± 0.04 V for a C{001} surface. Adsorbing H onto this surface and optimizing the structure yields 16.12 ± 0.04 V and adsorbing O and optimizing the structure yields 22.19 ± 0.04 V. Changing the bulk-terminated surface facet changes the mean inner potential, and optimizing the surface structure changes the mean inner potential more. However, the largest change is when surface adsorbates are added; the difference between the H and O adsorbate cases is very large, and potentially experimentally detectable. This surface dependence is likely most attributable to changes related to the work function. If a two-surface model is examined (a “nanowire” as opposed to a “thin film”), surface effects can be envisioned for bulk-like terminations, including fringing fields. The effect of the fringing fields appears reduced when the atom positions are relaxed; however, only one simulation showing this for C and one simulation for Si is presented, so conclusions should be limited. This work discusses several of the general implications of this surface dependence.

In short, if the mean inner potential DFT simulations correctly model the experimental mean inner potential, then experimentally measured mean inner potentials would include specimen surface state information. If the simulations are accurately reflecting the surface dependence, then the surface state can modify the measured mean inner potential by large amounts without this surface state being directly measured during the measurement. This would mean that any quantitative

mean inner potential measurements that do not consider the potential states of the specimen surface would be possibly problematic. This surface dependence might help explain the range of mean inner potential values cited in the literature for measurements on the same materials, and the reason the literature mean inner potential values do not vary as much as some of the simulations would predict may be due to the effects of relaxation, as seen with fringing field suppression in the DFT nanowire simulations.

Attempting Surface Modification and Nanowire Dissolution

Finally, surface modification on an InAs nanowire is attempted, based on a surface-preparation recipe reported previously under different conditions. The procedure does not yield surface modification, but nanowire dissolution instead. This dissolution is briefly outlined, and the effects of temperature and the TEM electron beam are examined, but a full quantification of these effects is outside the scope of this work. Increasing temperature appears to increase dissolution rate, and increasing the electron beam dose rate also appears to increase the dissolution rate, but full quantification is difficult due to the two-dimensional nature of the images acquired, the effect of the support and the thin shell surrounding the wire, and the dynamic nature of the dissolution behavior.

Future work on nanowire dissolution might prove productive in understanding the dynamics of this system at elevated temperature. While the dissolution conditions are not the growth conditions, analyzing the results from these dissolution experiments might yield insight into the growth kinetics of the nanowire system and their properties, as long as clear routes to quantification can be established. The dissolution behavior might also provide more insight into the properties of the nanowire, such as the nature of the thin amorphous shell that surrounds the wire.

Future Work

Two avenues for future work present themselves. First, further comparisons between tilt-controlled holograms and image simulations of amplitude and phase would allow for improved image simulations and improved fundamental understanding of image formation mechanisms and *e.g.* the Stobbs factor (discussed in Chapter 2 as the mismatch between simulation and experiment for high-resolution TEM). Second, further exploration of the surface dependence of the mean inner potential would potentially allow for surface-sensitive mean inner potential measurements, and more fundamental knowledge about the mean inner potential.

Additional future work, not elaborated here, could include several other topics. Improved tomographic algorithms and non-tomographic techniques, including possibly more complete error determination and elaboration for RENOIR and DT, would aid in quantitative thickness determination. Improved goniometer sensitivity would be helpful for tilt-controlled experiments. For dissolution experiments, energy-filtered imaging and other one-shot thickness-determination techniques might be useful, and the role of different factors on nanowire dissolution seen here should be examined.

Future Work: Amplitude and Phase Characterization

The work done here shows correspondence between experimental measurements of amplitude and phase using electron holography, and their simulated counterparts. Tilt-controlled holography allows for reciprocal-space maps of amplitude and phase. This

is performed for one specimen in this work; future work might focus on repeating the experiment for different specimens, and improving the simulation framework to match the experimental features through *e.g.* improved absorption parameter modeling.

Building upon the foundation outlined here, similar experiments could be performed for a range of additional specimens. By repeatedly applying this tilt-controlled holography methodology to different materials – linearly and densely sampling a large portion of reciprocal space about a large diffraction feature, with many electron holograms – and then comparing with simulations, several extensions to current models might be possible. The change in absorption parameter with tilt angle seen here can be investigated, possibly leading to better absorption models. After a better absorption model is applied, the simulated and experimental phases might be more amenable to quantitative matching. However, this may also require improving the specimen orientation determination to the point where each orientation is known especially for highly tilt-sensitive regions (*e.g.* near zone axes or systematic rows).

If this is done, it could be theoretically interesting. By acquiring more tilt-controlled holography datasets that can be compared with simulations following the process applied to the tilt-controlled data in Chapter 4, additional future experiments may contribute to the fundamental understanding and improvements of TEM image simulation techniques and knowledge, including the aforementioned Stobbs factor, allowing for improved quantitative comparison between experiment and simulation.

Future Work: Surface Dependence Of The Mean Inner Potential

There is clear scope for future measurements and simulations of the mean inner potential, but the surface dependence predicted in this work should be explored. As this work shows, it is possible to measure the mean inner potential using electron holography systematically and quantitatively, including accounting for diffraction effects using image simulations, and this can be done in the future for materials of interest. However, the role of the surface, largely neglected in the literature but explored in this work using DFT simulations, should be explored experimentally, and the surface dependence predictions in this work should be investigated. However, if this surface effect is verified experimentally, then electron holography could be used to quantitatively answer questions about surface structure.

If validated, this surface dependence is critical because it questions the fundamental limits of mean inner potential measurements. If the mean inner potential of a specimen is surface dependent, even to an unknown degree, then the mean inner potential is not a fundamental material property, but a property of the material plus its surface. It would be insufficient information to only specify the material.

Control of surface formation (*e.g.* native oxides) for mean inner potential specimens should be attempted. While the DFT simulations predict a wide range of possible mean inner potentials for a given material depending on surface condition, the spread in experimental measurements of mean inner potential is generally narrower. This relative convergence of the experimental data may be due to similar surface preparations inadvertently being used for experimental specimens – *e.g.* if every cleaved-wedge specimen surface is a native oxide due to air exposure after cleaving, but the exact environmental conditions under which this native oxide grows are not the same, it is perhaps reasonable to assume that the oxide structure at the cleaved-

wedge surfaces might be different. Even if the properties of this oxide are unknown, if it is systematically prepared, it might yield precise data. Therefore, controlling the conditions under which the specimen surfaces are formed may improve reproducibility of mean inner potential measurements, even if the surfaces are not understood.

Understanding and applying the hypothesized strong surface dependence of the mean inner potential predicted in the DFT simulations needs both additional DFT simulation and experimental verification. Additional DFT simulations, with the best possible surface structure predictions, should be used to screen and identify systems for mean inner potential measurement, as well as to understand the factors affecting this surface effect. DFT is an ideal tool to continue to use for this, because it allows for the continued study of mean inner potentials using realistic electronic structure while simultaneously knowing the exact surface structure under study.

Measuring this surface dependence would require one of three cases. Most preferably, this would need a system whose surface atom positions can be determined *in situ* – perhaps using a known surface-preparation recipe, an ultrahigh-vacuum TEM, and a specimen surface that is resistant to beam damage. The second-best alternative is a system that can have a surface modification *in situ*, allowing for a before-and-after comparison. The third-best alternative would be to have a system whose surface can be characterized *ex situ* through *e.g.* scanning tunneling microscopy before being transferred to the TEM for characterization. In all of these cases, DFT simulation is required to understand the effects on the mean inner potential. However, the first case can be simulated from the acquired surface atom position data – using the atom positions and chemical composition as-is is ideal. In the second case, a range of different surface-dependent simulations are necessary to understand the mean inner potential's possible range; however, this might be acceptable if a dramatic change in the mean inner potential is observed – this is no guarantee of success, given that the exact surface structure is not known. In the third case, this might be acceptable, if the surface itself did not change between the *ex situ* characterization and insertion into the microscope, or under irradiation. However, to reiterate, all of these cases require mean inner potential simulation using DFT to understand the surface structure effect on the mean inner potential for the given systems.

However, if this posited surface dependence is experimentally verified, it changes how electron holography can be used to characterize materials. In this case, DFT should be used for pre-screening of materials before mean inner potential measurements are undertaken, to ensure that the magnitude of the surface effects on the mean inner potential is known. This is a fundamentally different approach to the one taken in the past. If there is an unknown strong surface-dependent mean inner potential for a given material, and this surface dependence is not known, then a single mean inner potential measurement cannot be accurately compared to other mean inner potential measurements.

Not only would this surface dependence improve the theoretical understanding of the mean inner potential and electron holography, but also this strong surface effect could inhibit some experiments, and facilitate others. Experiments that would be inhibited are those that rely on a mean inner potential measurement when the surface is unknown. If DFT simulations are performed for all possible surface configurations, including those possibly generated under irradiation, then the data generated from

these experiments might be more interpretable and understandable. In light of this, semiconductor materials may be suboptimal for mean inner potential characterization if their surfaces are not known.

Novel experiments that would be possible are those where the surface is cleanable, or the atom positions in the entire structure are known, *e.g.* for nanoparticles. Cleanable surfaces would present a known target for DFT simulation to match. Ideally, the system would be one that has also been studied with DFT extensively, to match with the electron holography focus. However, these specimens would come with their own specific challenges – among others, thickness determination and orientation determination.

Summary Of Future Work

In summary, mean inner potentials are unavoidably affected by several factors: thickness determination, diffraction contrast, and perhaps surface effects. The best possible mean inner potential measurements include optimized signal-to-noise ratios, and optimized, verifiable, quantitative thickness measurements and orientation measurements for diffraction contrast modeling. Tilt-controlled electron holography can shed light on factors affecting the simulation of amplitude and phase. This diffraction contrast modeling should be improved, for both amplitude and phase. With improved modeling, quantitative comparison between experimental images and image simulation may be improved, allowing for new possibilities in both these fields.

If, as predicted from DFT simulations, the mean inner potential is experimentally verified to be surface dependent, future experimental mean inner potential measurements will require careful reflection. Either control or knowledge of surface properties would be required, along with mean inner potential DFT simulations. Both simulation and experiment must include transparent error estimates to allow for inherent inaccuracy and imprecision. If these are done, then mean inner potential measurements might be extended to quantitative surface property characterization, opening new possibilities for density functional theory and electron holography in the coming years.

For improved accuracy and precision of mean inner potentials, a combination of tilt-controlled holography and image simulations may provide improved precision, and a combination of DFT calculations and surface-controlled holography may provide both improved precision and improved accuracy. This will make mean inner potential measurements more complex, but also more useful for answering key questions about both the properties of materials and the fundamental nature of electron holography.

Appendix A: Models for Nanowire Charging

Outline

In this appendix, the calculation of the effect on the phase images of charging in the nanowire is derived. This calculation is used in the data that appears in the text, but is not necessarily of refined scientific interest. It is based on work done by Marco Beleggia, as related by Rafal Dunin-Borkowski.

The Problem

Assume an infinite wire of radius R_b with an outer shell far away of radius R_c. The inner shell has a linear charge density of λ, and the outer shell a linear charge density of –λ. Consider a two-dimensional cross-sectional profile of this structure. Set Cartesian coordinates (x,y) whose origin is the center of the wire, and set a polar coordinate system where the distance from the center of the wire is r (and the angle θ relative to (x,y) as θ=0 along the x-axis).

The charge on the wire may be distributed as a "surface charge" (only at r=R_b) or as a "volume charge" (evenly for r<R_b).

First, the objective is to derive the electric potential $V(r)$ in the three regions r<R_b, R_b<r<R_c, and r>R_c for these two charge cases (surface and volume charge).

Then the objective is to project this two-dimensional cross-sectional electric potential profile into a one-dimensional line profile (call this along the x-axis). These are the phase shifts that are applied to determine the correction due to charging.

Derivation of Electric Potential

Several starting points are required, from the textbook [139].

Equations: Notably, consider Gauss' Law, which relates electric field $\vec{E}$ and charge enclosed by a surface integral q_{enc}, in differential form, $\nabla \cdot \vec{E} = \frac{q_{enc}}{\varepsilon_0}$ or, in surface-integral form, $\oint_S \vec{E} \cdot d\vec{A} = \frac{q_{enc}}{\varepsilon_0}$. Also, consider the relation between electric field and electric potential V, in differential form $\vec{E} = -\nabla V$ or, in line-integral form, $V = -\int \vec{E} \cdot d\vec{l}$.

Boundary conditions: On either side of a boundary, V_{out}=V_{in}, and $\hat{n} \cdot [\vec{E}_{out} - \vec{E}_{in}] = \frac{\sigma}{\varepsilon_0}$, where σ is the charge on the boundary. Therefore, the potential is continuous, and the electric field (normal to the boundary) is discontinuous due to the charge on the boundary.

At infinity: Set the electric potential at $V(\infty) = 0$.

V(r), r > R_c

The same derivation holds for both surface and volume charges.

In the trivial case of $r>R_c$, the total charge enclosed is $q_{enc}= \lambda + (-\lambda) = 0$, so the electric field is zero, and the electric potential is therefore 0, because the potential is 0 at infinity.

$$V(r) = 0 \quad \left[r > R_c \right]$$

V(r), $R_b < r < R_c$

The same derivation holds for both surface and volume charges.

To calculate the electric field, create a Gauss-Law cylinder of length L and radius r, and evaluate the field on that cylinder. There is only a radial electric field component (from the wire whose edge is at R_b to the outer cylinder at R_c). Using $\oint_S \vec{E} \cdot d\vec{A} = \frac{q_{enc}}{\varepsilon_0}$, where $q_{enc}= \lambda L$ and, because the electric field is not dependent on the area of the cylinder:

$$\iint_A E_r \, dA = \frac{\lambda L}{\varepsilon_0} \quad \Rightarrow \quad E_r \iint_A dA = \frac{\lambda L}{\varepsilon_0} \quad \Rightarrow \quad E_r = \frac{\lambda L}{\varepsilon_0 A} \quad \Rightarrow \quad E_r = \frac{\lambda L}{\varepsilon_0} \frac{1}{2\pi r L} \quad \Rightarrow \quad E_r = \frac{\lambda}{2\pi\varepsilon_0 r}$$

To calculate the electric potential,

$$V(r) = -\int \vec{E} \cdot d\vec{l} = -\int \vec{E}_r \, dr = -\int \frac{\lambda}{2\pi\varepsilon_0 r} dr = -\frac{\lambda}{2\pi\varepsilon_0} \int \frac{1}{r} dr = -\frac{\lambda}{2\pi\varepsilon_0} \ln r + C$$

Now, to determine the constant of integration C, evaluate at the boundary condition $V(R_c)=0$:

$$0 = -\frac{\lambda}{2\pi\varepsilon_0} \ln R_c + C \quad \Rightarrow \quad C = \frac{\lambda}{2\pi\varepsilon_0} \ln R_c \quad ; \quad V(r) = \frac{\lambda}{2\pi\varepsilon_0} \ln R_c - \frac{\lambda}{2\pi\varepsilon_0} \ln r$$

$$V(r) = \frac{\lambda}{2\pi\varepsilon_0} \ln \frac{R_c}{r} \quad \left[R_b < r < R_c \right]$$

V(r), $r < R_b$, Surface Charge

The electric potential $V(r<R_b)$ will be different for the surface- and volume-charge cases.

The surface-charge case is relatively trivial. Creating a Gauss-Law cylinder inside the wire shows that $q_{enc}=0$. Therefore, the electric potential is constant inside the wire, and, since the electric potential is continuous, it has the same value inside the wire as it does at the boundary:

$$V(r) = \frac{\lambda}{2\pi\varepsilon_0} \ln \frac{R_c}{R_b} \quad \left[r < R_b; \text{ Surface Charge} \right]$$

V(r), $r < R_b$, Volume Charge

The volume charge case is slightly more complicated. The charge density inside the wire, assuming constant charge distribution over the area of the cross-section, is $\frac{\lambda}{\pi R_b^2}$. Therefore, for a Gaussian cylinder drawn inside the wire, $q_{enc} = \frac{\lambda}{\pi R_b^2} \pi r^2 = \frac{\lambda r^2}{R_b^2}$. Using the same sort of cylinder as for the $R_b<r<R_c$ case (length L, radius r):

$$E_r A = \frac{q_{enc}}{\varepsilon_r \varepsilon_0} \quad \Rightarrow \quad E_r \cdot \left(2\pi r L \right) = \frac{\lambda}{\varepsilon_r \varepsilon_0} \frac{r^2}{R_b^2} L \quad \Rightarrow \quad E_r = \frac{\lambda}{\varepsilon_r \varepsilon_0} \frac{r^2}{R_b^2} \frac{L}{2\pi r L} \quad \Rightarrow \quad E_r = \frac{\lambda}{2\pi\varepsilon_r\varepsilon_0} \frac{r}{R_b^2}$$

Where ε_r is the dielectric constant of the wire (14.6 for InAs [113]). And, as above, to obtain $V(r)$:

$$V(r) = -\int E_r\, dr = -\int \frac{\lambda}{2\pi\varepsilon_r\varepsilon_0}\frac{r}{R_b^2}dr = -\frac{\lambda}{2\pi\varepsilon_r\varepsilon_0 R_b^2}\int r\, dr = -\frac{\lambda}{4\pi\varepsilon_r\varepsilon_0 R_b^2}r^2 + C = V(r)$$

Setting the boundary condition of V_{out}=V_{in} leads to $V(R_b) = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{R_b}$:

$$V(R_b) = -\frac{\lambda}{4\pi\varepsilon_r\varepsilon_0 R_b^2}R_b^2 + C = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{R_b} \quad \Rightarrow \quad C = -\frac{\lambda}{2\pi\varepsilon_0}\left(\ln\frac{R_c}{R_b} + \frac{1}{2\varepsilon_r}\right)$$

Therefore:

$$V(r) = \frac{\lambda}{2\pi\varepsilon_0}\left(\frac{1}{2}\left[\frac{1}{\varepsilon_r} - \frac{r^2}{R_b^2}\right] + \ln\frac{R_c}{R_b}\right) \quad \left[r < R_b;\ \text{Volume Charge}\right]$$

Projected Phase From Electric Potential

Now the two-dimensional electric potential profile $V(r)$ will be projected into a one-dimensional phase profile $\phi(x)$. As noted above, this will use Cartesian coordinates (x,y) whose origin is the center of the wire, and will set a polar coordinate system where the distance from the center of the wire is r (and the angle θ relative to (x,y) as θ=0 along the x-axis). Therefore, $r = \sqrt{x^2 + y^2}$.

This projection will require each electric potential to be integrated over its thickness. Therefore, $\phi(x) = C_E \cdot V \cdot t = C_E \cdot \int V(x,y)\,dy$. Where the previous section considered radius r, this section projects into the position coordinate x, which may contain some regions with different radius-regimes of potential (see above).

There are four cases. The first case is trivial - where x>R_c, the projected potential is 0. The other three cases are where R_b<x<R_c (same for surface- and volume-charge cases), where x<R_b for the surface-charge case and where x<R_b for the volume-charge case.

$\phi(x)$, $R_b < x < R_c$

The electrostatic potential above will need to be integrated over the range y=[-R_c,R_c]. From above, $V(r) = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{r} \quad \left[R_b < r < R_c\right]$. Since $r = \sqrt{x^2 + y^2}$:

$$V(x,y) = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{\sqrt{x^2+y^2}}$$

Let $a = \sqrt{R_c^2 - x^2}$, and integrate (taking advantage of the symmetry about $x = 0$):

$$\int_{-a}^{a} V(x,y)\,dy = 2\int_0^a V(x,y)\,dy = 2\int_0^a \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{\sqrt{x^2+y^2}}\,dy$$

Which integrates to

$$\phi(x) = \frac{\lambda}{\pi\varepsilon_0}\left[y\left(1 + \ln R_c\right) - x\tan^{-1}\frac{y}{x} - \frac{1}{2}y\ln\left(x^2 + y^2\right)\right]_{y=0}^{y=\sqrt{R_c^2 - x^2}}$$

While y=0 evaluates to 0 for all terms,

$$\phi(x) = \frac{\lambda}{\pi\varepsilon_0}\left[\sqrt{R_c^2 - x^2}\left(1 + \ln R_c\right) - x\tan^{-1}\frac{\sqrt{R_c^2 - x^2}}{x} - \frac{1}{2}y\ln\left(x^2 + R_c^2 - x^2\right)\right]$$

Which, after simplifying, becomes:

$$\phi(x) = \frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_c^2 - x^2} - x\tan^{-1}\frac{\sqrt{R_c^2-x^2}}{x}\right) \quad \left[R_b < x < R_c\right]$$

ϕ(x), x < R_b

Part of this process is the same for the surface- and volume-charge cases. For the $x<R_b$ case, the same mirror symmetry about the x-axis applies. However, it is perhaps easiest to decompose the integral into three components:

$$\phi(x) = 2\int_0^{\sqrt{R_c^2-x^2}} V(R_b < r < R_c)\,dy - 2\int_0^{\sqrt{R_b^2-x^2}} V(R_b < r < R_c)\,dy + 2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy$$

The two $\int V(R_b < r < R_c)\,dy$ terms are identical to what was derived above, but with different limits, yielding:

$$\begin{aligned}\phi(x) \quad = \quad & \frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_c^2 - x^2} - x\tan^{-1}\frac{\sqrt{R_c^2-x^2}}{x}\right) \\ & -\frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_b^2 - x^2} - x\tan^{-1}\frac{\sqrt{R_b^2-x^2}}{x}\right) \\ & +2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy\end{aligned}$$

The $2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy$ term is different between the two cases, and will now be evaluated. Call this term $F(x) = 2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy$

ϕ(x), x < R_b, Surface Charge

The $F(x) = 2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy$ equation is combined with the surface charge equation from above: $V(r) = \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{R_b} \quad \left[r < R_b;\ \text{Surface Charge}\right]$, yielding:

$$F(x) = 2\int_0^{\sqrt{R_b^2-x^2}} \frac{\lambda}{2\pi\varepsilon_0}\ln\frac{R_c}{R_b}\,dy = \frac{\lambda}{\pi\varepsilon_0}\cdot\left(\ln\frac{R_c}{R_b}\right)\cdot\sqrt{R_b^2 - x^2}$$

For a final result of:

$$\begin{aligned}\phi(x) \quad = \quad & \frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_c^2 - x^2} - x\tan^{-1}\frac{\sqrt{R_c^2-x^2}}{x}\right) \\ & -\frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_b^2 - x^2} - x\tan^{-1}\frac{\sqrt{R_b^2-x^2}}{x}\right) \quad ; \quad \left[x < R_b; \quad \text{Surface Charge}\right] \\ & +\frac{\lambda}{\pi\varepsilon_0}\cdot\left(\ln\frac{R_c}{R_b}\right)\cdot\sqrt{R_b^2 - x^2}\end{aligned}$$

ϕ(x), x < R_b, Volume Charge

The volume-charge case is slightly more complicated, but not impossibly so.

$$F(x) = 2\int_0^{\sqrt{R_b^2-x^2}} V(r < R_b)\,dy \ ; \ V(r) = \frac{\lambda}{2\pi\varepsilon_0}\left(\frac{1}{2}\left[\frac{1}{\varepsilon_r} - \frac{r^2}{R_b^2}\right] + \ln\frac{R_c}{R_b}\right) \quad \left[r < R_b; \text{Volume Charge}\right]$$

Substitution into the integral yields: $2\int_0^{\sqrt{R_b^2-x^2}} \frac{\lambda}{2\pi\varepsilon_0}\left(\ln\frac{R_c}{R_b}+\frac{1}{2}\left[\frac{1}{\varepsilon_r}-\frac{x^2+y^2}{R_b^2}\right]\right)dy$

The first term is identical to the surface-charge case, yielding: $\frac{\lambda}{\pi\varepsilon_0}\cdot\left(\ln\frac{R_c}{R_b}\right)\cdot\sqrt{R_b^2-x^2}$

The second term is not, but simplifies to:

$$\frac{\lambda}{2\pi\varepsilon_0}\int_0^{\sqrt{R_b^2-x^2}}\left(\left[\frac{1}{\varepsilon_r}-\frac{x^2+y^2}{R_b^2}\right]\right)dy = \frac{\lambda}{2\pi\varepsilon_0}\left[\int_0^{\sqrt{R_b^2-x^2}}\left(\frac{1}{\varepsilon_r}-\frac{x^2}{R_b^2}\right)dy-\int_0^{\sqrt{R_b^2-x^2}}\frac{y^2}{R_b^2}dy\right]$$

$$= \frac{\lambda}{2\pi\varepsilon_0}\sqrt{R_b^2-x^2}\left[\left(\frac{1}{\varepsilon_r}-\frac{x^2}{R_b^2}\right)-\frac{1}{3R_b^2}\left(R_b^2-x^2\right)\right]$$

When all of this is put together, it yields:

$$\begin{aligned}\phi(x) \quad = \quad & \frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_c^2-x^2}-x\tan^{-1}\frac{\sqrt{R_c^2-x^2}}{x}\right)\\ & -\frac{\lambda}{\pi\varepsilon_0}\left(\sqrt{R_b^2-x^2}-x\tan^{-1}\frac{\sqrt{R_b^2-x^2}}{x}\right) \qquad ; \quad \left[x<R_b; \quad \text{Volume Charge}\right]\\ & +\frac{\lambda}{\pi\varepsilon_0}\sqrt{R_b^2-x^2}\left[\left(\ln\frac{R_c}{R_b}\right)+\frac{1}{2}\left(\frac{1}{\varepsilon_r}-\frac{x^2}{R_b^2}\right)-\frac{1}{6R_b^2}\left(R_b^2-x^2\right)\right]\end{aligned}$$

Appendix B: Algebraic Tomography

Outline

In this appendix, algebraic tomography (AT) is applied to the InAs nanowire tilt series detailed in chapter 3. AT, as implemented here, does not put constraints on the density or intensity distribution of the object, but solves it generally. The resulting cross-section, however, is not interpretable.

In this appendix, three different AT algorithms are applied to the tilt series of images – the weighted back-projection (WBP), the algebraic reconstruction technique (ART), and the simultaneous iterative reconstruction technique (SIRT), as implemented in the TomoJ software package [107]. However, the tomographically determined cross-sections require processing for interpretability, but some of this processing renders the output of WBP, ART, and SIRT not quantitatively interpretable.

AT Processing Steps

The three cross-sections of interest from the key region of the nanowire were each examined. Data processing on the input data is discussed in chapter 3. Individual cross-sections were used in this instance, yielding three reconstructed images for each technique. The red-green-blue color scheme of the cross-sections used here (see *e.g.* Figure B.1) follows that of Figure 3.1. These reconstructed cross-sections were forward-projected to yield line profiles, both at 0° and at the 7° specimen tilt angle that formed the "plateau-like" center of the holography span. The reconstructed cross-sections were then thresholded to match the width of the wire in the reconstruction to the known width of the wire from the HAADF images, and forward-projectioned cross-sectional line profiles at 0° and 7° were examined. Thresholding is required because it allows for the quantification of thickness from the forward-projected line profile.

ART will be examined first, as it introduces an additional component to the analysis, followed by WBP, and finishing with SIRT. As already mentioned, all three reconstruction algorithms used the TomoJ software package for reconstruction [107]. The latter is the most successful of the three, but none of the three AT algorithms is ultimately useful for quantitative cross-sectional determination.

ART

The ART-reconstructed nanowire cross-sections can be seen in Figure B.1. The red-green-blue color scheme of the cross-sections used here follows that of Figure 3.1. 25 ART iterations were used, which provided the optimal convergence according to the program output. All three cross-sections are similar enough for direct comparison. The reconstructed cross-sections in Figure B.1 show noise inside and outside the wire. However, there is a clearly distinguishable nanowire surrounded by vacuum present in Figure B.1. The edges along the left and right sides are more clearly defined than the edges at the top and the bottom, which are blurred due to the missing-wedge of incomplete angular span in the tilt series. The hexagonal facets and rounded corners are both visible on the left and right.

The as-reconstructed cross-sections in Figure B.1 are then forward-projected, yielding the line profiles seen in Figure B.2a for 0°. The intensity level in vacuum is nearly 0, which is consistent with the original forward-projected profiles. The profiles are very

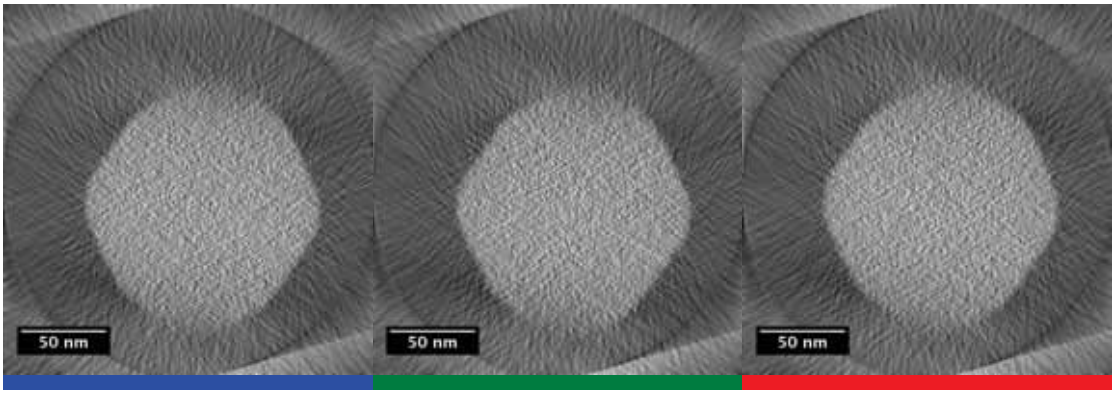


Figure B.1: ART reconstruction of nanowire, as-reconstructed. Three reconstructed slices of the region of interest, reconstructed using 25 iterations of TomoJ ART reconstruction (see text). The blue, green, and red lines correspond to the profiles seen in Figure 3.1.

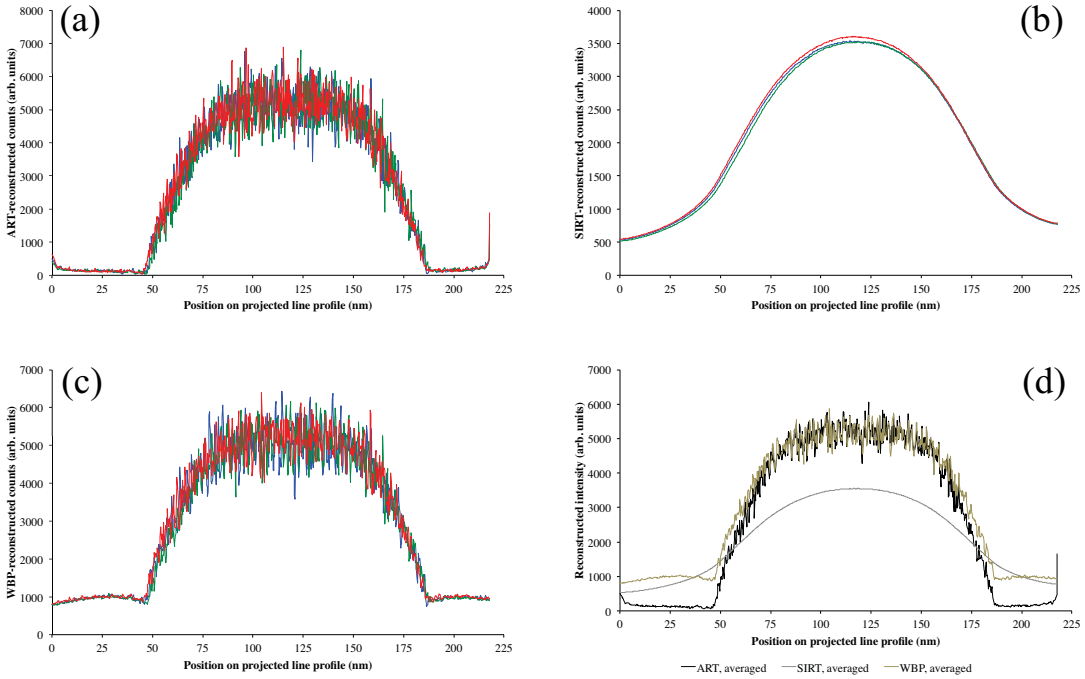


Figure B.2: Forward-projected line profiles from ART, SIRT, and WBP tomography data, as-reconstructed. (a), (b), and (c) show the results of forward-projecting the reconstructed specimen cross-sections corresponding to the blue, green, and red line profiles in Figure 3.1. (d) shows the results of averaging together each reconstruction technique's colored cross-sections. Note the vacuum levels for WBP and SIRT.

noisy, but similar; averaging over all of them yields the ART line profile in Figure B.2d. This is still quite noisy, unfortunately, but might be interpretable.

Thresholding this reconstruction yields the data seen in Figure B.3. This shows intensity inside the wire, but there is still noise both inside and outside the wire. Perhaps most problematically, outside of the circular middle region, there is significant intensity. When this thresholded reconstruction is forward-projected, the graph seen in Figure B.4a is obtained. It shows that about 30% of the vacuum region is above the threshold intensity for the nanowire, and the nanowire itself is only ~40% of the image at maximum intensity. This is quite suboptimal, and it would be difficult to remove this background while maintaining a verifiable nanowire cross-section and features.

There is a circular feature surrounding the wire, inscribed on the square of the image borders (meaning it is likely an artifact of the reconstruction technique, arising from how it determines intensity outside the images). There is intensity outside this circle that might be causing reconstruction artifacts. This circular feature does ensure that, when the wire is rotated for forward-projection at the holography orientation, the corners of the reconstructed cross-section outside the circle have no effect. Therefore, this region can be masked out, as shown in Figure B.5, yielding the line profiles seen in Figure B.6a for individual profiles, and B.6d for the three profiles averaged together. While there is no nanowire-linked intensity in this region, the removal of some intensity drops the intensity level in vacuum below zero, making it much more difficult to interpret this profile.

Since the thresholding gives large non-zero intensity in vacuum, and since, without thresholding, it creates negative intensity in vacuum when the region of the reconstruction outside of the inscribed circle is masked out, ART is not suitable in this instance for forward-projection averaging. The discussion now turns to weighted back-projection (WBP).

WBP

The reconstructed weighted back-projection (WBP) cross-sections can be seen in Figure B.7. All three cross-sections are similar. These show noise inside and outside the wire. However, there is a clearly distinguishable wire present. The edges along the left and right sides are more clearly defined than the edges at the top and the bottom, which are blurred due to the missing-wedge of incomplete angular span in the tilt series. The hexagonal facets and rounded corners are both visible on the left and right.

When the as-reconstructed cross-section is forward-projected, this yields the line profiles seen in Figure B.2c and B.2d for 0°. Unlike the original images, the intensity in vacuum is non-zero, which shows that the forward-projection is not necessarily interpretable for the raw WBP reconstruction. This is to be expected, as WBP does not iteratively optimize the reconstruction to ensure this [57].

Figure B.8 shows the thresholded WBP reconstruction, showing the same problem as for ART – the intensity in vacuum is non-zero. This non-zero intensity pattern will make a forward-projection even harder to match to the real forward-projected profile, and does not allow for good intensity-thickness calibration from the forward-projection of the tomography data. When this thresholded reconstruction is forward-projected and plotted, the intensity profile in Figure B.4c is obtained, showing the same behavior as the ART reconstruction; significant non-zero intensity in vacuum.

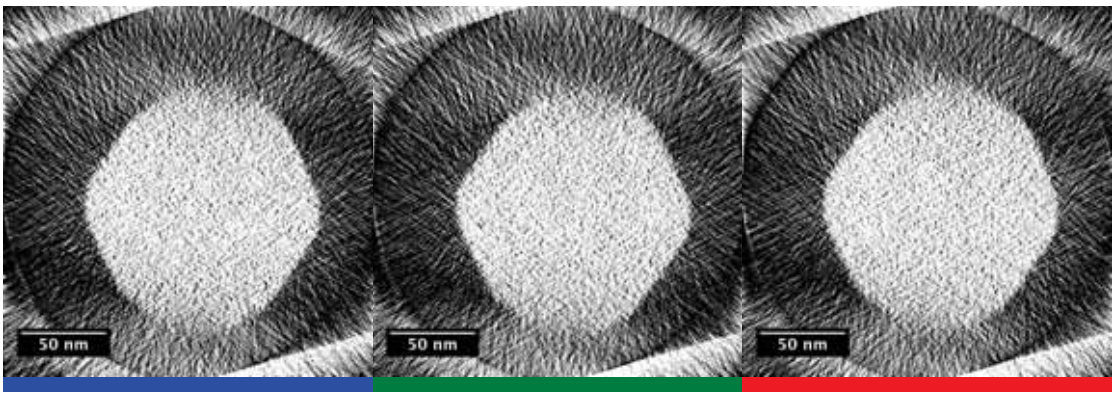


Figure B.3: ART reconstructions of nanowire from Figure B.1, thresholded. These are the cross-sections from Figure B.1, with thresholding as described in the text, to preserve the width of the wire and allow for quantitative determination of the cross-sectional thickness.

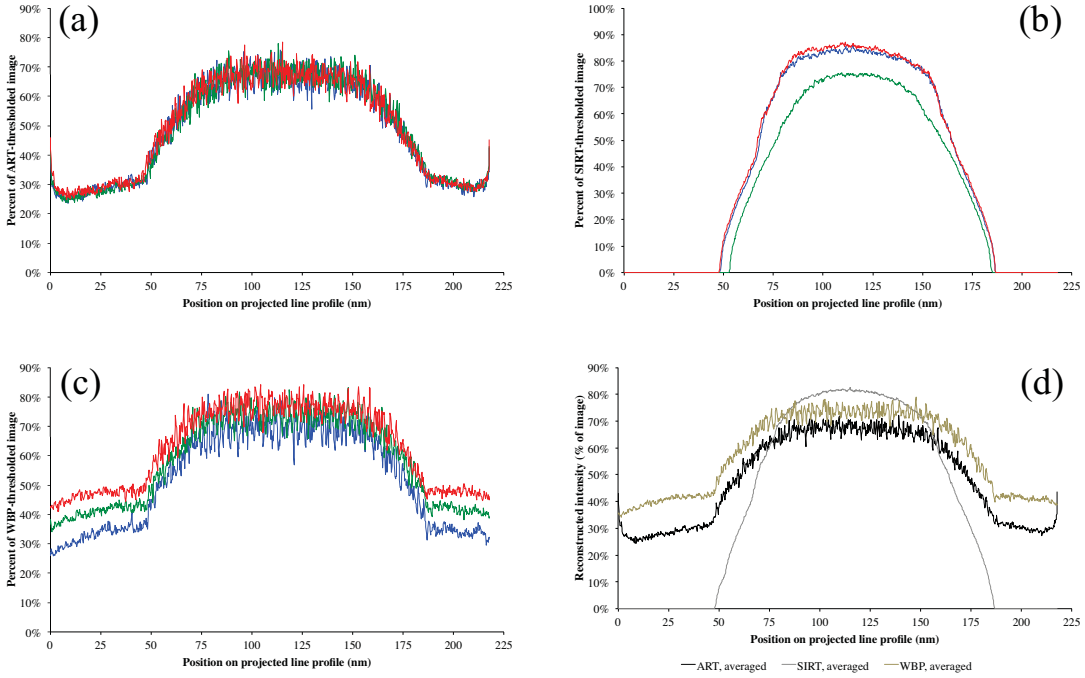


Figure B.4: Line profiles from thresholded reconstructions, unmasked and unrotated. These are the line profiles from Figure B.3, treated in the same manner as Figure B.2. See Figure B.2 for details.

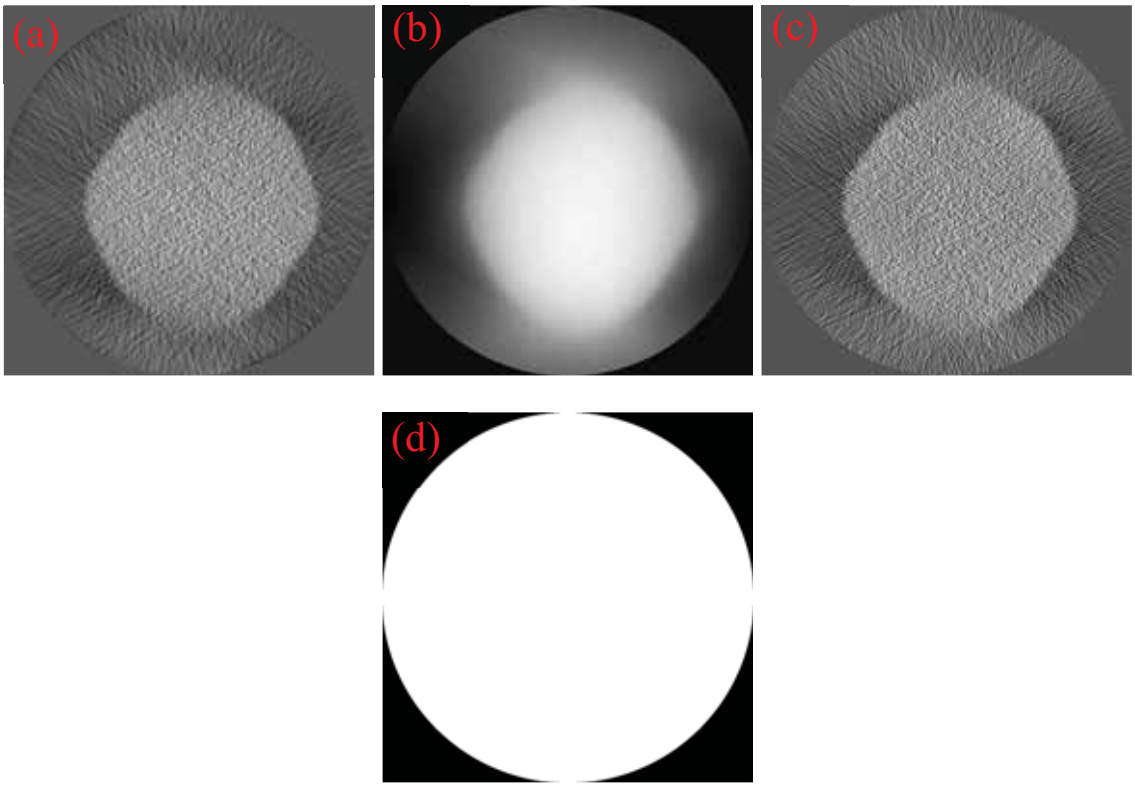


Figure B.5: Examples of masked algebraic tomography data. The same mask, seen in (d), was applied to all images. (a) is the mask applied to the red-profile ART reconstruction seen in Figure B.1, (b) is the mask applied to the red-profile SIRT reconstruction seen in Figure B.10, and (c) is the mask applied to the red-profile WBP images seen in Figure B.7.

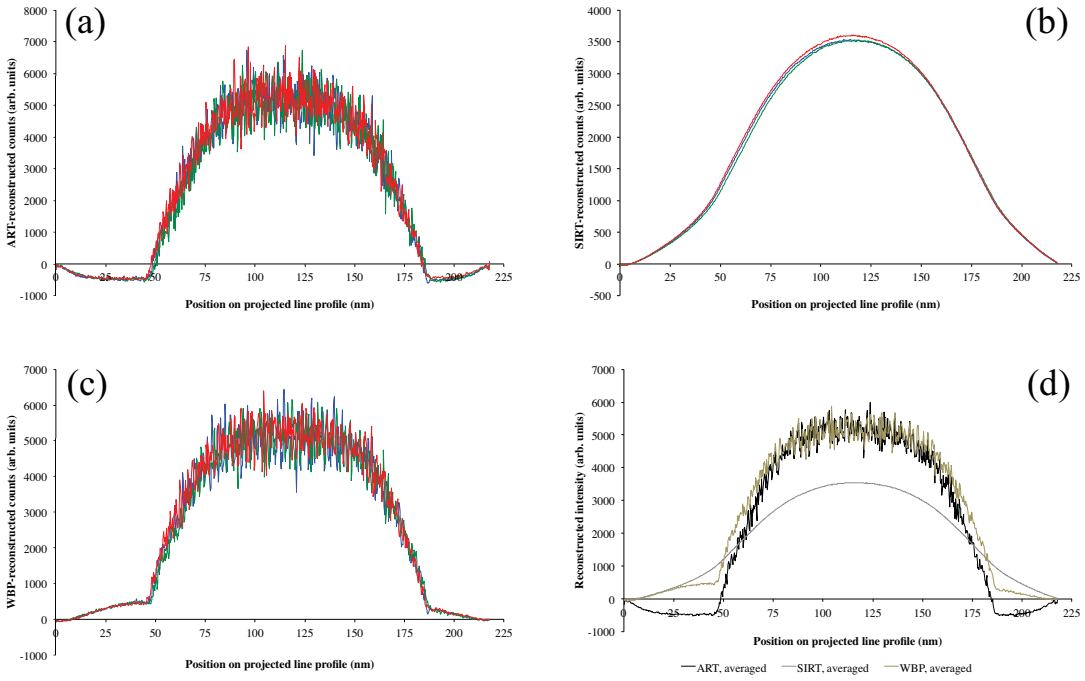


Figure B.6: Forward-projected line profiles, in the same manner as Figure B.2, created from masked tomography data (see Figure B.5). The same key applies as for Figure B.2. Note how the vacuum level for ART is <0, and the slopes in vacuum for SIRT and WBP.

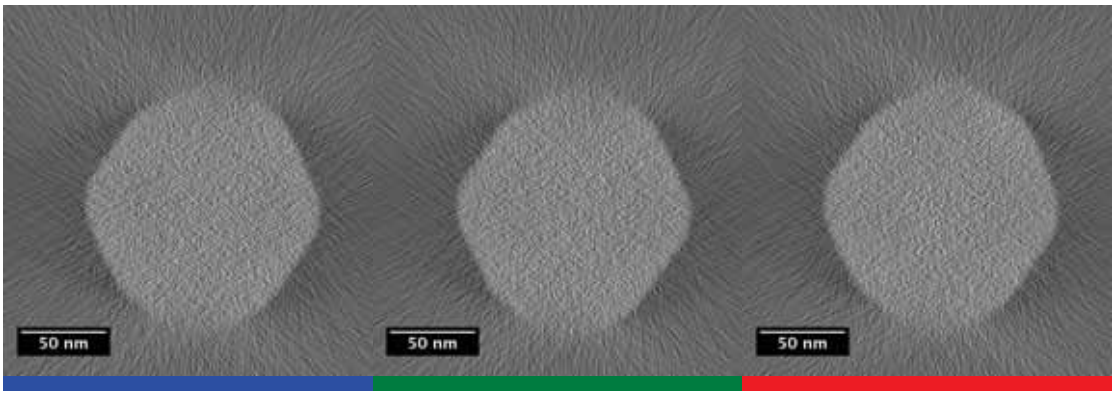


Figure B.7: Weighted back-projection reconstructed nanowire cross-sections, as-reconstructed. Created using the TomoJ software. This reconstruction is using the same coloring as the line profiles seen in Figure 3.1.

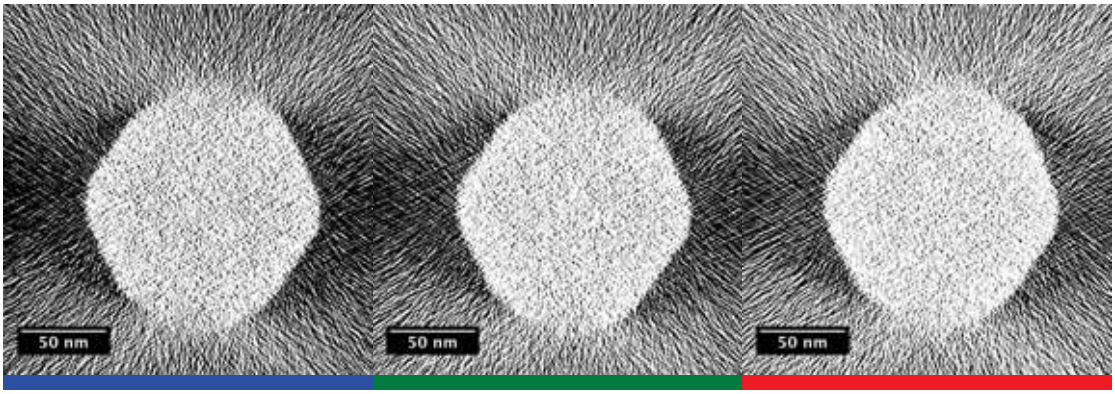


Figure B.8: Weighted back-projections from Figure B.7, thresholded in the same manner as Figure B.3. This might determine the correct width of the wire, but there is considerable intensity in vacuum.

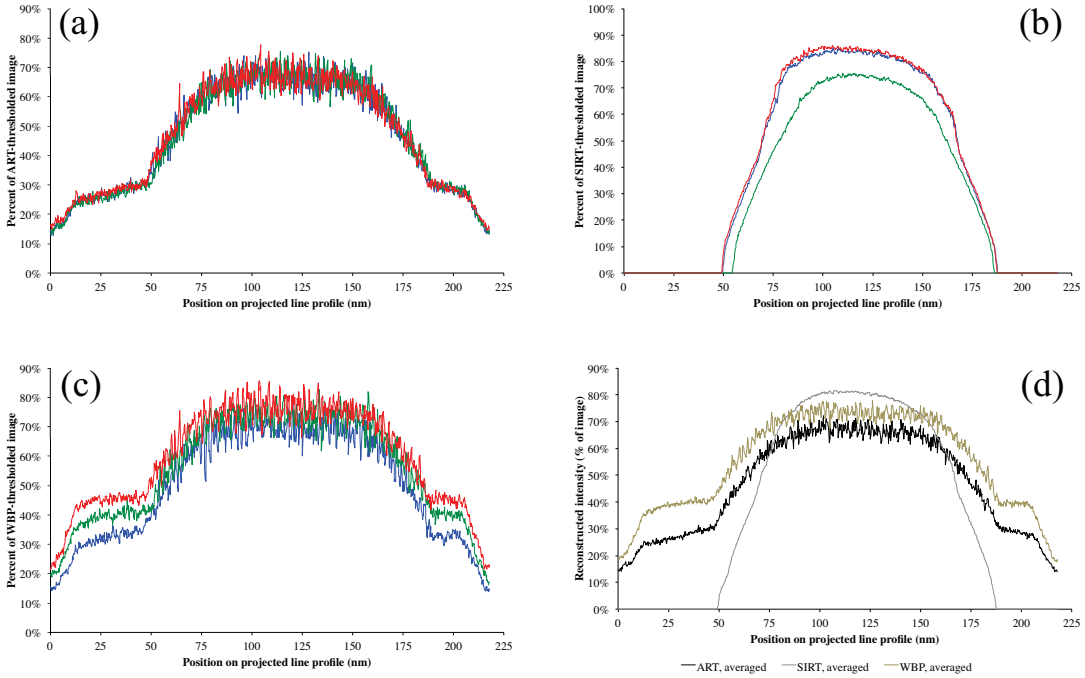


Figure B.9: Like Figure B.2, forward-projected line profiles, but from thresholded, unmasked reconstructions rotated to holography orientation.

The same circular masking that is applied to the ART reconstructions in the previous section is carried out on the WBP reconstruction. The results of the circular masking, shown in Figure B.5c for the image, Figure B.5d for the mask, and Figure B.6c for the forward-projection, show a decreasing-to-zero intensity in the vacuum region around the wire, but not a flat or an average-zero region. Therefore, as with the ART case, thresholding was performed to attempt to quantify the cross-section. However, as can be seen in Figure B.4c, the post-thresholding forward-projection of the WBP data encounters the same problems as the ART data in Figure B.4a. Even when averaged in Figure B.4d, or when rotated into the holography alignment in Figure B.9c and additionally averaged in Figure B.9d, there is a high background. Additionally, in both B.4c and B.9c, the three profiles all have different background intensity levels, unlike the ART reconstructions.

Therefore, the WBP reconstruction method does not appear appropriate for quantitative cross-sectional reconstruction. It does not yield a quantifiable cross-section due to the noise involved, although it does show the nanowire. The discussion now turns to the third AT algorithm – SIRT.

SIRT

The resulting object from the SIRT reconstruction is seen in Figure B.10 – B.10a for blue, B.10b for green, B.10c for red. 15 SIRT iterations provided optimal convergence. The first, most obvious comparison with the WBP and ART reconstructions examined in the earlier sections is how the noise level in the SIRT reconstruction appears low at first glance. The wire is clearly brighter than the surrounding low-intensity area. While the masking technique used here for the ART reconstructions might be useful, as there is a significant amount of intensity outside that mask radius, the SIRT reconstruction appears, at least at first glance, to be directly interpretable, apart from the top and bottom, where missing-wedge artifacts are also seen.

This initial appearance is deceptive, however, as Figure B.2b shows. The line profiles for these reconstructions not only have a non-zero intensity in vacuum, but the entire profile of the wire bears no clear resemblance to the experimental profile; due to the missing-wedge region of the tilt series (due to the limited angular span of the tilt series), the wire does not have a flat top, intensity extends farther out than the other reconstructions, the total intensity is lower by almost a factor of 2. This reconstruction, as-reconstructed, is clearly unsuitable. This may be a result of the TomoJ SIRT implementation, and may not be the case for other SIRT-using reconstruction programs.

Applying a circular mask to the SIRT reconstruction does not ameliorate the situation greatly, either, as Figure B.6b shows. While the intensity in vacuum drops to close to zero, there is no real improvement in the profile characteristics.

Applying a threshold, as seen in Figure B.11, does appear to make the nanowire much more interpretable, to an extent. The left and right sides of the wire have a well-defined appearance, but the top and bottom of the thresholded nanowire appear completely artifactual due to the missing wedge. The threshold used, as noted earlier, is chosen to yield the same width of the wire in the forward-projected reconstruction as measured experimentally. This forward projection is seen in Figure B.4b. While a different threshold might retrieve a different, more nominally interpretable profile of

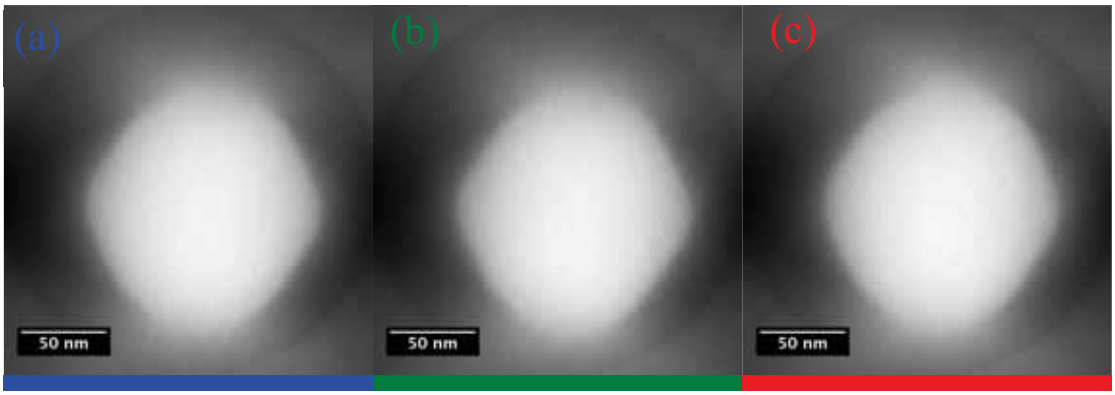


Figure B.10: SIRT reconstruction of nanowire, as-reconstructed. Reconstructed using TomoJ with 15 SIRT iterations. The SIRT reconstructions appear qualitatively to be closer to the expected cross-sectional profile than the ART (Figure B.1) or WBP (Figure B.7). There is less intensity outside the wire, outside of the missing-wedge region (top and bottom).

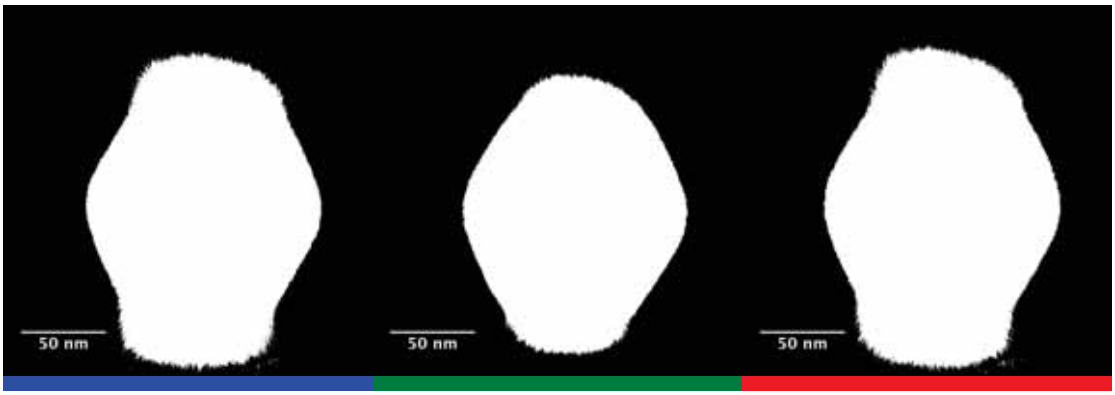


Figure B.11: SIRT reconstruction of nanowire from Figure B.10, thresholded. Compared with Figure B.10, the thresholded nanowire cross-sections are very different before and after thresholding. The threshold – chosen so that the width would be accurate – reveals that the intensity in the missing wedge region greatly affects the thresholded reconstruction.

the top and the bottom of the wire, this would yield an incorrect value for the known quantity, and would be inherently subjective. A different reconstruction program, such as FEI Inspect3D, might also yield a different SIRT reconstruction.

While the SIRT reconstruction appears better than ART or WBP reconstructions of this nanowire, it would only yield information only about a very restricted set of projected thicknesses, and must ignore the maximum projected thickness region entirely. Therefore, it is suboptimal for this experiment to use this SIRT reconstruction when other techniques may be available; for more details, see chapter 3.

Algorithmic Comparison

The output from different AT algorithms is comparable. Of the AT algorithms evaluated here, only SIRT provides a reconstruction without heavy noise, and the SIRT reconstruction itself has missing-wedge problems. While the AT reconstructions are not usable for quantitative nanowire cross-sectional determination, there are still lessons that can be drawn from AT. The nanowire appears faceted in a visual comparison of each of the pre-thresholded reconstructions, and, intuitively, the nanowire is present in the reconstruction. Qualitatively, AT appears to be a success.

Thresholding, while necessary, renders this success moot, however. The threshold is determined by taking the width of the nanowire at the appropriate tilt angle and thresholding to that intensity level. This is chosen because it creates a nanowire thickness profile that is verifiably true. If the nanowire width is incorrect, the cross-sectional profile itself cannot be correct, and would be unsuitable for use in chapter 4 for comparison with the holographic cross-sectional line profiles. Thresholding each of the AT reconstructions necessitates the inclusion of reconstruction artifacts that are clearly visible as intensity entering vacuum. The WBP and ART vacuum intensity make it impossible to accurately determine the scaling for the forward projection using the WBP and ART reconstructions, rendering them useless. The SIRT reconstruction appears to avoid vacuum intensity in the pre-thresholded image, but this is reversed when the threshold for the width is applied, leading to a center region that is clearly not interpretable. A different threshold may determine a different missing-wedge region of the wire, but would then fail to replicate the known nanowire width.

The missing wedge region provides part of the explanation, but not all. Certainly, it plays a role in the SIRT reconstruction's artifacts, especially post-thresholding. Increasing the angular range would reduce the missing-wedge artifacts, ultimately eliminating them for 180-degree tomography. This comes at the cost of potentially more damaging specimen preparation. Another program could be used that might have a different SIRT algorithm implementation. However, the ultimate problem is the exact nature of the reconstruction generated by AT.

For AT reconstructions, as discussed in [57] and in chapter 2, the specimen is a generalized density distribution, represented in the measured data as a greyscale image. The AT algorithm solves the inverse problem to reconstruct that specimen. This is its Achilles heel in this work – AT will reconstruct a specimen that is a generalized intensity profile, while the nanowire in this work is binary.

Why Not Algebraic Tomography?

As discussed in chapter 2, algebraic tomography is of general utility for reconstructing TEM specimens, but this generality comes at a price in this appendix and in Chapter 3. The objective of an AT algorithm is to reconstruct the arbitrary density distribution that leads to the measured intensity. It assumes, as prior knowledge, that the object is representable as an arbitrary density distribution. This is acceptable, and, indeed, even welcome, for a typical application such as dense catalyst particles on a sparse, convoluted support material [6]. An AT algorithm presents, as output, a greyscale picture of the density in the specimen.

This is because there are only three possible (and only likely two) densities in the nanowire specimen examined in chapter 3: InAs nanowire core, vacuum, and a thin oxide shell that does not appear in the HAADF images and likely is not captured in the reconstruction. It is sensible to recognize the nanowire as a binary object, and to determine the cross-section of this binary object. Therefore, to interpret the greyscale AT reconstruction as a cross-sectional thickness profile, the reconstructions for all three AT techniques must be thresholded. This fundamental compromise leads to the desire for a different technique that does include this binary information about the specimen, and, therefore, to either geometric tomography (GT) or discrete tomography (DT), both of which are treated in chapter 3.

Appendix C: HAADF Imaging Nonlinearity

The Measurement and Effect of Nonlinearity in HAADF Imaging

The phase and amplitude-derived t/λ discussed in chapter 4 are both linearly dependent on thickness, as derived in chapter 3. Therefore, if the holography data in chapter 4 appears nonlinearly dependent on thickness when turned into a wedge, then the source of the nonlinearity must be examined. In chapter 4, this is well compensated for by examining the shell around the nanowire. In this appendix, an alternative is considered: what if the nonlinearity in the thickness is due to the HAADF imaging process becoming nonlinearly dependent on thickness, as discussed in chapter 2? To investigate this possible effect, the 20-pixel-wide spatially averaged amplitude-derived t/λ line profile from a weakly diffracting condition discussed in chapter 4 was examined.

The amplitude and amplitude-derived t/λ line profiles across the nanowire can be seen in Figure C.1, and the wedge created with this line profile and the discrete tomography thickness data in Figure 4.11. The linear fit in Figure 4.11 initially appears qualitatively to be passably linear. There are some low-thickness oscillations, but these are also present in the hologram, and likely due to a defocus effect. However, a closer look reveals that, after approximately 90 nm in thickness, the linear behavior appears to change to sublinear behavior. This is shown in the right graph in Figure 4.11, where rolling linear fits are graphed, with 15, 20, and 30nm windows for the thinnest, medium-thickness, and thickest lines, respectively, meaning that at *e.g.* 70 nm, the thick line is fitting to the points from 40-70 nm, the medium line is fitting to the points from 50-70 nm, and the thinnest to the points from 55-70 nm. These rolling linear fits show that using a smaller window leads to more noise, and that lower thicknesses and smaller windows pick up those likely-defocus oscillations visible on the original images, both as expected. Critically, above approximately 100 nm, something changes: possibly a dramatic change takes place in the mean free path, the thickness data is sub-linear, or the shell effects are uncompensated. The third of these is considered in Chapter 4, and the second of these is considered in the following section.

Nonlinear Intensity-Thickness Response

One explanation is that the thickness data itself, derived from the HAADF experimental data, has a sublinear response at higher thicknesses. This would be plausible because intensity-thickness sublinearity has been observed for similar materials (*e.g.* InP) at thicknesses above approximately 150nm [55]. However, it is mildly surprising that this nonlinearity would be present at this relatively low thickness.

An informed-heuristic method is applied to model this sublinearity. From the extant experimental data, and the thickness curves seen in [55], the best fit that was found was to model the sublinear thickness using a threshold thickness value L for linear versus sublinear behavior, and modeling the sublinear behavior using a negative-exponential deviation from that threshold value. In other words, the real thickness t_r in terms of the measured thickness T_m:

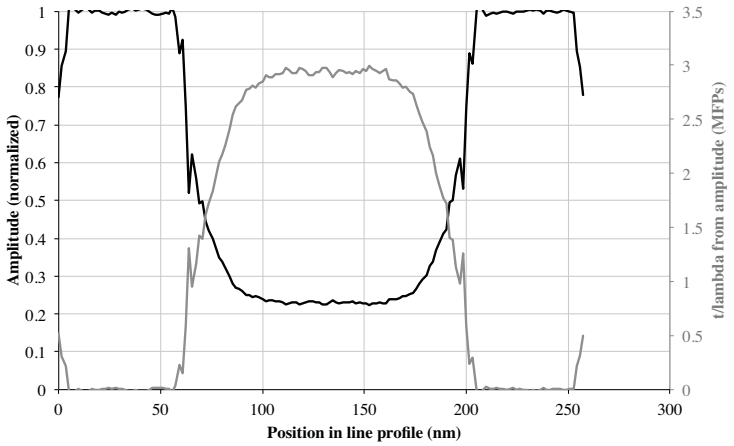


Figure C.1: Averaged 20-pixel amplitude and amplitude-derived t/λ line profile from [2.1,3]°. By using this weakly-diffracting condition and a 20-pixel wide window, the noise on the amplitude data seen in Chapter 4 can be reduced.

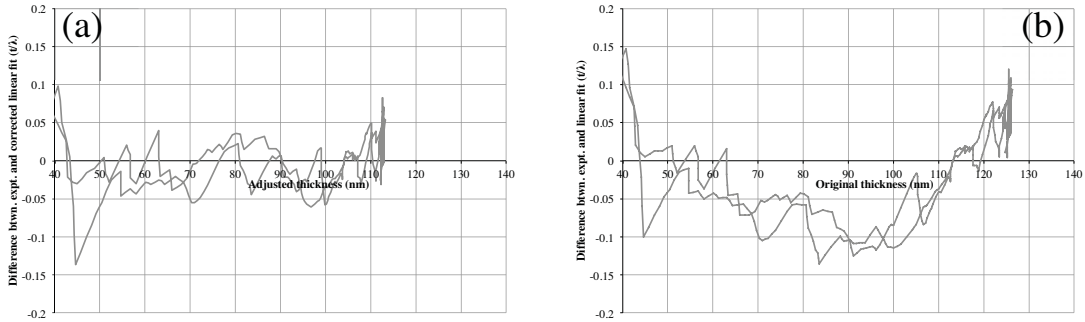


Figure C.2: Linearity of experimental amplitude-derived t/λ and linearity-adjusted thickness data, compared to original thickness data. (a) shows the difference between the experimental amplitude-derived t/λ data from the 20-pixel line profile seen in Figure C.1 and a linear fit to that t/λ data, as a function of the adjusted thickness. The thickness is adjusted to minimize the y-value of the function shown in (a). (b) is as (a), but using the original thickness data.

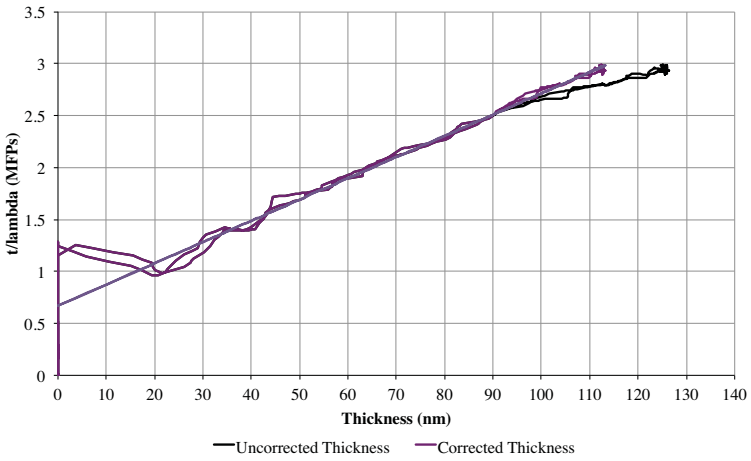


Figure C.3: Thickness profile with and without HAADF linearity correction. This shows the difference between the adjusted and unadjusted thicknesses.

$$T_m \le L: \qquad t_r(T_m) = T_m$$
$$T_m > L: \quad t_r(T_m) = T_m - (T_m - L)\left(A \cdot e^{-b\cdot(T_m - L)}\right)$$

This was done by adjusting the three parameters until they yielded the most linear fit, judged from balancing the deviations between the t/λ data and its linear fit above and below the linear fit. This was found at L = 90 nm, A = 0.45, and b = 0.006. The difference between this linear fit and the corrected experimental thickness data can be seen in Figure C.2a (and the difference between a linear fit and the original data is in Figure C.2b). For comparison, the difference between the original thickness function and the fitted version to it is in Figure C.3.

The results of this are visible in Figure C.4; C.4a and C.4b are the counterparts to 4.11 left and 4.11 right; the left graph in Figure 4.11 and Figure C.4a are plotted together in Figure C.5.

Consequences of Nonlinearity

Modeling the intensity as sublinear with thickness is not entirely satisfactory, however, because it consequently yields a nanowire thickness that is too low. If the nanowire were perfectly hexagonal, without truncated-facet corners, a width of 133.6 nm along the major axis would lead to a maximum thickness of 115.7 nm; the ~7 degree tilt of the nanowire out of the plane of the image leads to an additional multiplicative factor of 1/cos(7°), yielding a maximum thickness in this orientation of 116.5 nm. 116.5 nm would be the minimum "maximum thickness" allowed by a hexagonal cross-section. If the corners are truncated facets instead of points, then the maximum thickness would increase. The discrete tomography data yields a total maximum thickness of 125 nm, and appears to have truncated faceted corners. In contrast, the corrected thickness data in this scenario yields a maximum thickness of 112 nm, which is 3.8% below the nominal minimum of 116.5 nm.

Therefore, there is an intuitive problem with this approach. This could be corrected by simply changing the exponential function to one that started from 0 thickness, or by rescaling the entire thickness, but these methods are incompatible with the assumption, seen in Figure 4.11, that the HAADF intensity is at least linear at low thicknesses, before the onset of sublinear behavior, and the results of this do not yield a linear fit that is as good as the one detailed above. Rescaling the entire profile, without a nonlinearity onset thickness, also would raise another critical question – specifically, how to handle quantitatively the heuristic nature of the fitting procedure. If rescaling is done across the entire profile, then no part of the original thickness data is trustable, probably leading to interpretability problems.

Therefore, a different solution was required, returning to chapter 4 and the discussion of direct surface-layer compensation, which is shown in Chapter 4 (Figure 4.12) to be sufficient to compensate for the nonlinearity seen in Figure 4.11.

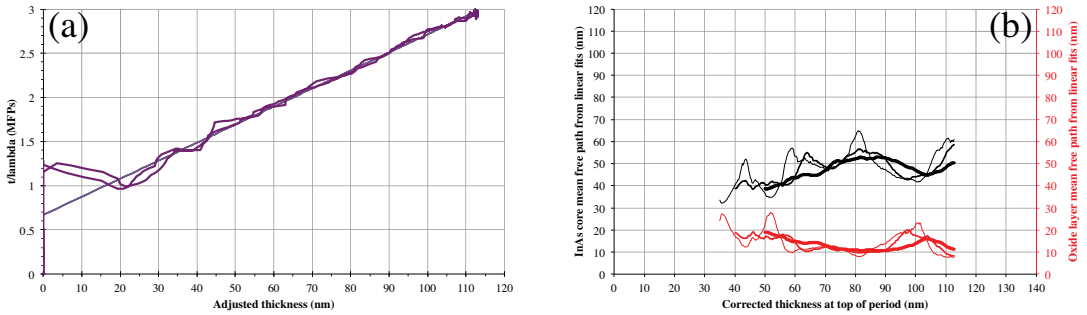


Figure C.4: Measure of linearity of adjusted thickness data. As Figure 4.11, but with the adjusted thickness data. Left: Intensity-thickness linearity is better than Figure 4.11, but still includes a non-zero intercept. Right: Mean free paths calculated from the slope (black) and intercept (red) of linear fits over 15 (thin), 20 (medium) and 30 (thick) nm thickness windows from left graph, showing the improved linearity as compared with Figure 4.11. This should be invariant with thickness, neglecting diffraction effects due to the weakly diffracting condition.

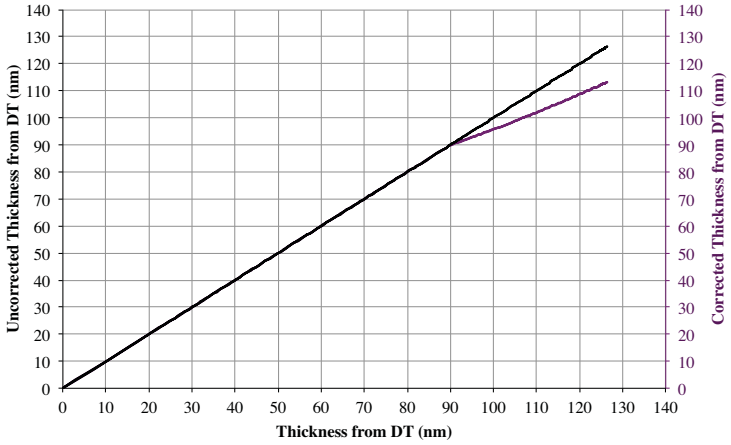


Figure C.5: HAADF-linearity-adjusted DT thickness as a function of original DT thickness. The black line shows the original thickness from DT as a function of itself (1:1 slope), while the purple line shows the HAADF-linearity-adjusted thickness from DT as a function of the original thickness from DT for the purple graphs in Appendix C. The correction created a new maximum thickness of 112 nm, which is too thin for a hexagonal cross-section model.

Appendix D: DFT Precisions for Si, Ge, GaAs

Outline

In this appendix are the DFT precision determination graphs for Si, Ge, and GaAs. These data – Figures D.1 to D.24 - are included in this dissertation because their results are used for quantitative analysis in chapter 5, but they are in an appendix because their details are sufficiently replicated by the graphs included in the main text.

The only difference between the data in this appendix and the Figures in chapter 5 is which material is used. The precision determinations are carried out identically to the diamond C precision determination outlined in Chapter 5, with two slight differences. A few data points that are missing in Figures D.1-D.24 but are present in Figures 5.4-5.11 are missing due to the simulations in question failing. Grid spacing tests are over different ranges, due to the different lattice parameters. These do not affect the overall quality of these figures.

The DFT mean inner potential precision-determination suite outlined in Chapter 5, as depicted in *e.g.* Figure 5.4 for diamond C – testing number of k-points, real-space grid sampling, the thickness of material, and the width of free space surrounding the material – is replicated here for Si, Ge, and GaAs.

Eight Figures are shown in this appendix for each material, with each figure in this appendix having a main-text counterpart. These are the figure numbers as given in the following table – where each column is under the same conditions, and each row using the same element. The main-text figures in chapter 5 are bolded.

Figure Relations Between Chapter 5 and Appendix D
(Figures 5.4-5.11 & Figures D.1-D.24)

Mode:	FD calculation mode				LCAO calculation mode			
Method:	Subtraction		Integration		Subtraction		Integration	
$a_?$	a_{REF}	a_{FD}	a_{REF}	a_{FD}	a_{REF}	a_{FD}	a_{REF}	a_{FD}
C	**5.4**	**5.5**	**5.6**	**5.7**	**5.8**	**5.9**	**5.10**	**5.11**
Si	D.1	D.2	D.3	D.4	D.13	D.14	D.15	D.16
Ge	D.5	D.6	D.7	D.8	D.17	D.18	D.19	D.20
GaAs	D.9	D.10	D.11	D.12	D.21	D.22	D.23	D.24

The 24 Figures presented in this Appendix are processed into the numerical values shown in Figure 5.12, Figure 5.13, Figure 5.14, and Figure 5.15 in the main text, and discussed further there.

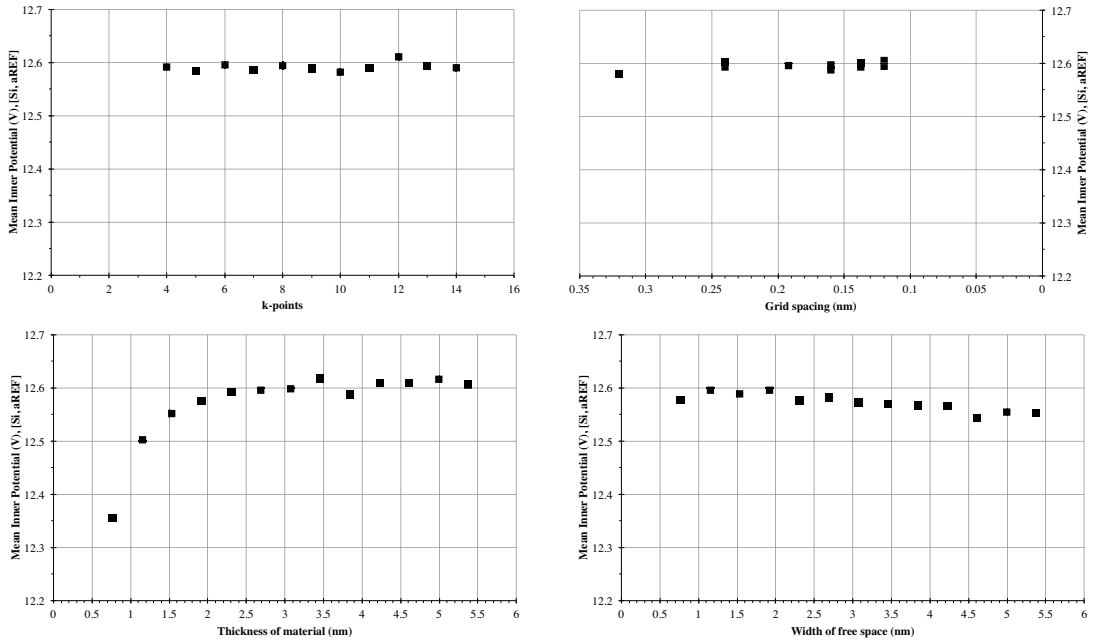


Figure D.1: DFT mean inner potential precision determination for Si, after Figure 5.4 for C. Like Figure 5.4, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

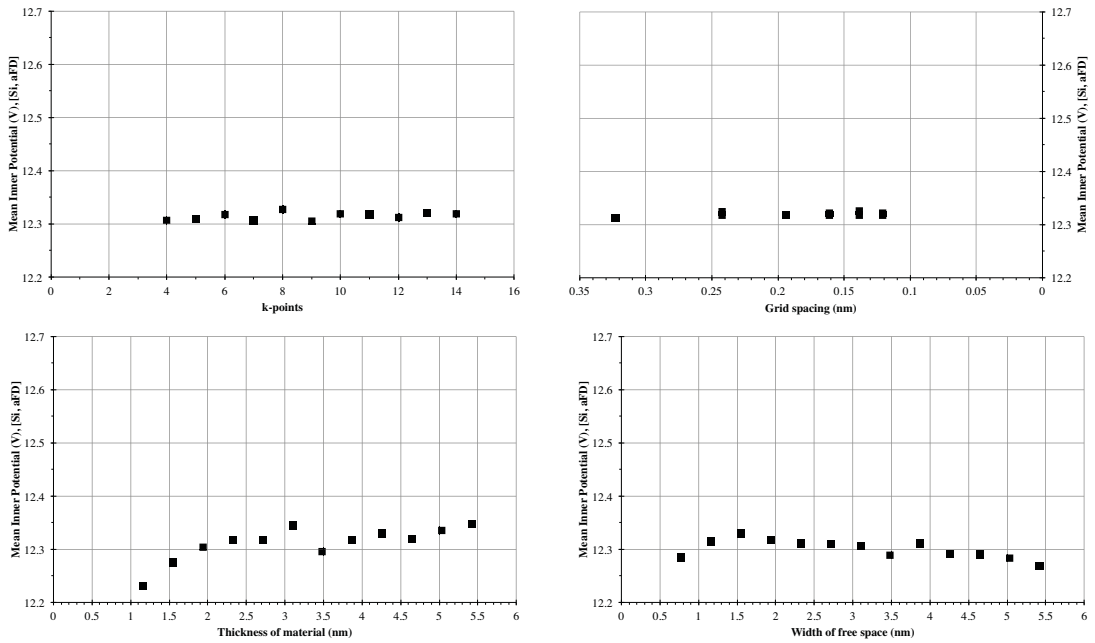


Figure D.2: DFT mean inner potential precision determination for Si, after Figure 5.5 for C. Like Figure 5.5, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

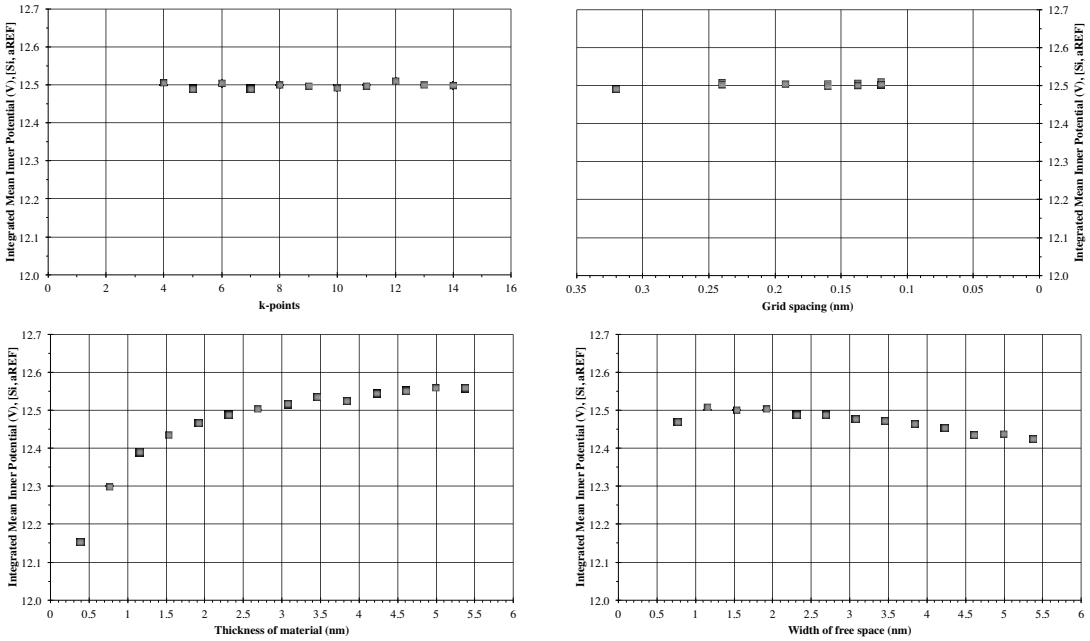


Figure D.3: DFT mean inner potential precision determination for Si, after Figure 5.6 for C. Like Figure 5.6, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

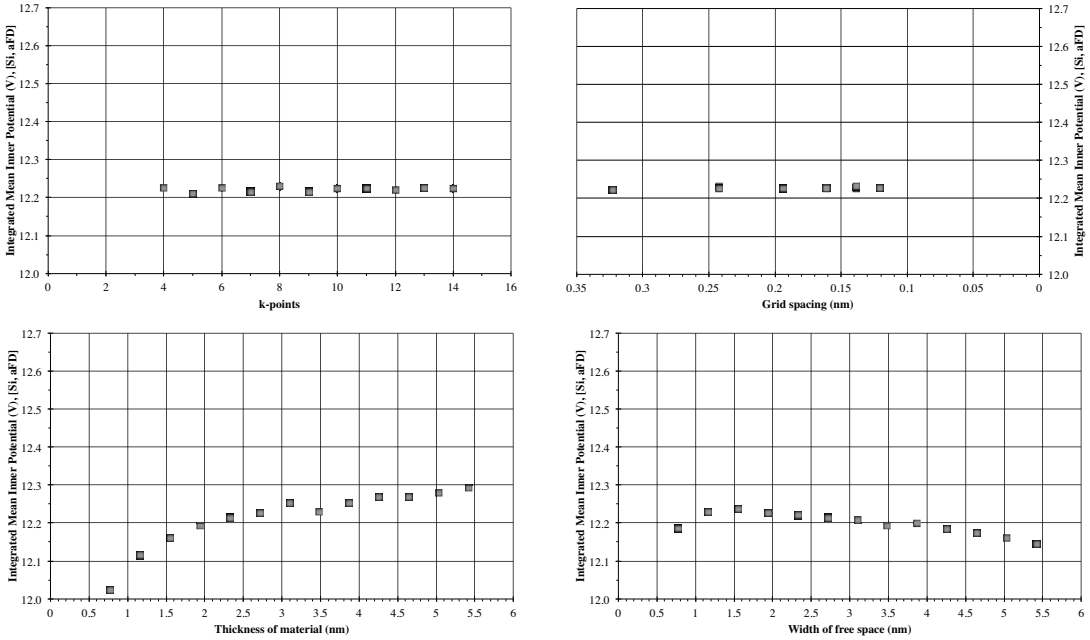


Figure D.4: DFT mean inner potential precision determination for Si, after Figure 5.7 for C. Like Figure 5.7, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

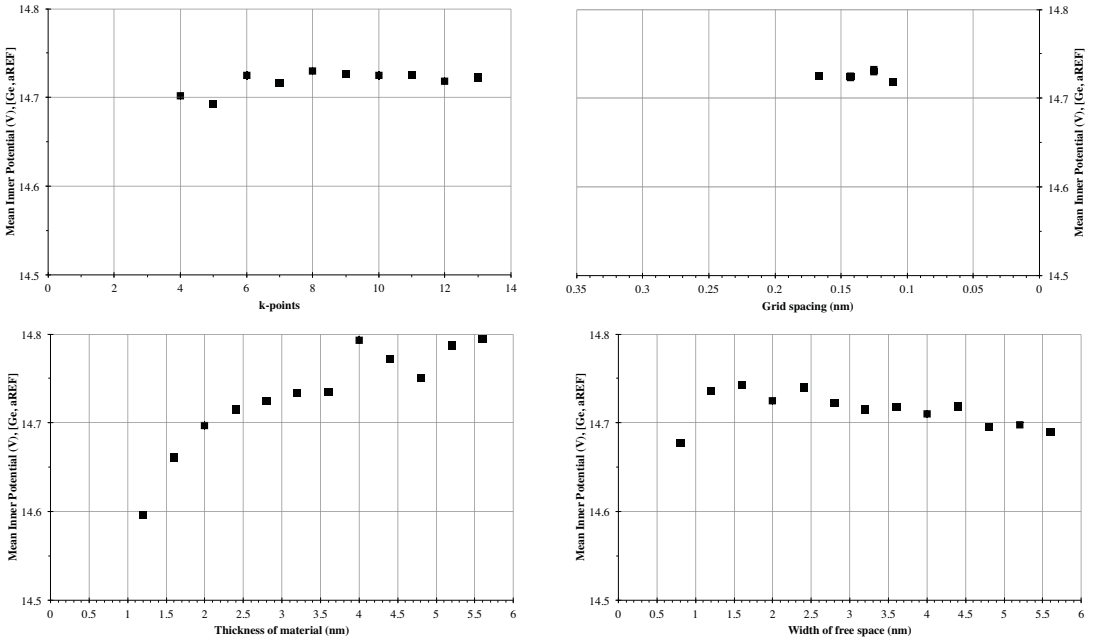


Figure D.5: DFT mean inner potential precision determination for Ge, after Figure 5.4 for C. Like Figure 5.4, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

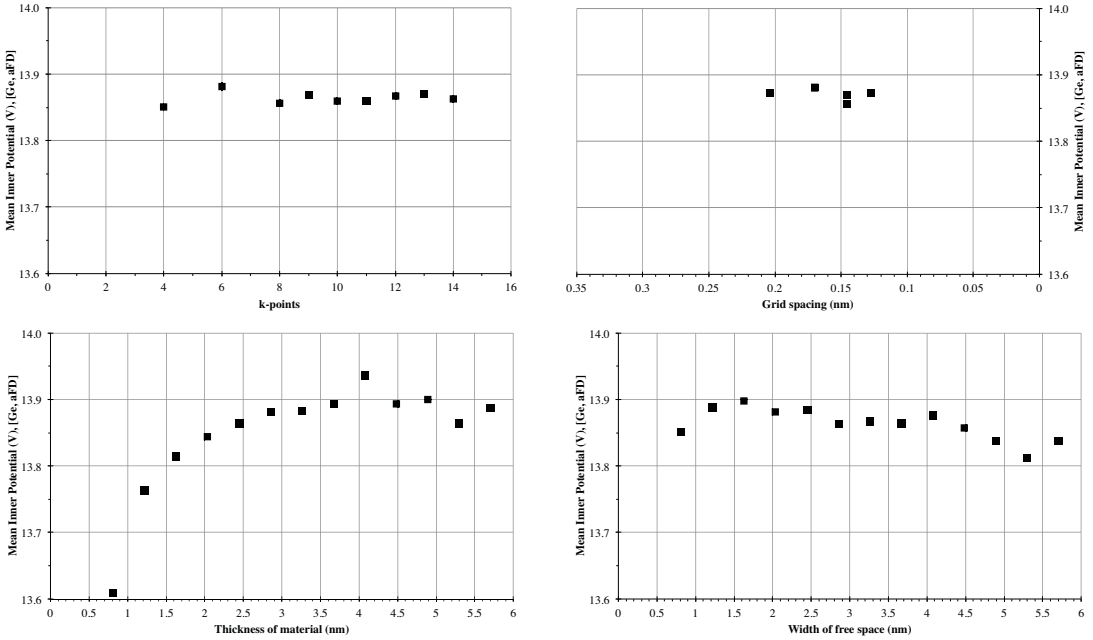


Figure D.6: DFT mean inner potential precision determination for Ge, after Figure 5.5 for C. Like Figure 5.5, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

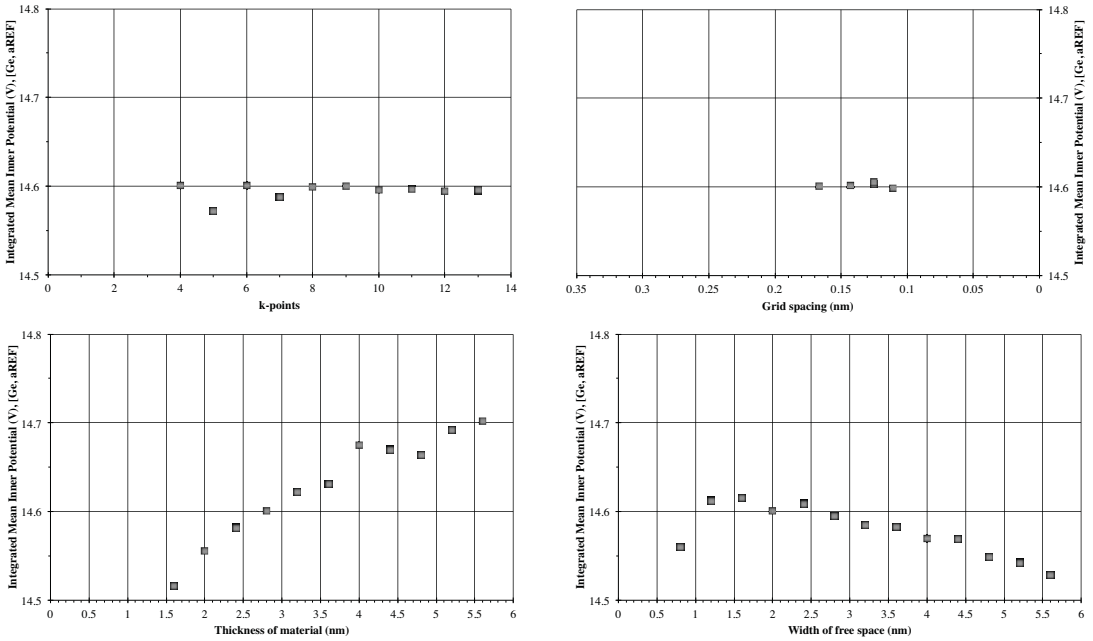


Figure D.7: DFT mean inner potential precision determination for Ge, after Figure 5.6 for C. Like Figure 5.6, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

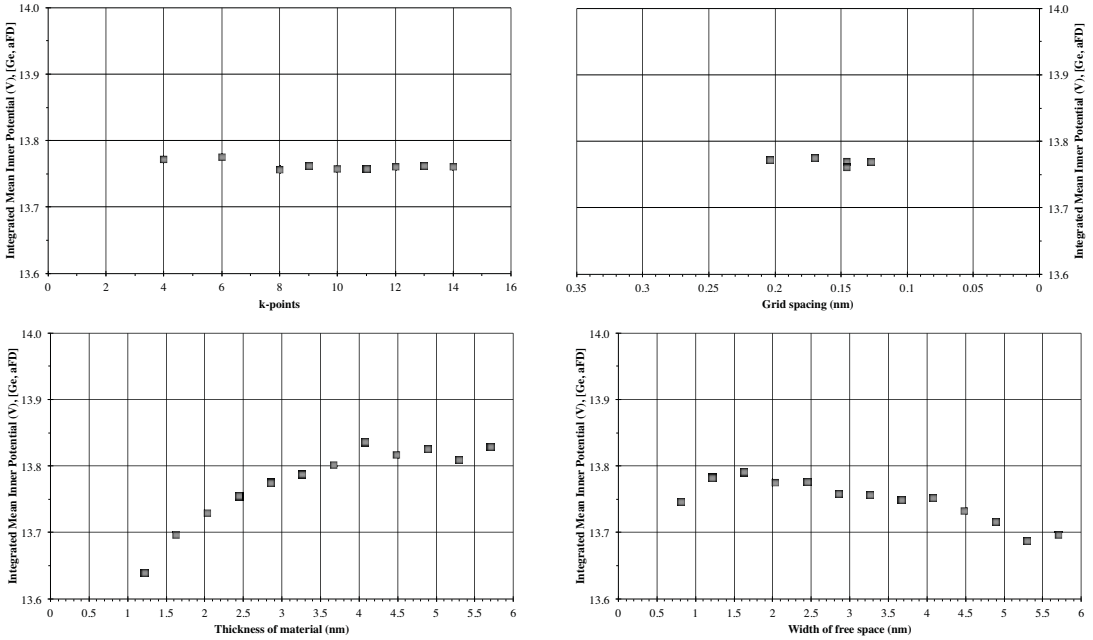


Figure D.8: DFT mean inner potential precision determination for Ge, after Figure 5.7 for C. Like Figure 5.7, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

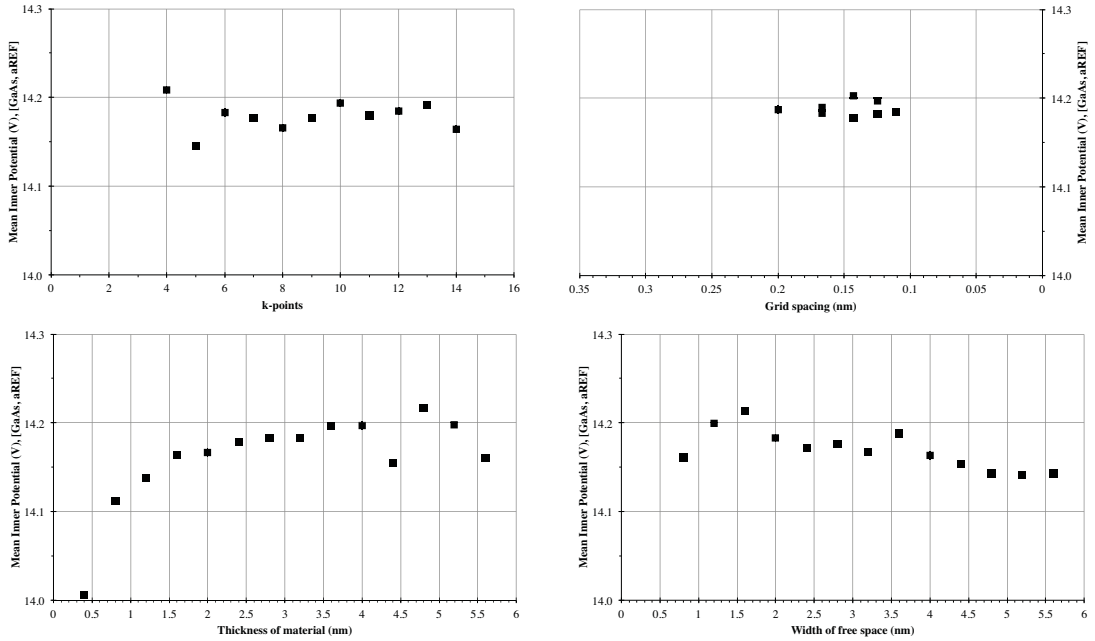


Figure D.9: DFT mean inner potential precision determination for GaAs, after Figure 5.4 for C. Like Figure 5.4, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

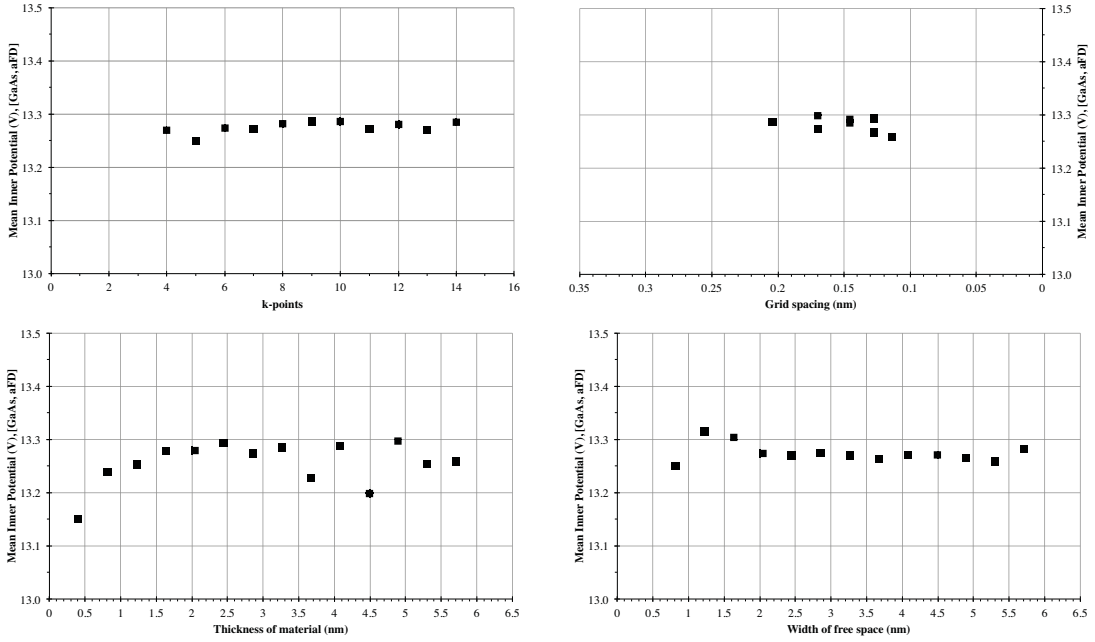


Figure D.10: DFT mean inner potential precision determination for GaAs, after Figure 5.5 for C. Like Figure 5.5, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and FD mode.

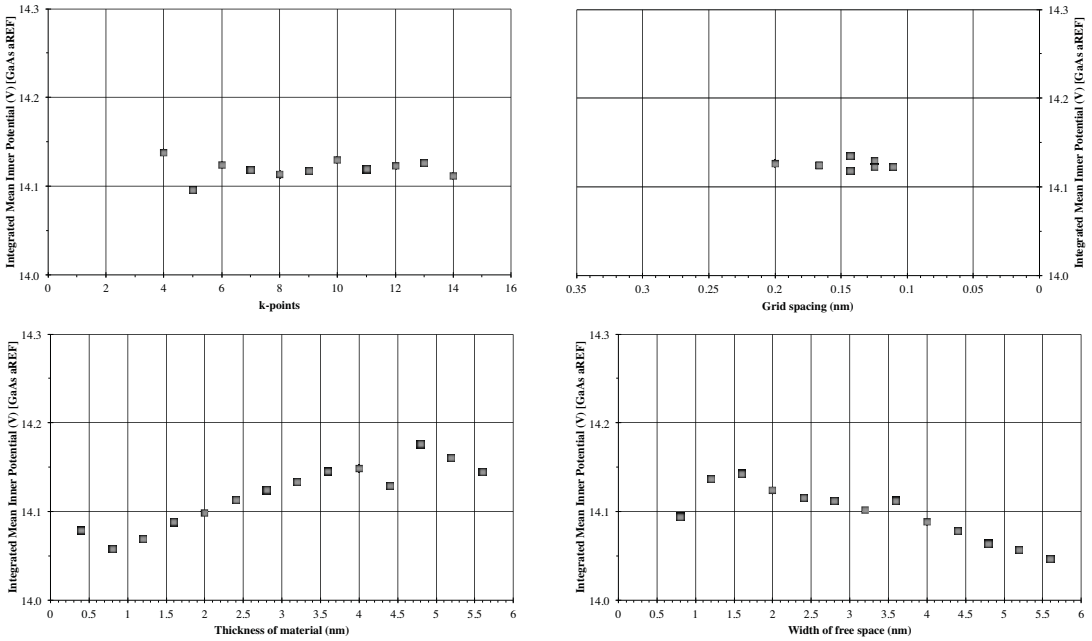


Figure D.11: DFT mean inner potential precision determination for GaAs, after Figure 5.6 for C. Like Figure 5.6, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

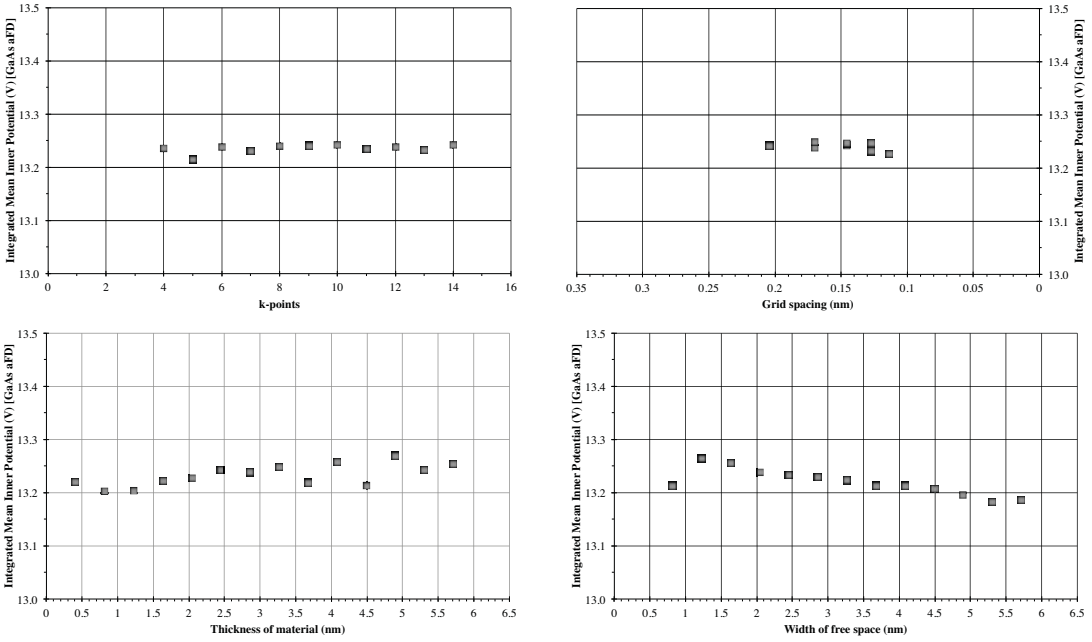


Figure D.12: DFT mean inner potential precision determination for GaAs, after Figure 5.7 for C. Like Figure 5.7, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and FD mode.

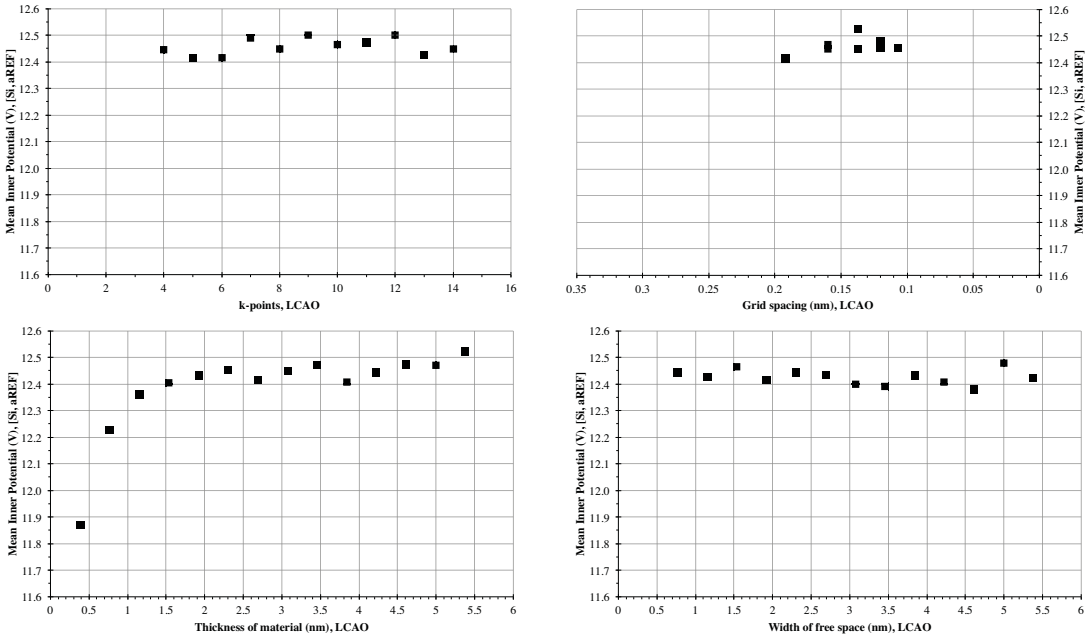


Figure D.13: DFT mean inner potential precision determination for Si, after Figure 5.8 for C. Like Figure 5.8, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

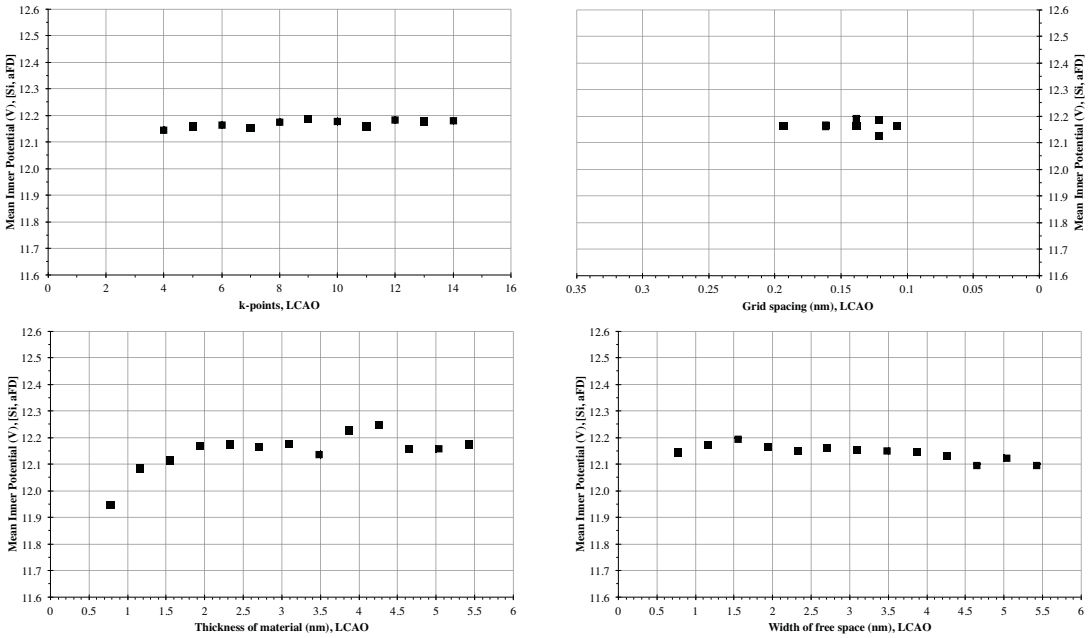


Figure D.14: DFT mean inner potential precision determination for Si, after Figure 5.9 for C. Like Figure 5.9, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

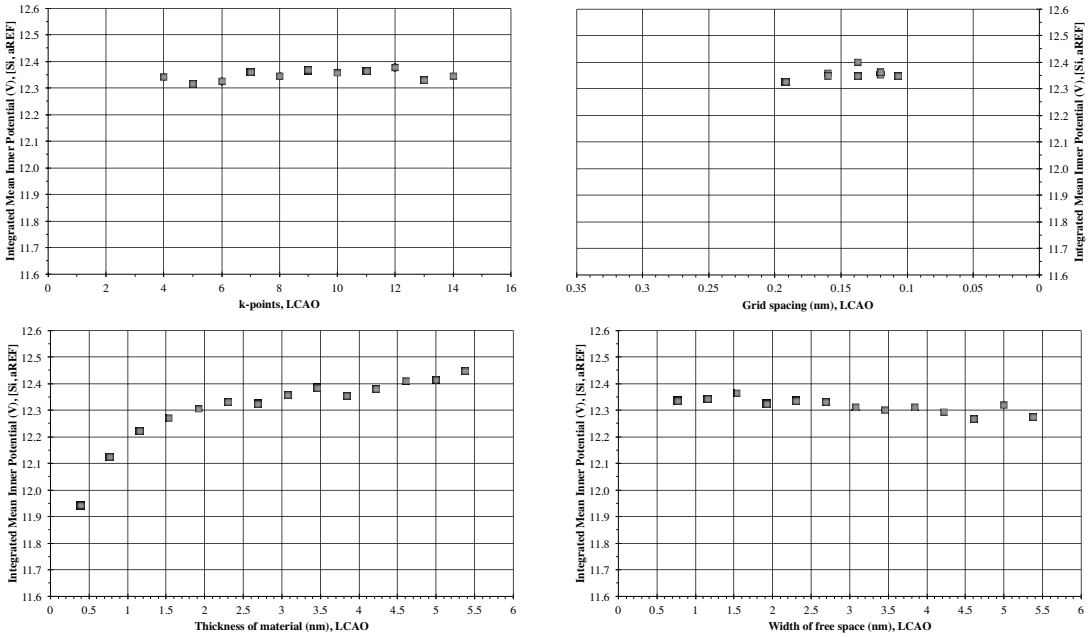


Figure D.15: DFT mean inner potential precision determination for Si, after Figure 5.10 for C. Like Figure 5.10, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

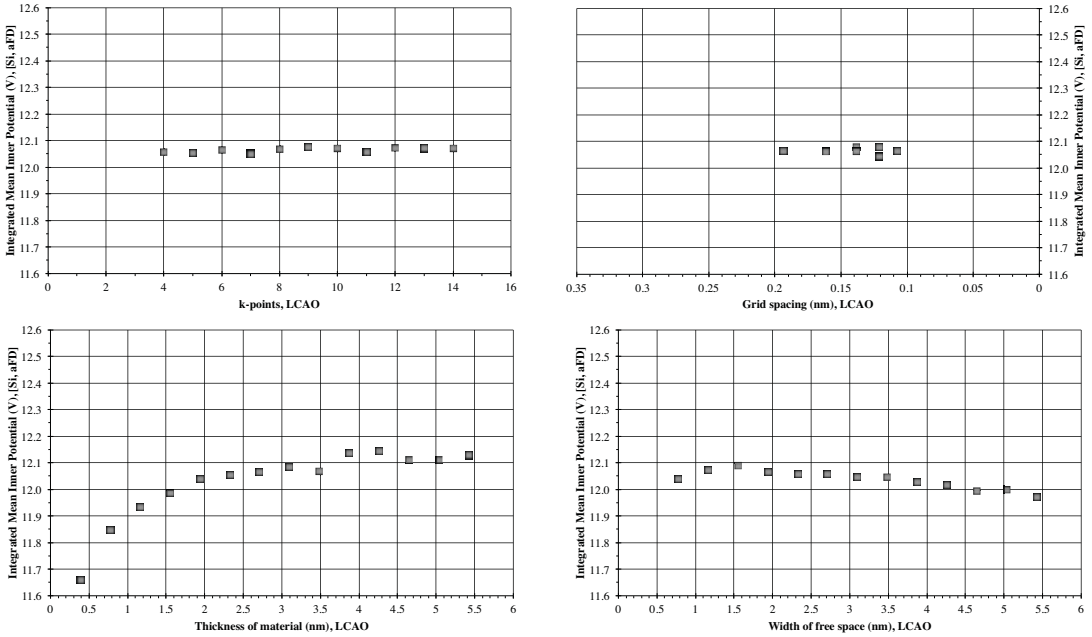


Figure D.16: DFT mean inner potential precision determination for Si, after Figure 5.11 for C. Like Figure 5.11, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

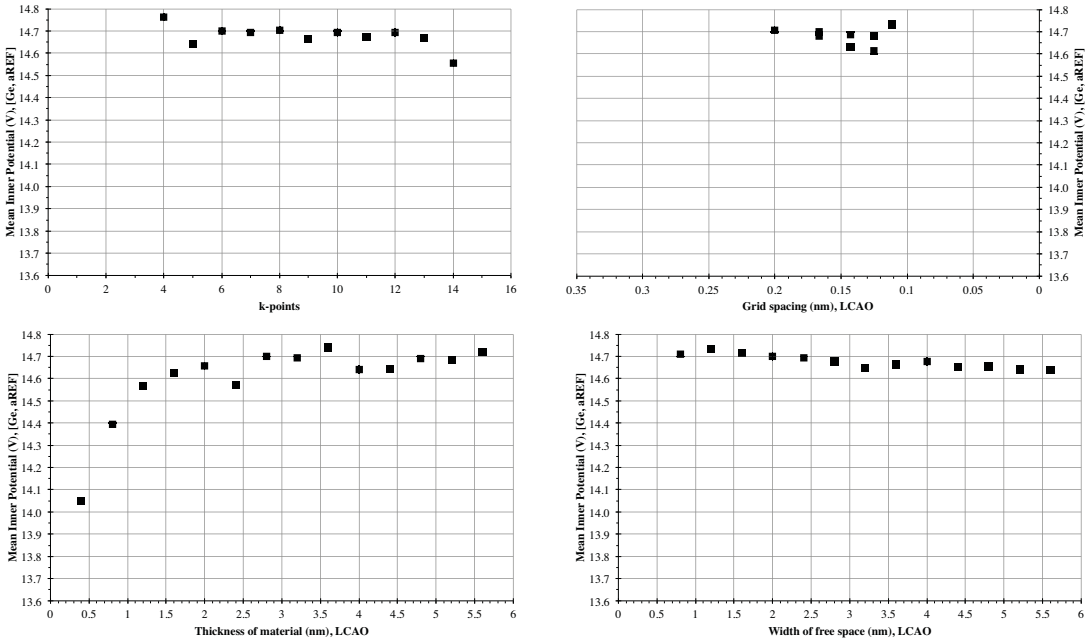


Figure D.17: DFT mean inner potential precision determination for Ge, after Figure 5.8 for C. Like Figure 5.8, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

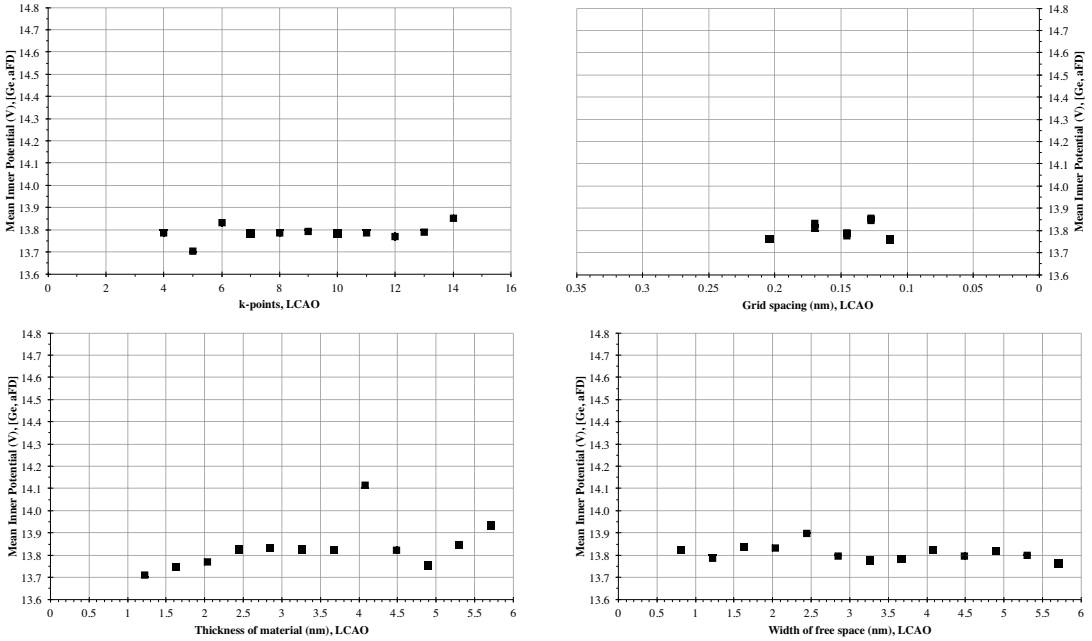


Figure D.18: DFT mean inner potential precision determination for Ge, after Figure 5.9 for C. Like Figure 5.9, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

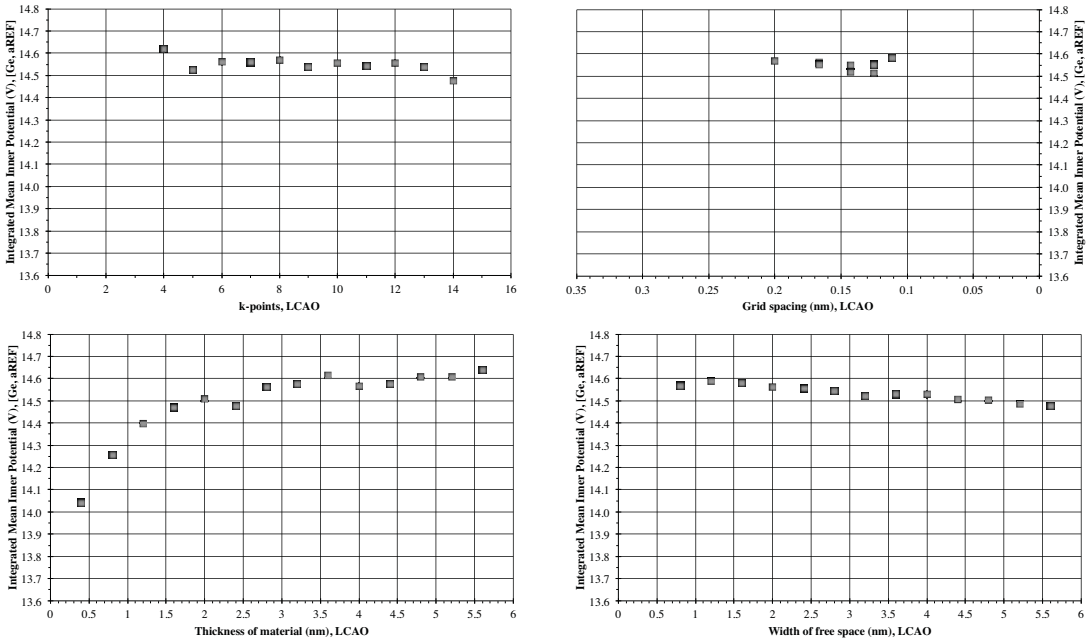


Figure D.19: DFT mean inner potential precision determination for Ge, after Figure 5.10 for C. Like Figure 5.10, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

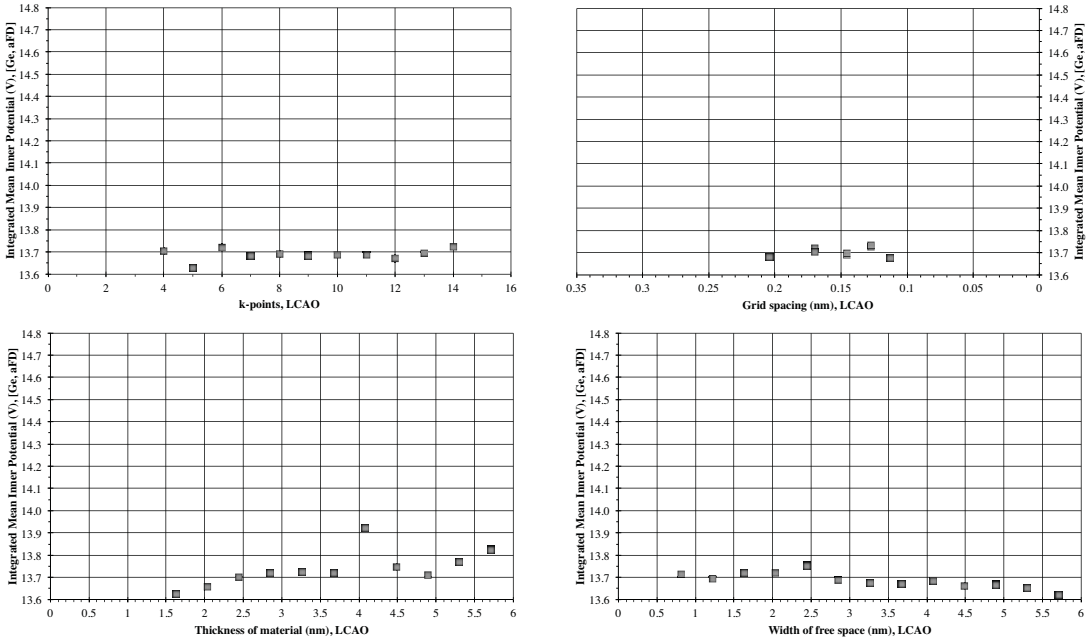


Figure D.20: DFT mean inner potential precision determination for Ge, after Figure 5.11 for C. Like Figure 5.11, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

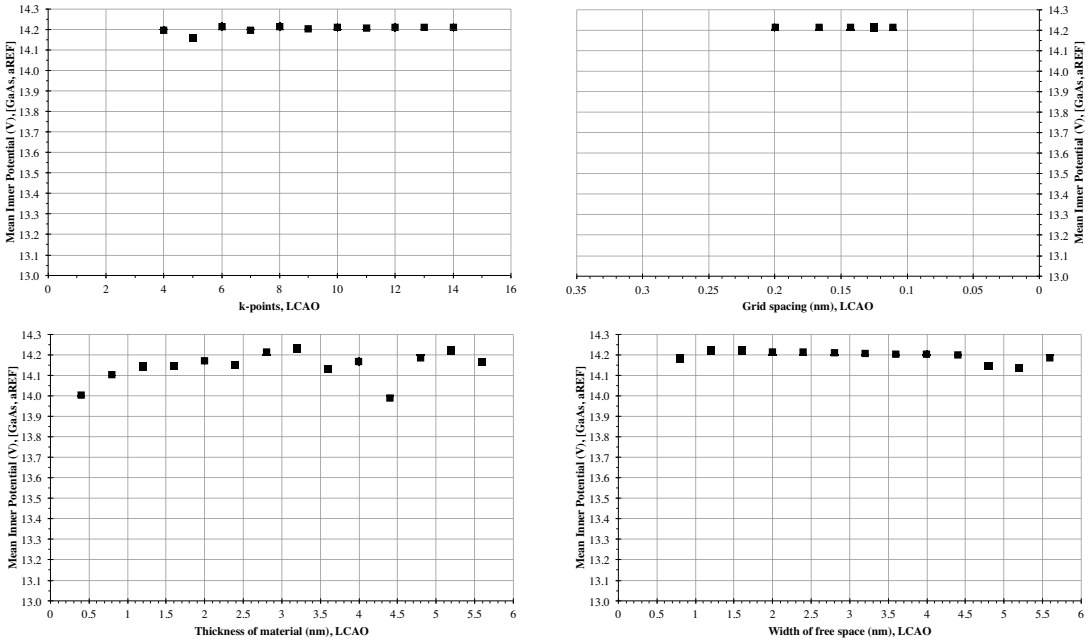


Figure D.21: DFT mean inner potential precision determination for GaAs, after Figure 5.8 for C. Like Figure 5.8, this is using the a_{REF} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

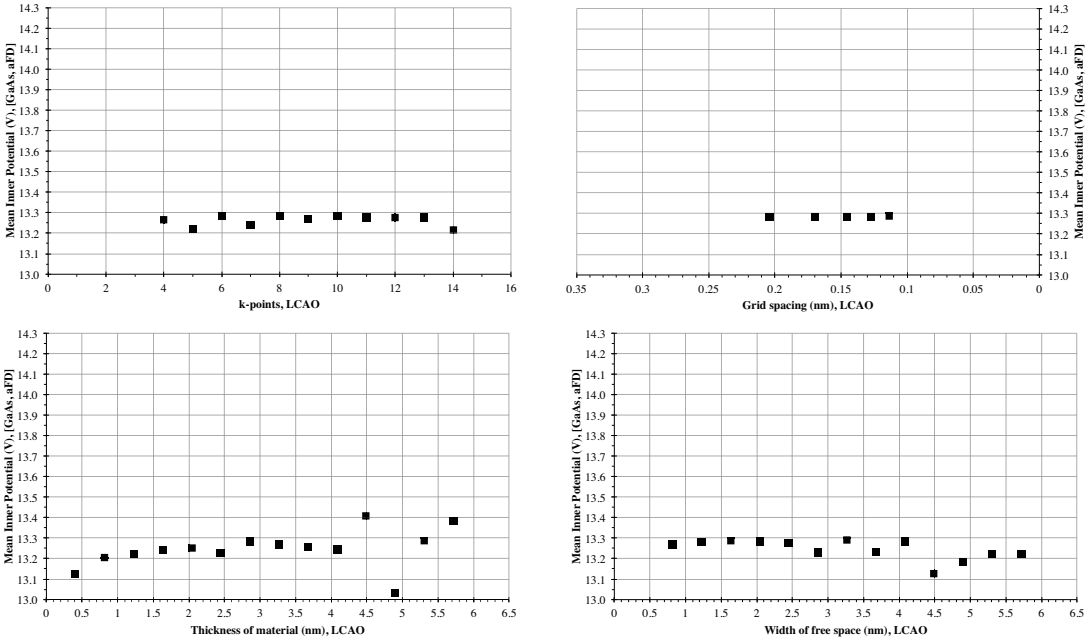


Figure D.22: DFT mean inner potential precision determination for GaAs, after Figure 5.9 for C. Like Figure 5.9, this is using the a_{FD} lattice parameter, the subtraction-method for calculating the mean inner potential, and LCAO mode.

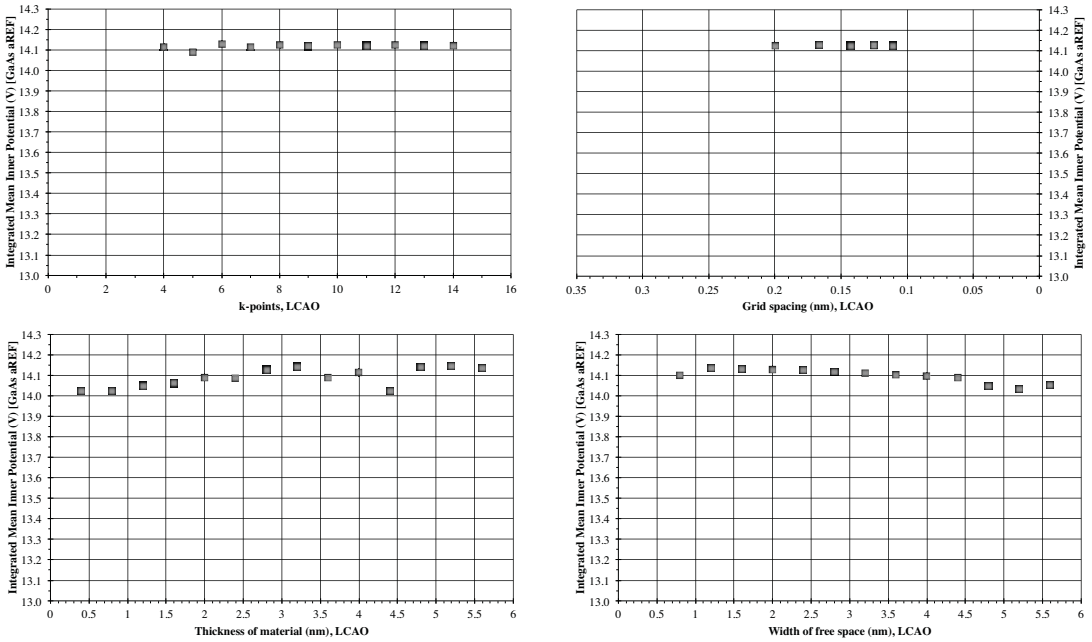


Figure D.23: DFT mean inner potential precision determination for GaAs, after Figure 5.10 for C. Like Figure 5.10, this is using the a_{REF} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

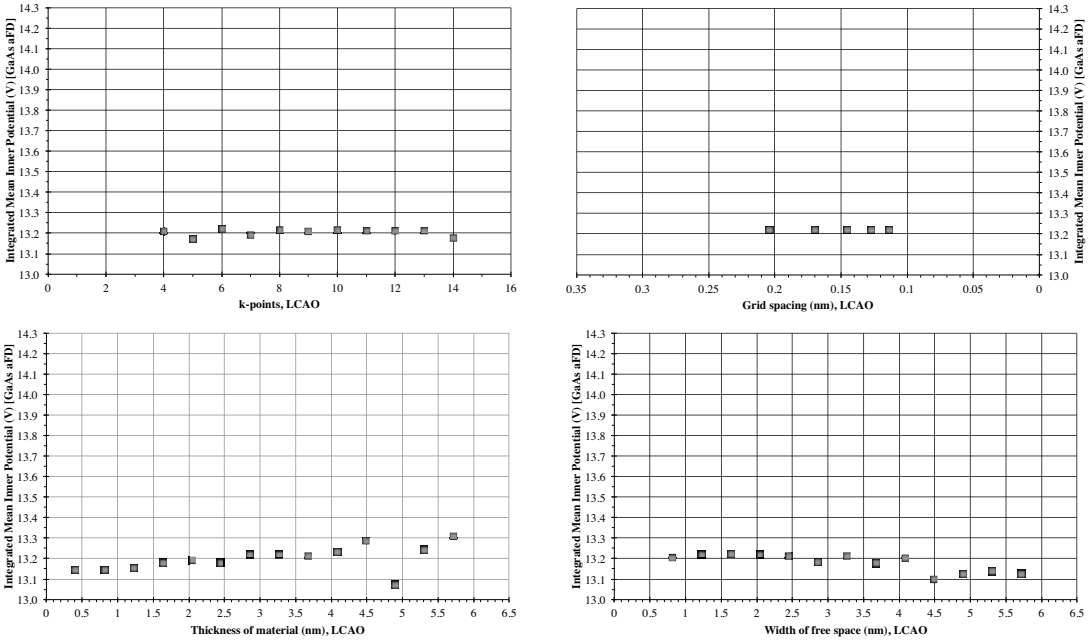


Figure D.24: DFT mean inner potential precision determination for GaAs, after Figure 5.11 for C. Like Figure 5.11, this is using the a_{FD} lattice parameter, the integration-method for calculating the mean inner potential, and LCAO mode.

List of Acronyms

ART	Algebraic reconstruction technique
AT	Algebraic tomography
ATEM	Analytical transmission electron microscope (One of the instruments installed at DTU Cen)
BFTEM	Bright-field TEM
CCD	Charge-coupled device
DART	Discrete algebraic reconstruction technique
DFT	Density functional theory
DFTEM	Dark-field TEM
DT	Discrete tomography
DTU Cen	Technical University of Denmark, Center for Electron Nanoscopy
ETEM	Environmental TEM (a technique, or one of the instruments installed at DTU Cen)
FD	Finite difference
FEG	Field emission gun
GGA	Generalized gradient approximation
GPAW	Grid-based projector-augmented wavefunction
GT	Geometric tomography
HAADF	High-angle annular dark-field
LAADF	Low-angle annular dark-field
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
MIP	Mean inner potential
RENOIR	Renormalized orthogonal-images routine
RMS	Root-mean-square
SIRT	Simultaneous iterative reconstruction technique
STEM	Scanning transmission electron microscope/microscopy
TEM	Transmission electron microscope/microscopy
WBP	Weighted back-projection

References

[1] Williams and Carter. *Transmission Electron Microscopy: A Textbook for Materials Science, 2nd Edition*. Springer. (2009)

[2] Gabor. A New Microscopic Principle, *Nature* (1948) vol. 161 (4098) pp. 777-778.

[3] Möllenstedt and Düker. Beobachtungen und Messungen an Biprisma-Interferenzen mit Elektronenwellen. *Z. Phys. A* (1956) vol. 145 (3) pp. 377-397

[4] McCartney *et al*.. Quantitative phase imaging of nanoscale electrostatic and magnetic fields using off-axis electron holography. *Ultramicroscopy* (2010) vol. 110 (5) pp. 375-382

[5] Hanszen. Methods of off-axis electron holography and investigations of the phase structure in crystals. *J. Phys. D: Appl. Phys.* (1986) vol. 19 (3) pp. 373-395

[6] Midgley and Dunin-Borkowski. Electron tomography and holography in materials science. *Nature Materials* (2009) vol. 8 pp. 271–280

[7] Edited by Völkl, Allard, Joy. *Introduction to Electron Holography*. Plenum Publishers. (1999)

[8] Kruse *et al*.. Determination of the mean inner potential in III–V semiconductors, Si and Ge by density functional theory and electron holography. *Ultramicroscopy* (2006) vol. 106 (2) pp. 105-113

[9] Howie. Hunting the Stobbs factor. *Ultramicroscopy* (2004) vol. 98 (2-4) pp. 73-79

[10] Nobelprize.org. "Ernst Ruska - Autobiography". 21 Nov 2011 http://www.nobelprize.org/nobel_prizes/physics/laureates/1986/ruska-autobio.html

[11] Nobelprize.org. "Dennis Gabor - Autobiography". 21 Nov 2011 http://www.nobelprize.org/nobel_prizes/physics/laureates/1971/gabor.html

[12] Haider *et al*.. A spherical-aberration-corrected 200 kV transmission electron microscope. *Ultramicroscopy* (1998) vol. 75 (1) pp. 53-60

[13] Batson *et al*.. Sub-ångstrom resolution using aberration corrected electron optics. *Nature* (2002) vol. 418 (6898) pp. 617-620

[14] Smith. Progress and perspectives for atomic-resolution electron microscopy. *Ultramicroscopy* (2008) vol. 108 (3) pp. 159-166

[15] Krivanek *et al*.. An electron microscope for the aberration-corrected era. *Ultramicroscopy* (2008) vol. 108 (3) pp. 179-195

[16] Schimmel. The succession of generations of electron microscopists. *Ultramicroscopy* (1986) vol. 20 (3) pp. 183-188

[17] Lichte. Electron holography: optimum position of the biprism in the electron microscope. *Ultramicroscopy* (1996) vol. 64 (1-4) pp. 79-86

[18] Fultz and Howe. *Transmission Electron Microscopy and Diffractometry of Materials, 2nd Edition*. Springer-Verlag. (2002)

[19] Edited by Lundqvist. *Nobel Lectures, Physics 1971-1980*. World Scientific Publishing Co. (1992)

[20] Lichte and Lehmann. Electron holography - basics and applications. *Reports on Progress in Physics* (2008) vol. 71 (1) pp. 016102

[21] LeBeau and Stemmer. Experimental quantification of annular dark-field images in scanning transmission electron microscopy. *Ultramicroscopy* (2008) vol. 108 (12) pp. 1653-1658

[22] Spence and Zuo. Does electron holography energy-filter?. *Ultramicroscopy* (1997) vol. 69 (3) pp. 185-190

[23] Missiroli *et al*.. Electron interferometry and interference electron microscopy. *J. Phys. E: Sci. Instrum.* (1981) vol. 14 pp. 649-671
[24] Cowley. Twenty forms of electron holography. *Ultramicroscopy* (1992) vol. 41 (4) pp. 335-348
[25] Hÿtch *et al*.. Nanoscale holographic interferometry for strain measurements in electronic devices. *Nature* (2008) vol. 453 (7198) pp. 1086-1089
[26] Béché *et al*.. Dark field electron holography for strain measurement. *Ultramicroscopy* (2011) vol. 111 (3) pp. 227-238
[27] Lichte. Performance limits of electron holography. *Ultramicroscopy* (2008) vol. 108 (3) pp. 256-262
[28] Harscher *et al*.. Interference experiments with energy filtered electrons. *Ultramicroscopy* (1997) vol. 69 (3) pp. 201-209
[29] Potapov *et al*.. Inelastic electron holography as a variant of the Feynman thought experiment. *Ultramicroscopy* (2007) vol. 107 (8) pp. 559-567
[30] Lichte *et al*.. Artefacts in electron holography. *Ultramicroscopy* (1996) vol. 64 (1-4) pp. 67-77
[31] Matteucci *et al*.. Interference electron microscopy in thin film investigations. *Thin Solid Films* (1979) vol. 62 (1) pp. 5-17
[32] Fukuhara *et al*.. Electron holography and magnetic specimens. *Phys. Rev. B* (1983) vol. 27 (3) pp. 1839-1843
[33] Osakabe *et al*.. Experimental confirmation of Aharonov-Bohm effect using a toroidal magnetic field confined by a superconductor. *Phys. Rev. A* (1986) vol. 34 (2) pp. 815-822
[34] Tonomura. Electron-holographic interference microscopy. *Advances in Physics* (1992) vol. 41 (1) pp. 59-103
[35] Cooper *et al*.. Quantitative off-axis electron holography of GaAs p-n junctions prepared by focused ion beam milling. *Journal of Microscopy* (2009) vol. 233 (1) pp. 102-113
[36] Rau *et al*.. Two dimensional mapping of pn junctions by electron holography. *Solid State Phenomena* (1998) vol. 63-64 pp. 525-528
[37] Matteucci *et al*.. Electron holography of long-range electric and magnetic fields. *J. Appl. Phys.* (1991) vol. 69 (4) pp. 1835-1842
[38] den Hertog *et al*.. Off axis holography of doped and intrinsic silicon nanowires: Interpretation and influence of fields in the vacuum. *J. Phys.: Conf. Ser.* (2010) vol. 209 pp. 012027
[39] Simon *et al*.. Electron holography of biological samples. *Micron* (2008) vol. 39 (3) pp. 229-256
[40] McCartney and Smith. Electron holography: phase imaging with nanometer resolution. *Annual Review of Materials Research* (2007) vol. 37 pp. 729-767
[41] McCartney and Gajdardziska-Josifovska. Absolute measurement of normalized thickness, t/λ_i, from off-axis electron holography. *Ultramicroscopy* (1994) vol. 53 (3) pp. 283-289
[42] Gajdardziska-Josifovska and McCartney. Elimination of thickness dependence from medium resolution electron holograms. *Ultramicroscopy* (1994) vol. 53 (3) pp. 291-296
[43] Hanszen. A simple reconstruction device for the interferometric detection of the object phase recorded in off-axis image field holograms. *Ultramicroscopy* (1982) vol. 9 (1-2) pp. 159-165
[44] Smith *et al*.. The importance of beam alignment and crystal tilt in high resolution electron microscopy. *Ultramicroscopy* (1983) vol. 11 (4) pp. 263-281

[45] Gajdardziska-Josifovska et al.. Accurate measurements of mean inner potential of crystal wedges using digital electron holograms. *Ultramicroscopy* (1993) vol. 50 (3) pp. 285-299
[46] Hanszen. Holography in electron microscopy. *Advances in Electronics and Electron Physics* (1982) vol. 59 pp. 1-77
[47] Ade. Digital techniques in electron off-axis holography. *Advances in Electronics and Electron Physics* (1994) vol. 89 pp. 1-51
[48] Formanek and Bugiel. On specimen tilt for electron holography of semiconductor devices. *Ultramicroscopy* (2006) vol. 106 (4-5) pp. 292-300
[49] Stern and Gervais. Inner potential measurements in electron diffraction. *Surface Science* (1969) vol. 17 (2) pp. 273-297
[50] Kruse *et al*.. Determination of the mean inner potential in III-V semiconductors by electron holography. *Ultramicroscopy* (2003) vol. 96 (1) pp. 11-16
[51] Lubk *et al*.. The effect of dynamical scattering in off-axis holographic mean inner potential and inelastic mean free path measurements. *Ultramicroscopy* (2010) vol. 110 (5) pp. 438-446
[52] Pennington *et al*.. Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials. *J. Phys.: Conf. Ser.* (2010) vol. 209 pp. 012030
[53] Hÿtch and Stobbs. Quantitative comparison of high resolution TEM images with image simulations. *Ultramicroscopy* (1994) vol. 53 (3) pp. 191-203
[54] Walther. A new experimental procedure to quantify annular dark field images in scanning transmission electron microscopy. *Journal of Microscopy* (2006) vol. 221 (2) pp. 137-144
[55] Findlay *et al*.. Atomic number contrast in high angle annular dark field imaging of crystals. *Materials Science and Technology* (2008) vol. 24 (6) pp. 660-666
[56] Yu *et al*.. Effects of specimen tilt in ADF-STEM imaging of a-Si/c-Si interfaces. *Ultramicroscopy* (2008) vol. 108 (5) pp. 494-501
[57] Kak and Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics. (2001)
[58] Gardner and Kiderlen. A new algorithm for 3D reconstruction from support functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2009) vol. 31 (3) pp. 556-562
[59] Batenburg and Sijbers. DART: a fast heuristic algebraic reconstruction algorithm for discrete tomography. *IEEE Proceedings - International Conference on Image Processing* (2007) vol. IV pp. 133-136
[60] Batenburg *et al*.. 3D imaging of nanomaterials by discrete tomography. *Ultramicroscopy* (2009) vol. 109 (6) pp. 730-740
[61] Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press. (2005)
[62] Kohn and Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* (1965) vol. 140 (4A) pp. A1133-A1138
[63] Perdew *et al*.. Generalized gradient approximation made simple. *Phys. Rev. Lett.* (1996) vol. 77 (18) pp. 3865-3868
[64] Dunin-Borkowski. The development of Fresnel contrast analysis, and the interpretation of mean inner potential profiles at interfaces. *Ultramicroscopy* (2000) vol. 83 (3-4) pp. 193-216
[65] Saldin and Spence. On the mean inner potential in high- and low-energy electron diffraction. *Ultramicroscopy* (1994) vol. 55 (4) pp. 397-406

[66] Doyle and Turner. Relativistic Hartree-Fock X-ray and electron scattering factors. *Acta Cryst*. (1968) vol. A24 (3) pp. 390-397
[67] Rez *et al*.. Dirac-Fock calculations of X-ray scattering factors and contributions to the mean inner potential for electron scattering. *Acta Cryst*. (1994) vol. A50 (4) pp. 481-497
[68] Buhl. Interferenzmikroskopie mit Elektronenwellen. *Z. Phys. A* (1959) vol. 155 (4) pp. 395-412
[69] Herring *et al*.. Determination of mean inner potential and thickness in wedge shaped specimens by interferometry and electron holography. *Proc. 50th Annual Meeting of the Electron Microscopy Society of America* (1992) pp. 990-991
[70] Gajdardziska-Josifovska *et al*.. Accurate measurements of mean inner potential of crystal wedges using electron holography. *Proc. 50th Annual Meeting of the Electron Microscopy Society of America* (1992) pp. 134-135
[71] de Ruijter *et al*.. Quantification of high-resolution lattice images and electron holograms. *Scanning Microscopy* (1992) vol. Supplement (6) pp. 347-359
[72] Rau *et al*.. Two-dimensional mapping of the electrostatic potential in transistors by electron holography. *Phys. Rev. Lett*. (1999) vol. 82 (12) pp. 2614-2617
[73] Li *et al*.. Semiconductor dopant profiling by off-axis electron holography. *Ultramicroscopy* (2003) vol. 94 (2) pp. 149-161
[74] Weiss *et al*.. Applications of electron holography to the study of interfaces. *Ultramicroscopy* (1993) vol. 50 (3) pp. 301-311
[75] Ravikumar *et al*.. Direct imaging of spatially varying potential and charge across internal interfaces in solids. *Phys. Rev. Lett*. (1995) vol. 75 (22) pp. 4063-4066
[76] Lü *et al*.. Study of band structure $In_xGa_{1-x}N$/GaN multiple quantum wells by high-resolution electron microscopy and electron holography. *Appl. Phys. Lett*. (2005) vol. 86 (4) pp. 041902
[77] Chung *et al*.. Off-axis electron holographic potential mapping across AlGaAs/AlAs/GaAs heterostructures. *J. Appl. Phys*. (2009) vol. 105 (1) pp. 014910-014910-4
[78] Chung *et al*.. Determination of the inelastic mean-free-path and mean inner Potential for AlAs and GaAs using off-axis electron holography and convergent beam electron diffraction. *Microscopy and Microanalysis* (2007) vol. 13 (05) pp. 329-335
[79] Li *et al*.. Determination of mean inner potential of germanium using off-axis electron holography. *Acta Cryst*. (1999) vol. A55 (4) pp. 652-658
[80] Weng *et al*.. The mean inner potential of GaN measured from nanowires using off-axis electron holography. *Materials Research Society Symposium Proceedings* (2006) vol. 892 pp. 209-214
[81] Müller *et al*.. Measurement of the mean inner potential of ZnO nanorods by transmission electron holography. *Appl. Phys. Lett*. (2005) vol. 86 pp. 154108
[82] Lin and Dravid. Mapping the potential of graphite nanotubes with electron holography. *Appl. Phys. Lett*. (1996) vol. 69 (7) pp. 1014-1016
[83] Wang *et al*.. Transmission electron holography of silicon nanospheres with surface oxide layers. *Appl. Phys. Lett*. (1997) vol. 70 (10) pp. 1296-1298
[84] Popescu *et al*.. Increase of the Mean Inner Coulomb Potential of Au in Au Clusters Induced by Surface Tension. *Microsc. Microanal*. (2007) vol. 13 (S03) pp. 138-139
[85] Popescu *et al*.. Increase of the mean inner Coulomb potential in Au clusters induced by surface tension and its implication for electron scattering. *Phys. Rev. B* (2007) vol. 76 (23) pp. 235411

[86] Wanner *et al*.. Electron holography of thin amorphous carbon films: Measurement of the mean inner potential and a thickness-independent phase shift. *Ultramicroscopy* (2006) vol. 106 (4-5) pp. 341-345
[87] Ortolani *et al*.. Surface electrostatic potentials in carbon nanotubes and graphene membranes investigated with electron holography. *Carbon* (2011) vol. 49 (4) pp. 1423-1429
[88] Yamamoto *et al*.. Hologram simulation for off-axis electron holography. *Ultramicroscopy* (2000) vol. 85 (1) pp. 35-49
[89] Kim *et al*.. Ab-initio LDA calculations of the mean Coulomb potential V_0 in slabs of crystalline Si, Ge and MgO. *Phys. Status Solidi A* (1998) vol. 166 (1) pp. 445-451
[90] Schowalter *et al*.. Ab initio computation of the mean inner Coulomb potential of wurtzite-type semiconductors and gold. *Appl. Phys. Lett.* (2006) vol. 88 (23) pp. 232108
[91] Schowalter *et al*.. First-principles calculations of the mean inner Coulomb potential for sphalerite type II-VI semiconductors. *Appl. Phys. Lett.* (2004) vol. 85 (21) pp. 4938-4940
[92] Schowalter *et al*.. Ab initio computation of the mean inner Coulomb potential of amorphous carbon structures. *Appl. Phys. Lett.* (2005) vol. 86 (11) pp. 112102
[93] O'Keeffe and Spence. On the average Coulomb potential Φ_0 and constraints on the electron density in crystals. *Acta Cryst.* (1994) vol. A50 (1) pp. 33-45
[94] Spence. On the accurate measurement of structure-factor amplitudes and phases by electron diffraction. *Acta Cryst.* (1993) vol. A49 (2) pp. 231-260
[95] Dick. A review of nanowire growth promoted by alloys and non-alloying elements with emphasis on Au-assisted III–V nanowires. *Progress in Crystal Growth and Characterization of Materials* (2008) vol. 54 (3-4) pp. 138-173
[96] Wagner *et al*.. Growth and segregation of GaAs–$Al_xIn_{1-x}P$ core-shell nanowires. *Journal of Crystal Growth* (2010) vol. 312 (10) pp. 1755-1760
[97] Givargizov and Sheftal. Morphology of silicon whiskers grown by the VLS-technique. *Journal of Crystal Growth* (1971) vol. 9 pp. 326-329
[98] Larsson *et al*.. Strain mapping in free-standing heterostructured wurtzite InAs/InP nanowires. *Nanotechnology* (2007) vol. 18 (1) pp. 015504
[99] Dick *et al*.. InAs nanowires grown by MOVPE. *Journal of Crystal Growth* (2007) vol. 298 pp. 631-634
[100] Zürner *et al*. Discrete tomography of demanding samples based on a modified SIRT algorithm. *Ultramicroscopy* (2012). In press.
[101] Abràmoff *et al*.. Image processing with ImageJ. *Biophotonics International* (2004) vol. 11 (7) pp. 36-43
[102] Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision* (1987) vol. 1 (2) pp. 167-187
[103] Houben and Bar Sadan. Refinement procedure for the image alignment in high-resolution electron tomography. *Ultramicroscopy* (2011) vol. 111 (9-10) pp. 1512-1520
[104] King. *CRC Handbook of Chemistry and Physics, 89th Edition: Crystal Structures and Lattice Parameters of Allotropes of the Elements*. CRC Press. (2008)
[105] McCaffrey and Baribeau. A transmission electron microscope (TEM) calibration standard sample for all magnification, camera constant, and image/diffraction pattern rotation calibrations. *Microscopy Research and Technique* (1995) vol. 32 (5) pp. 449-454

[106] Nepijko *et al*.. Dependence of lattice parameter on particle size. *Phys. Status Solidi A* (1980) vol. 61 (2) pp. 469-475
[107] Messaoudil *et al*.. TomoJ: tomography software for three-dimensional reconstruction in transmission electron microscopy. *Bmc Bioinformatics* (2007) vol. 8 pp. 288
[108] Gardner and Milanfar. Reconstruction of convex bodies from brightness functions. *Discrete and Computational Geometry* (2003) vol. 29 (2) pp. 279-303
[109] Cooper *et al*.. Medium resolution off-axis electron holography with millivolt sensitivity. *Appl. Phys. Lett.* (2007) vol. 91 (14) pp. 143501
[110] Stadelmann. EMS - a software package for electron diffraction analysis and HREM image simulation in materials science. *Ultramicroscopy* (1987) vol. 21 (2) pp. 131-145
[111] den Hertog *et al*.. Off axis holography of doped and intrinsic silicon nanowires: Interpretation and influence of fields in the vacuum. *J. Phys.: Conf. Ser.* (2010) vol. 209 pp. 012027
[112] den Hertog. Caractérisation de nanofils de silicium par microscopie électronique en transmission. PhD Thesis. (2009)
[113] Berger. *CRC Handbook of Chemistry and Physics, 88th Edition: Properties of Semiconductors*. CRC Press. (2007)
[114] Mortensen *et al*.. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* (2005) vol. 71 (3) pp. 035109
[115] Jakobsen and Bahn. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science and Engineering* (2002) vol. 4 (3) pp. 56-66
[116] Larsen *et al*.. Localized atomic basis set in the projector augmented wave method. *Phys. Rev. B* (2009) vol. 80 (19) pp. 195112
[117] Enkovaara *et al*.. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* (2010) vol. 22 (25) pp. 253202
[118] Kleinman. Comment on the average potential of a Wigner solid. *Phys. Rev. B* (1981) vol. 24 (12) pp. 7412-7414
[119] Kittel. *Introduction to Solid State Physics, Sixth Edition*. John Wiley & Sons. 1986.
[120] Sque *et al*.. Structure, electronics, and interaction of hydrogen and oxygen on diamond surfaces. *Phys. Rev. B* (2006) vol. 73 (8) pp. 085313
[121] Kurth *et al*.. Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs. *International Journal of Quantum Chemistry* (1999) vol. 75 (4-5) pp. 889-909
[122] Over *et al*.. Surface atomic geometry of Si(001)-(2×1): A low-energy electron-diffraction structure analysis. *Phys. Rev. B* (1997) vol. 55 (7) pp. 4731-4736
[123] Stallcup and Perez. Atomic structure of steps and defects on the clean diamond (100)-2×1 surface studied using ultrahigh vacuum scanning tunneling microscopy. *Appl. Phys. Lett.* (2002) vol. 81 (24) pp. 4538-4540
[124] Lauridsen *et al*. Structure determination of the Si(001)-(2×1)-H reconstruction by surface X-ray diffraction: weakening of the dimer bond by the addition of hydrogen. *Surface Science* (2000) vol. 453 (1-3) pp. 18-24
[125] *CRC Handbook of Chemistry and Physics, 90th Edition: Electron Work Function of the Elements*. CRC Press. (2009)
[126] Fall *et al*.. Deriving accurate work functions from thin-slab calculations. *J. Phys.: Condens. Matter* (1999) vol. 11 (13) pp. 2689-2696

[127] Fall *et al*.. Work functions at facet edges. *Phys*. *Rev*. *Lett*. (2002) vol. 88 (15) pp. 156802
[128] Fall *et al*.. Work functions and surface charges at metallic facet edges. *Phys*. *Rev*. *B* (2002) vol. 66 (7) pp. 075405
[129] Zangwill. *Physics At Surfaces*. Cambridge University Press. (1988)
[130] Brandes and Mills. Work function and affinity changes associated with the structure of hydrogen-terminated diamond (100) surfaces. *Phys*. *Rev*. *B* (1998) vol. 58 (8) pp. 4952-4962
[131] Lopinski *et al*.. Enhanced conductance of chlorine-terminated Si(111) surfaces: Formation of a two-dimensional hole gas via chemical modification. *Phys*. *Rev*. *B* (2005) vol. 71 (12) pp. 125308
[132] Hilner *et al*.. Direct atomic scale imaging of III–V nanowire surfaces. *Nano Letters* (2008) vol. 8 (11) pp. 3978-3982
[133] Hansen *et al*.. Aberration corrected and monochromated environmental transmission electron microscopy: challenges and prospects for materials science. *Materials Science and Technology* (2010) vol. 26 (11) pp. 1338-1344
[134] Hofmann *et al*.. Ledge-flow-controlled catalyst interface dynamics during Si nanowire growth. *Nature Materials* (2008) vol. 7 (5) pp. 372-375
[135] Hannon *et al*.. The influence of the surface migration of gold on the growth of silicon nanowires. *Nature* (2006) vol. 440 (7080) pp. 69-71
[136] Ross *et al*.. Sawtooth faceting in silicon nanowires. *Phys*. *Rev*. *Lett*. (2005) vol. 95 (14) pp. 146104
[137] Persson *et al*.. Solid-phase diffusion mechanism for GaAs nanowire growth. *Nature Materials* (2004) vol. 3 (10) pp. 677-681
[138] Pennington *et al*.. Advanced transmission electron microscopy of semiconductor nanowires for solar cell applications. (2010) Workshop Poster, European Graduate School for Sustainable Energy.
[139] Griffiths. *Introduction to Electrodynamics, Third Edition*. Prentice Hall, Inc.. (1999)

Included Papers

Paper I

Reconstruction of an InAs nanowire using geometric and algebraic tomography

Paper II

Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials

Paper III

Atomic resolution imaging of *in situ* InAs nanowire dissolution at elevated temperature

Reconstruction of an InAs nanowire using geometric and algebraic tomography

R S Pennington[1], S König[2], A Alpers[2], C B Boothroyd[1] and R E Dunin-Borkowski[3,1]

[1] Center for Electron Nanoscopy, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[2] Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, D-85747 Garching bei München, Germany
[3] Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, D-52425 Jülich, Germany

E-mail: rp@cen.dtu.dk

Abstract. Geometric tomography and conventional algebraic tomography algorithms are used to reconstruct cross-sections of an InAs nanowire from a tilt series of experimental annular dark-field images. Both algorithms are also applied to a test object to assess what factors affect the reconstruction quality. When using the present algorithms, geometric tomography is faster, but artifacts in the reconstruction may be difficult to recognize.

1. Introduction

Conventional algebraic tomography reconstruction algorithms such as the simultaneous iterative reconstruction technique (SIRT) [1] reconstruct a three-dimensional object of varying density from a tilt series of projected images. In contrast, geometric tomography (GT) [2] reconstructs only the shape of the object, using additional information such as e.g. convexity, if applicable.

Here, we reconstruct the cross-section of an InAs nanowire from a tilt series of experimental annular dark-field images using a GT algorithm [3], and compare to the results from a conventional SIRT reconstruction. Additional comparisons using a simulated test object are used to assess the fidelity of each reconstruction and the ease with which artifacts can be identified.

2. Conventional algebraic reconstruction

A high-angle annular dark-field scanning transmission electron microscopy (HAADF STEM) image series of an InAs nanowire was acquired using a probe-aberration-corrected FEI Titan 80-300 operated at 300 kV with a probe convergence semi-angle of 18 mrad and an inner detector semi-angle of 100 mrad. The images were acquired over a total angular range of 139° with a 1° tilt increment. The original 2048×2048 pixel images were then binned by a factor of 4 to reduce noise in the reconstruction. A bright-field TEM image of the nanowire and two representative HAADF STEM images and line profiles from the tilt series are shown in Fig. 1.

SIRT reconstruction was performed using the TomoJ software with 15 iterations [4, 5] after aligning the original tilt series using cross-correlation and a Canny-Deriche edge-detection filter. Cross-sections of the wire determined from reconstructions using 14 and 140 projections are shown in Fig. 2 and Fig. 3, respectively, for both of the sections of the nanowire marked in Fig. 1.

The SIRT reconstruction is blurred in the direction of the "missing wedge", contains artifacts in the region outside the wire, and requires a subjective choice of threshold to find the nanowire boundary.

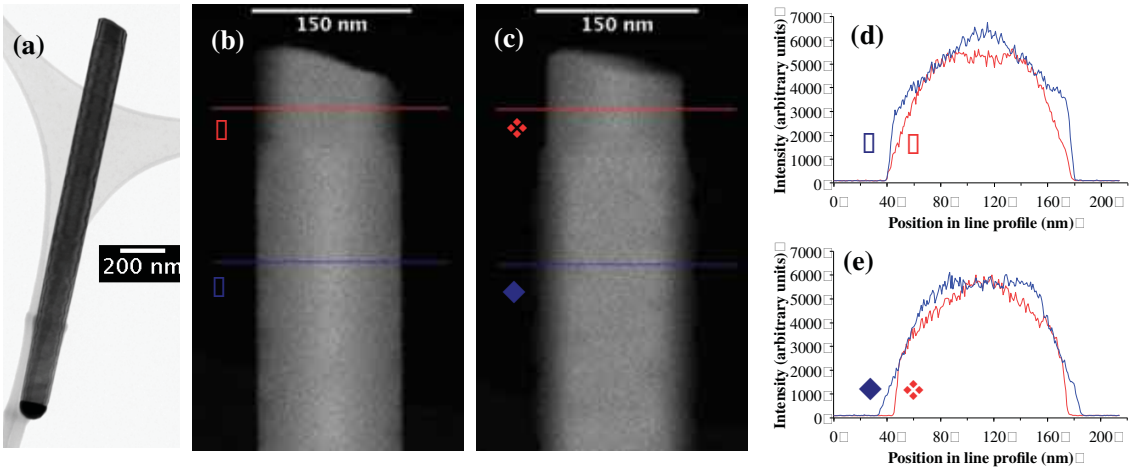


Figure 1. (a) is a bright-field TEM image of the InAs nanowire specimen used for tomography; (b) and (c) aligned binned HAADF STEM images taken from the tilt series of the wire at angles of 7° and 37°, respectively, with 8-pixel-wide line profiles from the marked regions shown in (d) and (e), exhibiting the roughly-hexagonal faceting expected from the reconstruction, but not noise levels. There is a growth defect between the red and blue line profile, leading to rotation in the blue profile relative to the red.

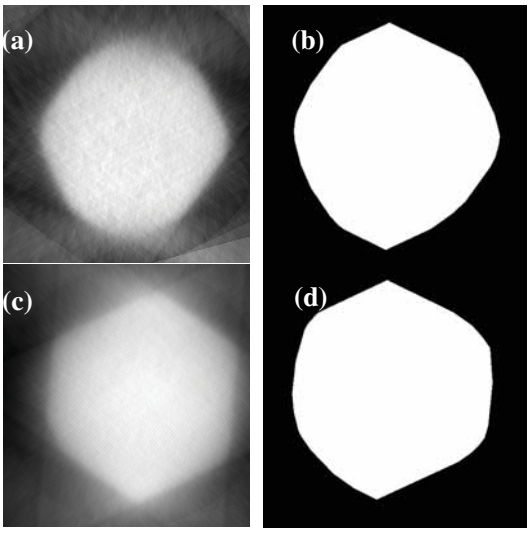


Figure 2. Reconstructed cross-sections from HAADF STEM tilt series using (a,c) SIRT and (b,d) GT algorithms from 14 equally-spaced projections over a tilt range of 139°. (a,b) are from the pale red line in Fig. 1, (c,d) from the dark blue line. Nanowire width is 140 nm measured from profiles in Fig. 1.

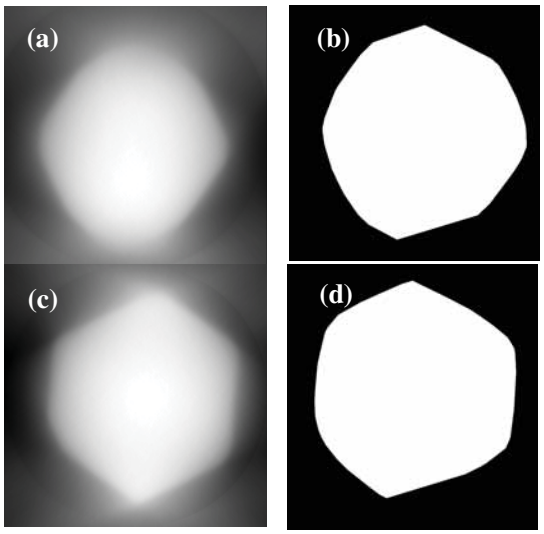


Figure 3. As Fig. 2, but reconstructed from 140 equally-spaced projections over a tilt range of 139°.

3. Geometric tomography from shadow outlines

The GT algorithm used in the present study, introduced by Gardner and Kiderlen [3], reconstructs only from the shadow outlines in each projection, neglecting any density variation. It is computationally simple: after thresholding, it fits a polygon to be consistent with the shadow outline from all projections. The algorithm used here converges as the number of projections approaches infinity [6]. Our implementation is more than an order of magnitude faster than SIRT [4,5]. Previous work has applied the GT concept to TEM images, but the back-projection algorithm used in [7] needs both post-reconstruction thresholding and is not proved to be convergent; the algorithm used in [8] is also not proved to be convergent. Our GT implementation [3] was applied to the same nanowire cross-sections (Fig. 1) that were used in section 2 after thresholding the input images at a value of 300 counts. For the present sample, our GT reconstructions from 14 and 140 projections are nearly identical, but the SIRT reconstructions improve with increased sampling. Our GT implementation and SIRT are in reasonable agreement outside the areas affected by the missing-wedge at the top and the bottom of the wire (Figs. 2, 3), where the SIRT reconstruction is blurred and our GT reconstruction is fully uninterpretable.

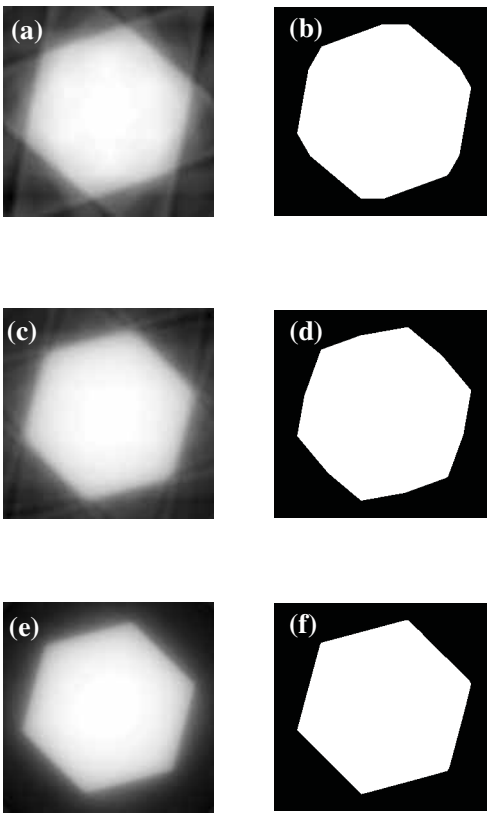


Figure 4. The effect of angular spacing on test data reconstruction for (a,c,e) SIRT and (b,d,f) geometric tomography for (a,b) 20°, (c,d) 10°, and (e,f) 1° with no noise and 180° tilt range.

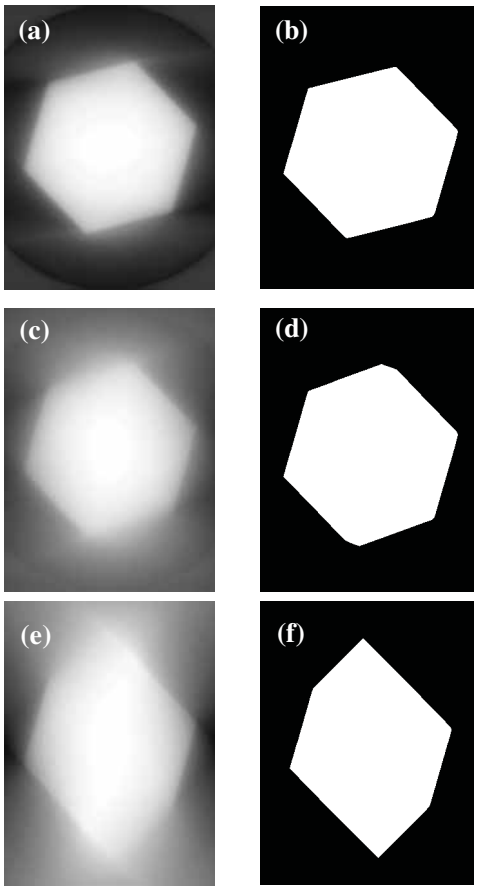


Figure 5. The effect of the missing wedge on test data reconstruction for (a,c,e) SIRT and (b,d,f) geometric tomography for (a,b) 160°, (c,d) 140°, and (e,f) 90° tilt range with no noise and 1° angular spacing.

4. Comparison of both geometric and algebraic algorithms using test data

To compare how individual factors affect this GT implementation and conventional SIRT, a simulated tilt series was generated from a regular, binary hexagon. The angular spacing between images (i.e. number of projections), tilt range and noise level were all individually varied, and image intensity fluctuations were used to simulate diffraction contrast.

Increasing the angular spacing creates artifacts outside the wire for the SIRT reconstruction, but a decrease in corner sharpness using GT (Fig. 4). Limiting the tilt range leads to missing-wedge effects in both algorithms, but the GT missing-wedge is more similar to the rest of the reconstruction (Fig. 5). The GT reconstruction can be sensitive to noise if it makes thresholding difficult, and this can lead to e.g. apparent facet artifacts. (Fig. 6), which may be a problem for the experimental GT reconstruction in section 3. For diffraction contrast, the SIRT reconstruction shows minimal change, but the GT reconstruction shows no change, because the model used here does not affect the outlines (Fig. 7).

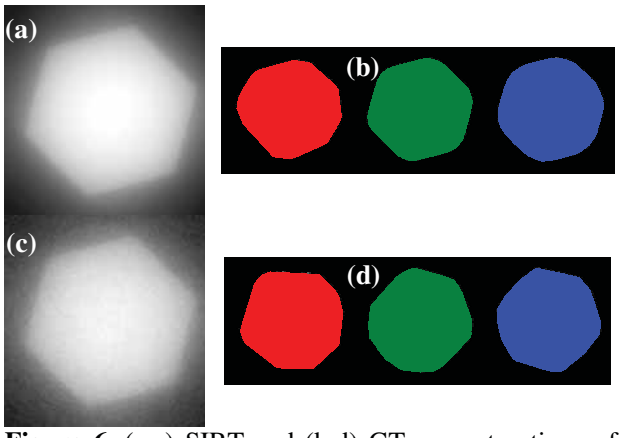


Figure 6. (a,c) SIRT and (b,d) GT reconstructions of the test hexagon with Gaussian noise level 10× (top) and 80× (bottom) the original intensity per object pixel. Color goes from red to green to blue (left to right) in (b,d) as the pre-reconstruction threshold T increases (b: T=300,500,700) (d: T=2000,2700,2900).

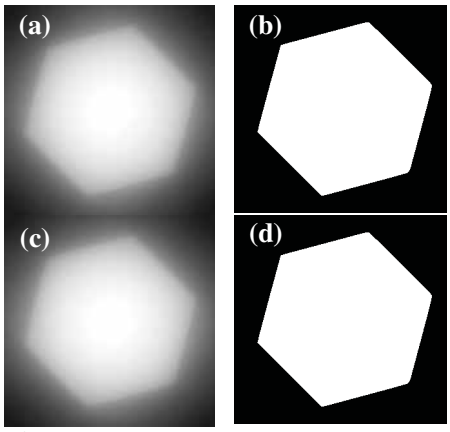


Figure 7. (a,c) SIRT and (b,d) GT reconstructions of the test hexagon for random diffraction contrast (50% chance of ≤25% intensity increase (a,b) or ≤75% intensity increase (c,d)).

5. Conclusions

A geometric tomography algorithm based on image outlines has been compared with algebraic tomography for an InAs nanowire with a convex cross-section. The missing wedge is a problem for both algorithms. Artifacts in the reconstruction may be more difficult to recognize for geometric tomography. Other geometric tomography algorithms can use the image intensity (e.g. [9]), instead of only the shadow outlines, potentially yielding a better reconstruction, especially in the missing-wedge region.

References

[1] Herman G T 2009 *Fundamentals of Computerized Tomography*, (London: Springer)
[2] Gardner R J 2006 *Geometric Tomography* (New York: Cambridge University Press)
[3] Gardner R J and Kiderlen M 2009 *IEEE Trans. Pattern Anal. Machine Intell.* **31** 556-562
[4] Abràmoff M D, Magalhães P J and Ram S J 2004 *Biophotonics International* **11** 36-43
[5] Messaoudil C, Boudier T, Sanchez Sorzano CO and Marco S 2007 *BMC Bioinformatics* **8** 288
[6] Gardner R J, Kiderlen M and Milanfar P 2006 *Ann. Statist.* **34** 1331-1374
[7] Saghi Z, Xu X and Möbus G 2008 *Journal of Microscopy* **232** 186-195
[8] Petersen T C and Ringer S P 2009 *J. Appl. Phys.* **105**, 103518
[9] Gardner R J and Milanfar P 2003 *Discrete and Computational Geometry* **29** 279-303

Journal of Physics: Conference Series **209** (2010) 012030 doi:10.1088/1742-6596/209/1/012030

Theoretical and experimental factors affecting measurements of semiconductor mean inner potentials

RS Pennington[1], JJ Mortensen[2], T Kasama[1], CB Boothroyd[1] and RE Dunin-Borkowski[1]

[1] Center for Electron Nanoscopy, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[2] Center for Atomic-Scale Materials Design, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

E-mail: robert.pennington@cen.dtu.dk

Abstract. We use density functional theory to explore the effect on calculations of semiconductor mean inner potentials of the presence of reconstructions, changes in lattice spacing and adsorbates on the surfaces of parallel-sided thin specimens. We also use electron holography to illustrate several factors that affect experimental measurements of mean inner potentials of semiconductor nanowires.

1. Introduction

Quantitative electron holography of semiconductors is of interest for measuring dopant distributions, specimen thickness profiles, compositions, and charge densities at interfaces and surfaces. For many applications, the mean inner potential V_0 of the semiconductor should be known accurately. However, measurements of V_0 are sensitive to many factors, including specimen charging, bonding, dynamical diffraction and the nature of the specimen surface [1]. In addition, the specimen orientation and thickness must be known accurately. Here, we assess some of the factors that affect predictions and measurements of V_0 by performing calculations of potentials in semiconductors with adsorbates and surface reconstructions and by using electron holography to record phase images of GaAs nanowires.

2. Calculation details and results

V_0 was calculated for selected semiconductors using the projector-augmented wavefunction approach within the density functional theory (DFT) formalism, implemented as GPAW [2]. The calculation volume contained the material of interest and a region with no material, as shown in Fig. 1 and also used by Kruse et al. [3]. The Hartree potential V_H was determined from the valence electrons, screened core electrons and nuclei for region B marked in Fig. 1 relative to that in region A. As will be detailed elsewhere, in order to take into account the unscreened core charges, a second potential V_{EC} was calculated by solving Poisson's equation for the unscreened charge according to the expression [2]

$$\rho(\mathbf{r}) = \tilde{\rho}(\mathbf{r}) + \sum_a \Delta\tilde{\rho}^a(\mathbf{r} - \mathbf{R}^a),$$

after which $V_0 = V_H + V_{EC}$.

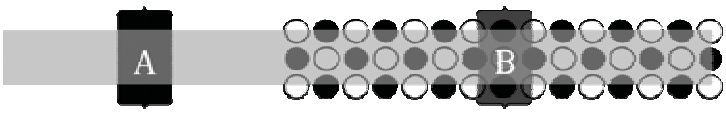


Fig. 1. Calculation geometry used in this work, with the material of interest and a region of vacuum.

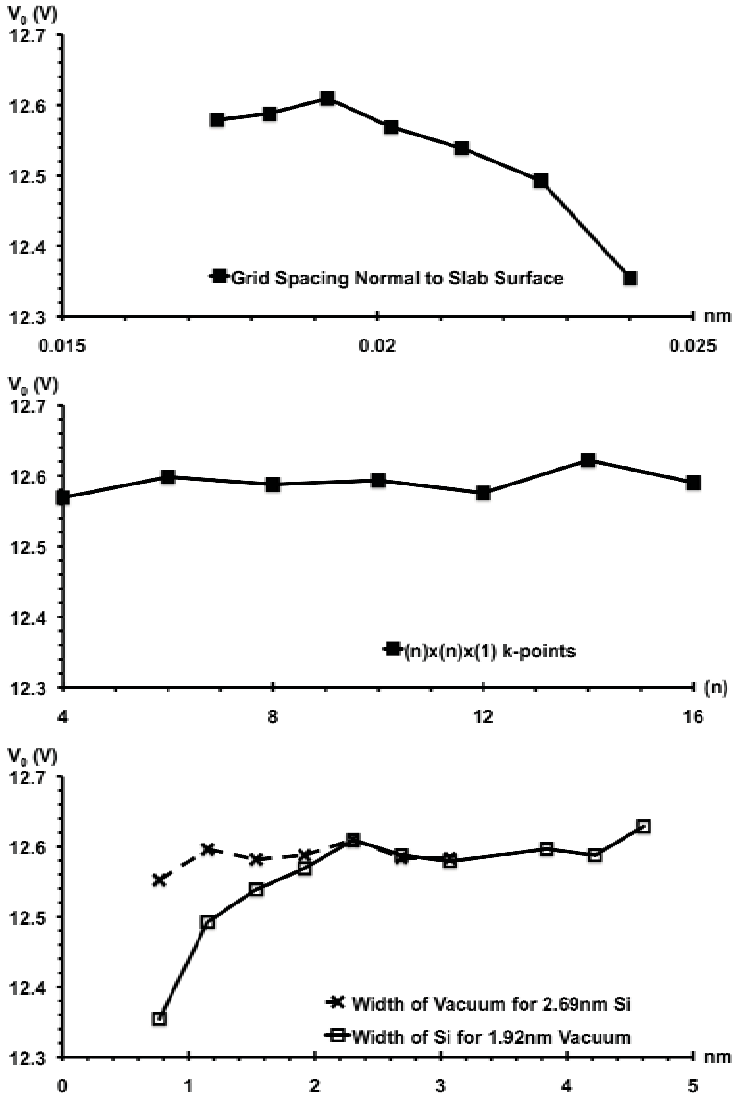


Fig. 2. Verification of parameter stability for calculations of V_0 for parallel-sided Si specimens with (110) surfaces. Grid spacings of $h_x = 0.0226$ nm and $h_y = 0.0192$ nm were used in the [001] and [1-10] directions parallel to the surface of the slab.

The sensitivity of V_0 to a range of input parameters is shown in Fig. 2 for parallel-sided Si specimens with (110) surfaces. As specimens containing only a few monolayers of material have decreased values of V_0, 2.69 nm of Si and 1.92 nm of vacuum were typically used in subsequent calculations for (110) surfaces. It was found that, in order to ensure stability, the atomic lattice of the crystal should be located on grid points in the direction normal to the surface of the slab. Unless otherwise stated, the calculation parameters used were $8 \times 8 \times 1$ k-points, 0.0192 nm per real-space grid point in the direction normal to the surface of the slab, and the PBE exchange-correlation functional [4], parallelized over 8 processor cores using domain decomposition.

2.1. Surface orientation and reconstruction

Table 1 shows the results of calculations of V_0 for parallel-sided Si slabs with different crystallographic orientations, with atom positions that are unchanged from those in a bulk Si crystal, and with the parameters optimized for each simulation. The difference between the calculated values for (110), (111) and (211) surfaces is below 0.05 V. However, the value for a (100) surface is 0.32 V above that for a (110) surface. Most significantly, the introduction of a (2×1) reconstruction on the (100) surface, with atom positions included from Ref. [5], changes V_0 by almost 0.6 V when compared to the unreconstructed (100) surface.

2.2. Adsorbates

Table 2 shows the results of calculations of V_0 for parallel-sided Si slabs with (110) surface orientations that have one atomic layer of either Si, C or Ge on their surfaces, both with the surface atoms located at positions that Si atoms would occupy on a perfect bulk Si lattice and then with the positions of either all of the atoms in the calculation (for a Si surface) or just the surface atoms (for C and Ge surfaces) relaxed until the force on the atoms was below 0.5 eV / nm. Although the arrangements of atoms in the surface layer may not reflect the positions of adsorbed atoms that would occur experimentally, the calculations illustrate the sensitivity of V_0 to the identities and positions of atoms in the surface layers, with a single relaxed adsorbed layer of C changing the calculated value of V_0 by more than 0.5 V.

2.3. Different semiconductors

Table 3 shows a comparison of DFT calculations of V_0 for different semiconductors with values that would be obtained from neutral atom electron scattering factors [6]. The fact that the present DFT values are close to those in Ref. [3] (also reproduced in Table 3) provides confidence in the applicability of the GPAW simulator for calculations of V_0 for a wide range of materials.

Table 1. V_0 for Si calculated as a function of surface orientation for optimized parameters.

Silicon surface	V_0 (V)
(110)	12.59
(111)	12.63
(211)	12.64
(100)	12.91
(100) (2×1) from [5].	12.32

Table 2. V_0 for Si calculated as a function of surface composition and relaxation.

Surface	V_0 (V)
Si (all-relaxed)	12.52
Si(C)	14.12
Si (C-relaxed)	12.01
Si (Ge)	12.46
Si (Ge-relaxed)	12.53

Table 3. V_0 calculated for different semiconductors with (110) surfaces using the present density functional theory (DFT) calculations, compared with values reproduced from the DFT calculations of Ref. [3] and obtained using neutral atom electron scattering factors (V_{fel}) taken from Ref. [6].

Material	V_0 (V) $_{DFT}$	V_0 (V) $_{Ref.\ [3]}$	V_0 (V) $_{fel}$
Si	12.59	12.57	13.76
Ge	14.69	14.67	15.60
AlP	11.40	11.39	13.42
AlAs	12.29	12.34	13.97
GaP	13.57	13.63	14.60
GaAs	14.15	14.19	15.38
InP	13.85	13.90	15.22
InAs	14.13	14.34	15.47
CdS	13.02	-	13.87

3. Electron holography results

Semiconductor nanowires are, in many ways, ideally suited for experimental measurements of V_0 as their orientations and cross-sections can be determined accurately and minimal specimen preparation is required. Fig. 3 shows part of a representative electron hologram of a GaAs nanowire that has a diameter of 160 nm, acquired at 300 kV using an FEI Titan 80-300 TEM operated in Lorentz mode. The figure also shows experimental phase profiles acquired from a wire oriented at the [110] zone axis and rotated about its long axis to several orientations. The phase profiles illustrate four of the difficulties of measuring V_0 experimentally, even from nanowires: the sensitivity of the line profiles to crystallographic orientation as a result of both changes in projected thickness and dynamical contributions to the phase shift, the difficulty of removing phase wraps at the edges of the wire where the sample thickness changes rapidly, the possibility of charging of the wire as a result of secondary electron emission during electron irradiation, resulting in a ramp in vacuum outside the wire for the 19.1° phase profile, and the unknown state of the nanowire surface.

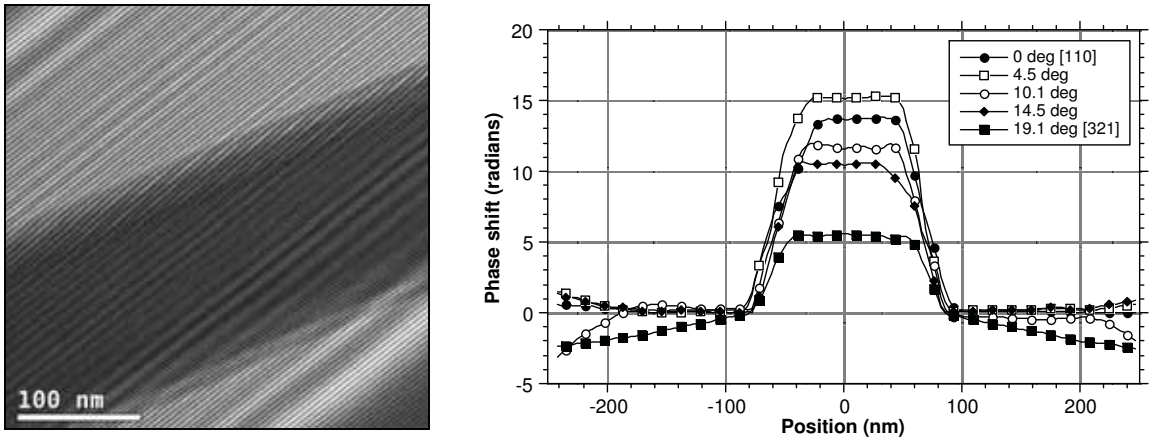


Fig. 3. Electron hologram and phase profiles for a GaAs nanowire examined at different orientations.

4. Conclusions

We have obtained preliminary density functional theory simulations and experimental measurements that illustrate several of the factors that may affect measurements of semiconductor mean inner potentials. The calculations illustrate the sensitivity of the mean inner potential to the presence of adsorbed or reconstructed layers on the specimen surface. The electron holography results illustrate several additional practical complications associated with experimental measurements of mean inner potentials, even for the relatively simple GaAs nanowire system.

Janusz Sadowski of Lund University is thanked for provision of the GaAs nanowire samples.

References

[1] Völkl E, Allard LF and Joy D (Eds) 1999 *Introduction to Electron Holography* (Springer)
[2] Mortensen JJ, Hansen LB and Jacobsen KW 2005 *Phys. Rev. B* **71**, 035109
[3] Kruse P, Schowalter, Lamoen D, Rosenauer A and Gerthsen D 2006 *Ultramicroscopy* **106**, 105
[4] Perdew JP, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77**, 3865
[5] Over H, Wasserfall J, Ranke W, Ambiatello C, Sawitzki R, Moritz W1997 *Phys. Rev. B* **55,** 4731
[6] Rez D, Rez P and Grant I 1994 *Acta Cryst. A* **50**, 481

Journal of Physics: Conference Series **209** (2010) 012013 doi:10.1088/1742-6596/209/1/012013

Atomic resolution imaging of *in situ* InAs nanowire dissolution at elevated temperature

RS Pennington[1], JR Jinschek[2], JB Wagner[1], CB Boothroyd[1] and RE Dunin-Borkowski[1]

[1] Center for Electron Nanoscopy, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[2] FEI Company, Achtseweg Noord 5, 5600 KA Eindhoven, Netherlands

E-mail: robert.pennington@cen.dtu.dk

Abstract. We present a preliminary study of the *in situ* heating of InAs nanowires in a gaseous environment in an environmental transmission electron microscope. Nanowire dissolution, accompanied by dynamic reshaping of crystalline Au-containing catalyst particles at the ends of the wires, is observed, accompanied by indium oxide crystallite formation nearby.

1. Introduction

Catalytic growth of semiconductor nanowires is of considerable interest [1], but is poorly understood at the atomic level. Some nanowires can be grown below the eutectic temperature of the catalyst-semiconductor alloy, leading to the proposal of a vapor-solid-solid growth mechanism [2]. Here, we examine InAs nanowires, whose growth by chemical vapor deposition is known to drop abruptly above ~500 ºC [3]. We image nanowires just above this temperature in the transmission electron microscope (TEM) in a gaseous environment, as a preliminary experiment aimed at understanding the behavior and imaging properties of semiconductor nanowires in an environmental TEM, in advance of attempting true *in situ* III-V nanowire growth. We observe a complex dynamic nanowire dissolution process accompanied by subsidiary crystallite formation.

2. Experimental Details

InAs nanowires containing InP barriers were grown by chemical beam epitaxy at 390ºC. A bright-field TEM image showing a selection of nanowires, which are each capped by a Au catalyst particle and supported on a C film for TEM imaging, is shown in Fig. 1. The nanowires were examined at 300 kV in an image-aberration-corrected FEI Titan 80-300 TEM equipped with a differentially-pumped environmental cell. Motivated by recent reports that the surfaces of similar nanowires can be cleaned using atomic hydrogen [4], the nanowires were examined at elevated temperature in a H_2 atmosphere.

3. Results

The InAs nanowires were heated to 520 ºC in 1.3 mbar H_2 *in situ* in the TEM. Unexpectedly, under these conditions, the Au-containing particles changed shape dynamically while remaining crystalline, as confirmed by the presence of lattice fringes in the image shown in Fig. 2. Dissolution of the wires proceeded, apparently at the catalyst-nanowire interface, with the wires shrinking in length. New crystallites were observed to form on the support film adjacent to the nanowires.

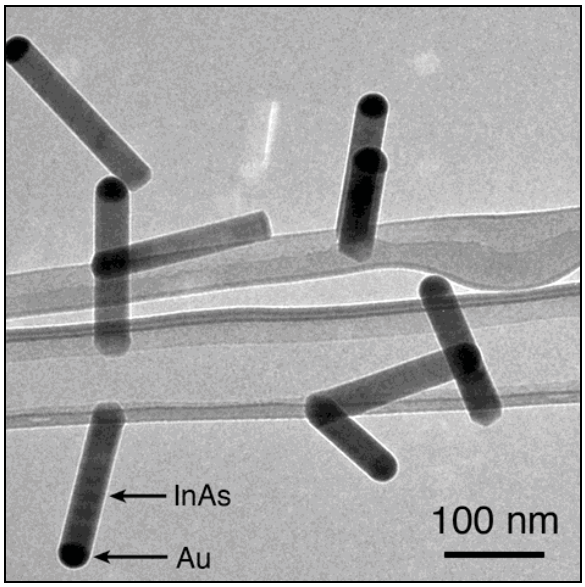


Fig. 1. Bright-field image of nanowires imaged at room temperature.

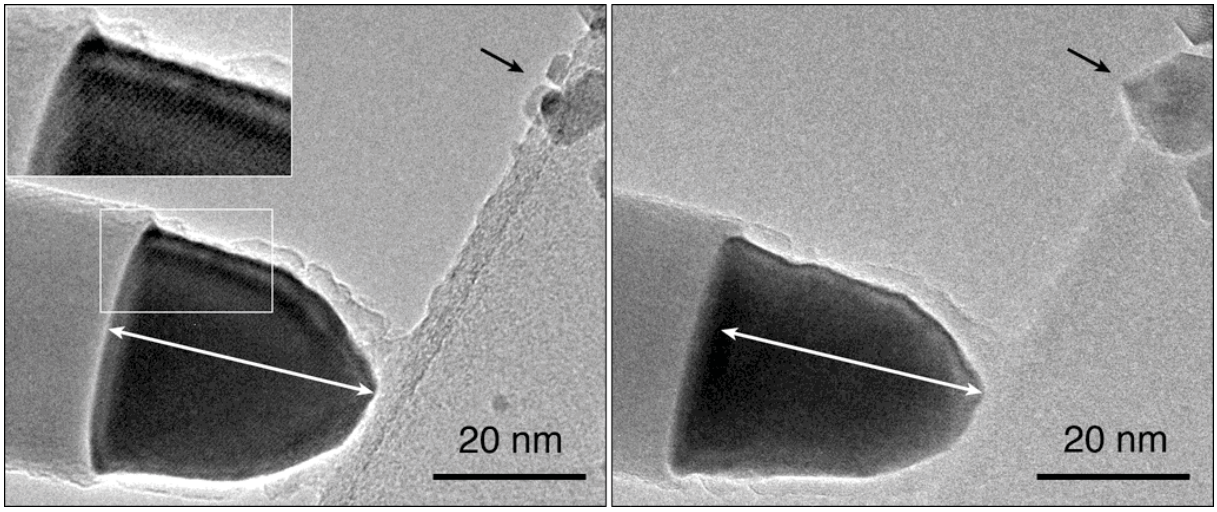


Fig. 2. Images acquired 3 minutes apart at 520 ºC in 1.3 mbar H_2, showing lattice fringes in a Au-containing particle at the end of an InAs nanowire. The white arrows, which have the same length, illustrate the elongation of the catalyst particle. New crystallites form close to the wire at the position indicated by the black arrow.

Fig. 3 shows three snapshots taken from a movie illustrating dynamic changes in the shape of a crystalline Au-containing particle during dissolution of an InAs nanowire. The brighter region in the wire is a segment of InP. Dissolution of the InP segment is observed to proceed faster than that of the InAs. Crystallites are observed to form in the upper-right corner of each frame. Lattice fringe contrast in both the particle and the nanowire is visible in Fig. 4, which shows the presence of a ledge at the interface during nanowire dissolution. Fig. 5 shows high-resolution lattice images acquired both during and after crystallite formation, as well as an electron energy-loss spectrum, confirming that the crystallites are In_2O_3. Neither As nor P were detected in the crystallites. The original thin amorphous coating on the nanowire surface, which is visible in Figs 2 and 3, remained as a product next to the Au particle. This coating was subsequently identified using electron energy-loss spectroscopy as being composed primarily of C.

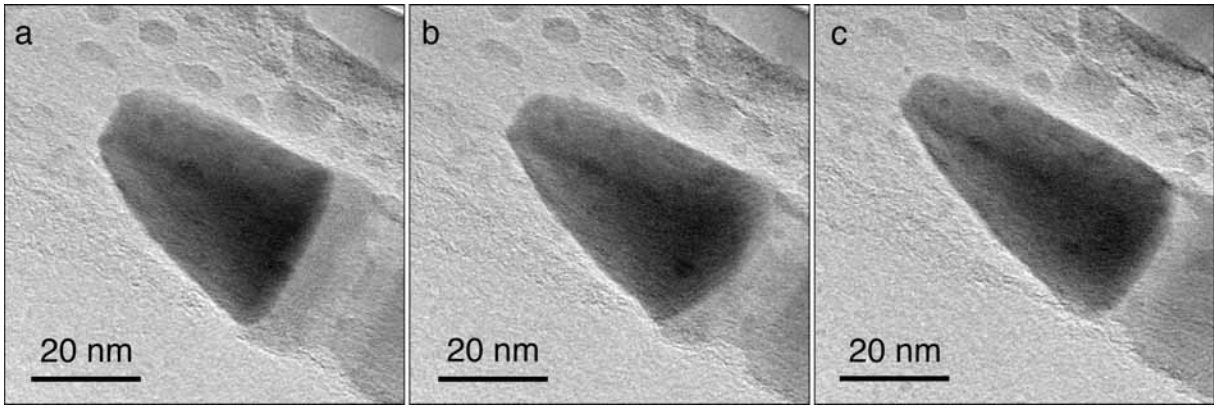


Fig. 3. Three frames, separated by 7 seconds, taken from a movie of InAs nanowire dissolution. A short segment of InP is visible as a lighter-contrast segment of the wire.

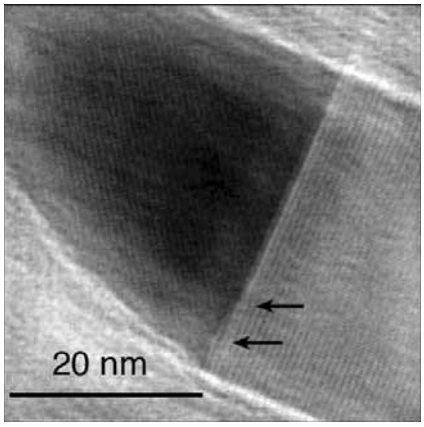


Fig. 4. One Fourier-filtered frame taken from a movie acquired during InAs nanowire dissolution. Lattice fringe contrast is visible both in the Au-containing particle and in the InAs nanowire, indicating that they are both crystalline at this temperature (520 ºC). The image also confirms that lattice fringe contrast can be observed in the presence of 1.3 mbar of H_2 in the differentially pumped section of the electron microscope column. Ledges are visible at the particle-nanowire interface (arrowed).

4. Discussion

The crystalline nature of the Au-containing particle at the temperature used in the present experiments, which is above the growth temperature of the InAs nanowires, suggests that the particle may also have been crystalline during nanowire growth at lower temperature. The interaction between the Au-containing particle and the support film may play a role in the observed shape changes. The formation of In_2O_3 crystallites is surprising as there is no intentional source of oxygen in the microscope. Some native oxide on the nanowire is likely, but residual gas in the environmental cell is a more likely source given the amount required for particle formation. The present observations illustrate the complexity of even a nominally straightforward heating experiment carried out in an environmental TEM. Parameters that may affect the observed dissolution process include specimen temperature, gas environment, the interaction between the Au particle and the carbon support film and the effect of the incident electron beam both on the specimen and on the surrounding gas.

5. Conclusions

Preliminary results from *in situ* heating of InAs nanowires in H_2 in an environmental TEM reveal, at atomic resolution, a complicated nanowire dissolution process, which appears to proceed at the nanowire-catalyst interface. During this process, the Au-containing catalyst particle remains crystalline but changes shape dynamically. At the same time, In_2O_3 crystallites form adjacent to the wire. All aspects of such a simplified experiment should ideally be understood before proceeding to the additional complication of introducing reactive gases to grow III-V nanowires *in situ* in the TEM.

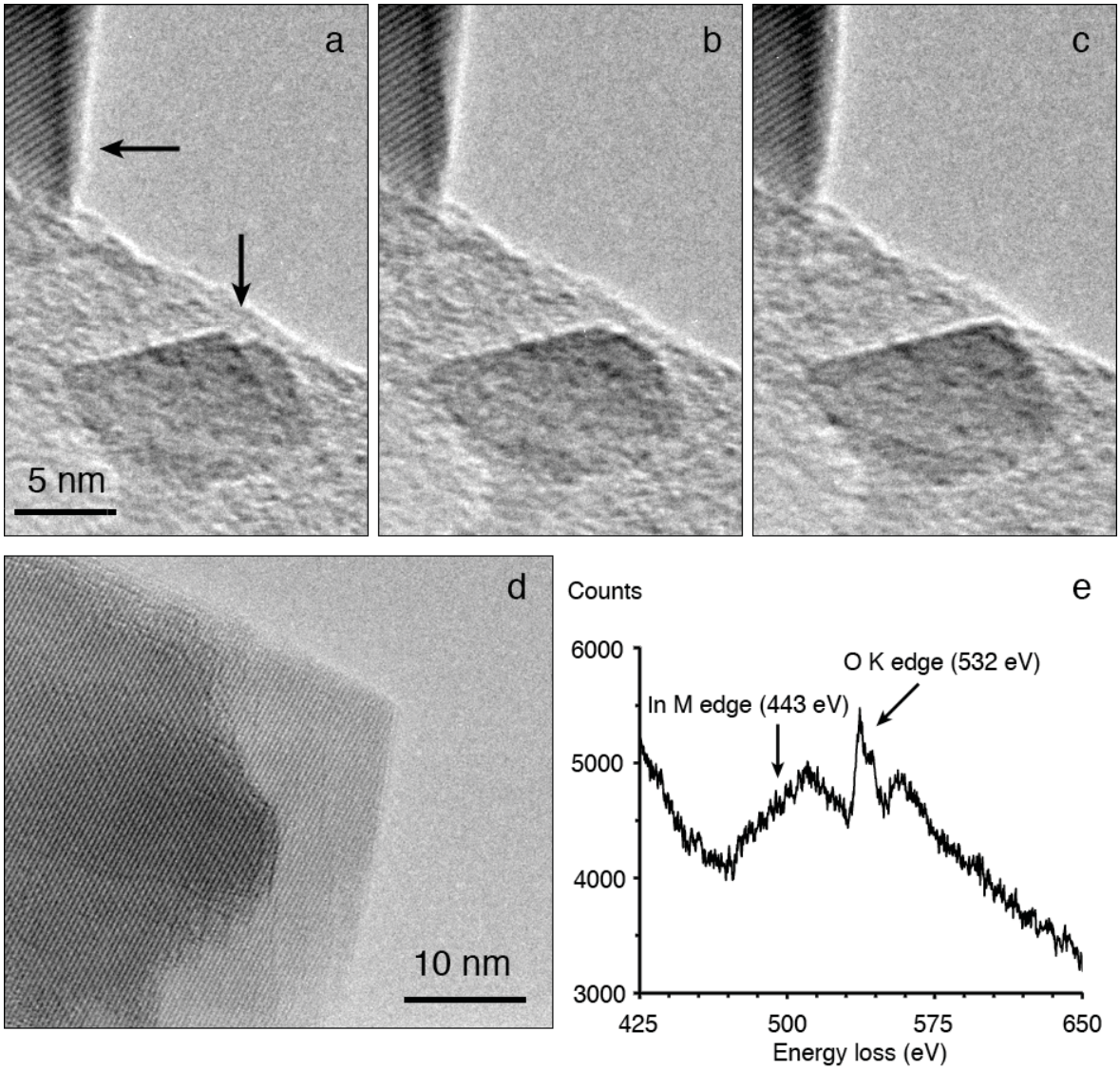


Fig. 5. a) - c) High-resolution lattice images showing crystallite formation adjacent to a nanowire. d) High-resolution lattice image and e) electron energy-loss spectrum acquired from crystallites after nanowire dissolution, with In and O peak edges indicated [5], confirming the formation of In_2O_3.

Magnus Borgström and Linus Fröberg from Lund University are acknowledged for useful discussions and for growing and providing the nanowires used for this study, respectively.

References

[1] K A Dick 2008 *Progress in Crystal Growth and Characterization of Materials* **54**, 138
[2] K A Dick, K Deppert, T Mårtensson, B Mandl, L Samuelson and W Seifert 2005 *Nano Letters* **5**, 761
[3] A I Persson, M W Larsson, S Stenström, B J Ohlsson, L Samuelson and L R Wallenberg 2004 *Nature Materials* **3**, 677
[4] E Hilner, U Håkanson, L E Fröberg, M Karlsson, P Kratzer, E Lundgren, L Samuelson and A Mikkelsen 2008 *Nano Letters* **8**, 3978
[5] C C Ahn and O L Krivanek 1983 *EELS Atlas* (Gatan)