Quantum Noise Effects in e-Beam Lithography and Metrology

Verduin, Thomas

DOI
10.4233/uuid:f214f594-a21f-4318-9f29-9776d60ab06c

Publication date
2017

Document Version
Final published version

Citation (APA)
Verduin, T. (2017). *Quantum Noise Effects in e-Beam Lithography and Metrology*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:f214f594-a21f-4318-9f29-9776d60ab06c

Quantum Noise Effects

in e-Beam

Lithography and Metrology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 18 januari 2017 om 10:00 uur

door

Thomas VERDUIN

Master of Science in Natuurkunde en Meteorologie & Fysische Oceanologie,
geboren te Rotterdam

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. P. Kruit
copromotor: dr. C.W. Hagen

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. P. Kruit,	Technische Universiteit Delft
Dr. C.W. Hagen,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. Y.M. Blanter,	Technische Universiteit Delft
Prof. dr. W.M.J.M. Coene,	Technische Universiteit Delft
Prof. dr. ir. H.X. Lin,	Technische Universiteit Delft
Prof. dr. F. Salvat,	Universitat de Barcelona
Overige leden:	
Dr. ir. E.R. Kieft,	FEI Company

Printed by: Sieca Repro, Delft

ISBN 978-94-6186-782-7

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

“The purpose of computing...

is insight, not numbers.”

— Richard W. Hamming

Contents

Samenvatting

Geïntegreerde schakelingen (ICs) worden op een wereldwijde schaal geproduceerd met nanolithografie. De halfgeleider industrie wordt gedreven door het handhaven van de wet van Moore bij het steeds kleiner maken van de elektronische schakelingen. Het verkleinen van de elektronische schakelingen kan echter de doorvoer van een lithografisch systeem reduceren, d.w.z. het aantal wafers dat per uur geproduceerd kan worden. De doorvoersnelheid is een belangrijke parameter omdat het de kosteneffectiviteit van de productie van geïntegreerde schakelingen bepaalt. Lithografische systemen zijn daarom geoptimaliseerd voor hoge doorvoersnelheden. De doorvoersnelheid kan op verschillende manieren worden verhoogd. In optische lithografie is bijvoorbeeld het gebruik van een krachtigere lichtbron een oplossing. In e-beam lithografie is een toename van het aantal parallelle bundels een oplossing. Nu is het zo dat in beide gevallen het gebruik van een gevoeliger chemisch-versterkt resist (engels: Chemically Amplified Resist, oftewel CAR) in een vermindering van de vereiste belichtingsdosis en derhalve een vermindering van de belichtingstijd van de wafers resulteert. Het lijkt voor de hand te liggen om de doorvoersnelheid te maximaliseren door het meest gevoelige chemisch-versterkte resist met de daarbij horende laagst mogelijke belichtingsdosis te kiezen. In deze limiet is echter een stijging van de ruwheid van een lijn (engels: Line Edge Roughness, oftewel LER) geconstateerd. Dit resulteert in onvoldoende controle over de karakteristieke dimensies (engels: Critical Dimension, oftewel CD) in de productie van elektronische schakelingen. De toename van de LER wordt hoofdzakelijk veroorzaakt door fundamentele kwantumruis effecten (ook wel bekend als hagelruis, engels: shotnoise). Deze shotnoise effecten compliceren de verkleining van de dimensies van de transistors en dat staat de vooruitgang van de toekomstige generaties ICs in de weg. Dit onderzoek heeft als doelstelling om de fundamentele grenzen van de kwantumruis effecten te onderzoeken en om de theoretische kennis voor de productie van kleinere ICs te verbeteren.

De resultaten van dit onderzoek kunnen als volgt worden samengevat,

1. Ontwikkeling van een nieuwe meettechniek voor de LER in tweedimensionale CD-SEM beelden (hoofdstuk 2).

2. Ontwikkeling van een snelle GPU-versnelde Monte-Carlo elektron-materie interactie simulator. Deze simulator is gebaseerd op de meest geavanceerde modellen uit de vaste stof fysica (hoofdstuk 3, 4 en 6).

3. Een uitgebreide toepassing van de GPU-versnelde simulator: De karakterisering van de ruwheid van driedimensionale patronen vanuit gesimuleerde CD-SEM beelden (hoofdstuk 5 en 6).

4. Ontwikkeling van een Monte-Carlo bouwwerk voor de ab-initio berekening van shotnoise effecten in e-beam direct-write lithografie (hoofdstuk 7).

Shotnoise effecten in lithografische systemen ontstaan door de fluctuaties in het aantal elektronen (of fotonen in optische lithografie) tijdens het belichtingsproces. Deze effecten zijn het beste waar te nemen in een reeks patronen van lijnen waarbij opeenvolgende lijnen aan een steeds kleinere (lithografische) belichtingsdosis worden blootgesteld. De shotnoise effecten worden vervolgens gemeten en afgebeeld door gebruik te maken van een CD-SEM. Van de verkregen beelden van de lijnen wordt vervolgens de LER bepaald. Wanneer we de belichtingsdosis verlagen, nemen de fluctuaties in het aantal elektronen toe (kwantum effect) en dat resulteert in een dramatische toename van de LER. Hoewel dergelijke metingen alom bekend zijn in de wetenschappelijk literatuur, is de methode van de LER bepaling allesbehalve triviaal te noemen. Zo zijn de nauwkeurigheid en betrouwbaarheid van de bestaande metingen op zijn minst twijfelachtig te noemen. De huidige bepaling van de LER stelt een limiet aan de maximaal toegestane beeldruis in de afbeelding van een patroon met een CD-SEM. Dit betekent dat de beelden bijvoorbeeld moeten worden verkregen met een hogere dosis elektronen. Een andere oplossing is om de afbeelding van de CD-SEM te filteren vóórdat de LER wordt bepaalt. Echter, door het

verhogen van de dosis lopen we het risico dat patronen beginnen te krimpen door de impact van de elektronenbundel. Dit betekent dat de patronen veranderen terwijl we aan het meten zijn. De meest voor de hand liggende oplossing is om de beeldruis te reduceren door het toepassen van een ruis-onderdrukkend filter vóórdat de LER bepaling begint. Dit laatste is dan ook de gebruikelijke manier in de halfgeleiderindustrie. Echter, we laten in dit onderzoek zien dat het gebruik van een dergelijk filter een systematisch afwijking introduceert in de LER bepaling. Onze eerste pogingen zijn gebaseerd op het optimaliseren van een elliptisch ruis-onderdrukkend filter voor experimentele CD-SEM beelden. Wat we zien is dat het effect van het filter in de transversale en longitudinale richtingen niet kunnen worden genegeerd, zelfs niet wanneer gekozen wordt voor een geoptimaliseerde sterkte van het filter. We hebben dit probleem aangepakt door op een nieuwe en speciale manier de LER te bepalen. Deze bepaling is zódanig, dat een ruis-onderdrukkend filter niet meer nodig is. Daartoe modelleren we de signaal-respons van de CD-SEM om de ruwheid van de lijn te bepalen. De signaal-respons wordt verkregen door een experimenteel CD-SEM beeld van een patroon met lijnen te integreren. Het signaal wordt vervolgens gefit met een niet-lineaire kleinste kwadraten methode tegen de ruwe (ongefilterde) amplitudes van het CD-SEM beeld. Hoewel de lokale rand verplaatsingen op zich al een directe meting van de ruwheid geven, wordt de LER het beste geanalyseerd door gebruik te maken van de de spectrale vermogensdichtheid (engels: Power Spectral Density, oftewel PSD). De ruwheid wordt vervolgens gekarakteriseerd door het model van Palasantzas met vier parameters (de LER, correlatie lengte, ruwheids-exponent en sterkte van het beeldruis) aan de verkregen PSD te fitten. Met deze filter-vrije methode kan de LER nauwkeurig en éénduidig bepaalt worden, zelfs uit de meeste ruizige CD-SEM beelden zonder (1) het risico van inkrimpen van de patronen en (2) zonder systematische afwijkingen die te wijten zijn aan het toepassen van een filter. In een theoretische studie tonen we aan dat de LER afgeleid kan worden door slechts twee elektronen per pixel gemiddeld te gebruiken in de belichting. Dit laatste komt neer op zo'n $10\,\mu C/cm^2$

en geeft zeer ruizige afbeeldingen. Om de statistische fout in de PSD te onderdrukken moeten vele lijnen geanalyseerd worden. Echter, we laten zien dat zelfs één enkel beeld met 12 lijnen al een redelijke schatting van de LER oplevert met een relatieve fout van slechts 10%. Het verder verhogen van de belichtingsdosis dan 20 elektronen per pixel resulteert in geen verdere verbetering in de bepaling van de LER. Tenslotte hebben we de methode toegepast op experimentele, zeer ruizige CD-SEM beelden. In het experiment hebben we een vast patroon van lijnen herhaaldelijk afgebeeld met de laagst mogelijke instelling van een CG4000 CD-SEM van Hitachi. De beelden zijn verkregen door een elektronenbundel te gebruiken met een voltage van 300 eV, stroom van 10 pA en hoogstens één integratie frame. De resulterende belichtingsdosis is gelijk aan ongeveer één elektron per vierkante nanometer. Alhoewel dit een extreem lage belichtingsdosis is, accumuleert deze dosis desalniettemin bij het nemen van opeenvolgende beelden. Het gevolg daarvan is dat het patroon bij opeenvolgende beelden steeds meer gaat inkrimpen. We hebben twee verschillende soorten resists geanalyseerd en gebruiken een multi-exponentieel model om de krimp te karakteriseren. Wat we hebben waargenomen is dat de meeste krimp gebeurt in de eerste frames van de beeldopname en is niet-lineair van aard. Bovendien constateren we dat de LER langzaam en geleidelijk minder wordt bij toenemende belichtingsdosis. We concluderen dat het gebruik van een lage voltage en een lage dosis in CD-SEM beeldanalyse een noodzakelijke voorwaarde is voor een betrouwbare karakterisering van de CD en de LER.

De ware grootte, vorm en kenmerken van de ruwheid worden niet volledig onderzocht in de analyse van tweedimensionale CD-SEM beelden. In werkelijkheid zijn de ruwe patronen complexe driedimensionale structuren. De karakterisering van de ruwheid van een lijn gaat dan over op de analyse van de ruwheid van de zijwand (engels: Sidewall Roughness, oftewel SWR). Deze ruwheid kan gemeten worden met bijvoorbeeld een atomic force microscoop (AFM). Echter, gezien het grote aantal wafers dat geproduceerd wordt in een typische productieomgeving, wordt de AFM niet gezien als geschikt instrument voor inspectie. Een ander probleem met de AFM betreft

de complexiteit van de meting waarbij de exacte vorm van de tip een cruciale rol speelt. Een andere mogelijkheid is om virtuele patronen van lijnen met ruwheid te creëren en de beeldopname van een CD-SEM te simuleren. Dergelijke numerieke studies zijn moeilijk uitvoerbaar, omdat de details van de ruwheid tot atomair niveau gedefinieerd moeten worden. Simulaties van CD-SEM beelden van patronen met ruwe lijnen zijn zeer tijdrovend en worden daarom feitelijk vermeden. De tijdspanne van een simulatie kan verkort worden door sterk vereenvoudigde fysische modellen te gebruiken. Het nadeel is dat er een grote onzekerheid zit in de geldigheid en nauwkeurigheid van de vereenvoudigde modellen. Onze zoektocht om de rekentijd van CD-SEM beelden te reduceren heeft ons doen overwegen of graphics processing units (GPUs) daar een rol in kunnen spelen. Het is ons gelukt om een geavanceerde Monte-Carlo simulator voor CD-SEM beelden te ontwikkelen. Onze simulator is, naar ons beste weten, de eerste en enige simulator voor CD-SEM beelden die volledig draait op een GPU. We hebben gebruik gemaakt van de meest geavanceerde modellen uit de vaste stof fysica: Mott-verstrooiing voor de elastische processen, diëlektrisch functie model voor de inelastische bulk en oppervlakte verstrooiingsprocessen, akoestische en optische fonon-verstrooiing, het opsluiten van elektronen in isolatoren, alsmede effecten op de grensvlakken van materialen door gebruik te maken van kwantummechanische berekeningen voor transmissie en reflectie van elektronen. De modellen in de GPU-versnelde simulator worden geverifieerd door meetbare uitkomsten te vergelijkingen met experiment, zoals de secundaire en backscatter elektronen-yield en spectra van het energieverlies van de elektronen. We hebben als toepassing voor de simulator gekozen voor het bepalen van de ruwheid uit gesimuleerde CD-SEM beelden van kunstmatig gecreëerde driedimensionale ruwe lijnen. De voorgestelde simulatie is buitengewoon complex en groot van opzet omdat het (1) noodzakelijk is om een groot aantal CD-SEM beelden te verwerven voor de statistiek, (2) vele parameters te variëren (karakteristieke dimensie, ruwheidsparameters, hoogte van het lijntje en de energie van de elektronenbundel) en (3) het lijntje (ter grootte van één micrometer) gedefinieerd moet worden met een

ruwheid op de schaal van nanometers. In eerste instantie voeren we de simulatie uit op een GeForce GTX480 van NVIDIA. De simulatie wordt tevens gedupliceerd op een CPU-gebaseerd programma, waarvoor wij een Intel Xeon X5650 hebben gebruikt. We concluderen dat, buiten de willekeurige statistieken in de simulatie, geen verschil kan worden aangewezen tussen de CPU- en GPU-gesimuleerde resultaten. Dit kan echter niet gezegd worden over de rekentijd. We hebben vastgesteld dat de GTX480 de CD-SEM beelden (afhankelijk van de energie van de primaire elektronenbundel) 387 tot 894 maal sneller dan een enkele thread op een Intel X5650 CPU genereert. Deze prestatieverbetering wordt als volgt bereikt. Allereerst hebben wij een speciale manier om de geometrie te definieëren. Het vacum, de sample en de vorm van de detectoren worden gedefinieerd in een speciale driedimensionale octree van kubusvormige cellen. De detectoren en de grensvlakken van de materialen wordt opgedeeld in driehoeken. Elke cel in de octree heeft een lijst van precies die driehoeken die overlappen met deze specifieke cel. Dit resulteert in een aanzienlijke vermindering in het bepalen van het aantal elektron-driehoek doorsnijdingen tijdens de simulatie van een complexe geometrie. Ten tweede, elektronen met soortgelijke gebeurtenissen worden gegroepeerd met behulp van een parallelle radix sorteermethode, die eveneens op een GPU draait. Door het sorteren van de elektronen naar gebeurtenis, verminderen we het risico van instructie divergentie binnen de rekeneenheden van de GPU. Tot slot verzadigen we de GPU door voldoende elektronen parallel te traceren. In de praktijk heeft de GTX480 niet voldoende geheugen om alle elektronen van de CD-SEM tegelijk te verwerken. In plaats daarvan moet een batch proces gebruikt worden om de GPU met regelmaat te voorzien van primaire elektronen. De versnellingsfactor in het verkrijgen van de gesimuleerde CD-SEM beelden betekent dat bijvoorbeeld vele parameters van een ruw patronen (karakteristieke dimensie, ruwheid, hoogte, ...) kunnen worden gevarieerd in een acceptabele tijdspanne. Bovendien kunnen ten bate van de statistieken een groot aantal CD-SEM beelden gesimuleerd worden.

Als laatste onderzoeken we hoe de vorming van SWR tot stand komt. In de nanolithografie wordt eerst een resistlaag aangebracht op een substraat. De resistlaag wordt dan optisch (met fotonen) of met een elektronenbundel belicht. In beide gevallen resulteert de blootstelling aan zowel fotonen als aan elektronen in het vrijkomen van zuren in de resistlaag. Daarop volgt een fase (de zogenaamde post-exposure bake, owel PEB) waarbij de zuren gaan diffunderen en maken of breken daarbij moleculaire verbindingen. Door het maken of breken van verbindingen worden delen van het resist oplosbaar of juist onoplosbaar. Het doel van de computationele lithografie is het voorspellen van de effecten zodat het lithografische proces kan worden geoptimaliseerd. Eén van de belangrijkste uitdagingen van dit onderzoek is om te begrijpen hoe de shotnoise effecten vanaf de eerste blootstelling de uiteindelijke SWR beïnvloeden. We hebben, voor het bestuderen van de vorming van shotnoise-geïnduceerde SWR, een zeer snel 3D Monte-Carlo bouwwerk ontwikkeld. Als voorbeeld geven we een berekening waarbij een resistlaag van 100 nm dik bovenop een oneindig dikke laag silicium wordt blootgesteld aan een elektronenbundel. We gebruiken de eerder genoemde GPU-versnelde Monte-Carlo simulator voor elektron-materie interactie, echter nu ten behoeve van lithografie. Een patroon van een geïsoleerde lijn wordt in de resistlaag geschreven door het scannen van de elektronenbundel (met een energie van 20 keV) over een gebied van 32 nm$\times$1 µm (breedte maal lengte). Tijdens de belichting gebruiken we een spotgrootte van 20 nm, een stapgrootte van 4 nm en een Poisson-verdeelde belichtingsdosis van 80 µC/cm^2, 60 µC/cm^2 en 40 µC/cm^2. Tijdens de belichting registreren we de locaties van de inelastische gebeurtenissen in de resistlaag. De verdeling van de vrijgekomen zuren wordt bepaald aan de hand van de vereenvoudigde aanname dat elke inelastische gebeurtenis leidt tot het vrijkomen van een zuur. We maken vervolgens een driedimensionaal beeld van de (on)oplosbaarheid van de resistlaag in een gebied van van 128 nm(256px) breed, 800 nm(1024px) lang en 100 nm(128px) hoog. De (on)oplosbaarheid wordt verkregen door het optellen van de bijdragen van alle zuren in de resistlaag voor elke voxel. We veronderstellen

dat het proces van diffunderen van de zuren in de resistlaag een Gaussische kansverdeling volgt met $\sigma_{x,y,z} = r_D = 5\,\mathrm{nm}$. Vervolgens bepalen we het grensvlak tussen oplosbaar en onoplosbaar door een drempelwaarde te kiezen. De gemiddelde LER wordt verkregen door het berekenen van de standaardafwijking van de linker en rechter begrenzing van de xy-segmenten. Hierin worden alle segmenten, van de top van de resistlaag tot op de bodem van het substraat, de gemiddelde LER als functie van de diepte van de resistlaag verkregen. Bij het verlagen van de belichtingsdosis zien we de shotnoise effecten toenemen. Tevens zien we een versterkt effect van shotnoise nabij de grensvlakken met het vacuüm en het substraat. Een oorzaak daarvoor is terug te vinden in het daadwerkelijke aantal zuren dat effectief een bijdrage kan leveren aan de oplosbaarheid. Het aantal zuren is vanwege de elektronen verstrooiing minder in de buurt van een grensvlak dan in de bulk van de resistlaag. Een andere oorzaak komt voort uit het feit dat er geen zuren kunnen worden vrijgemaakt aan de vacuümzijde noch aan de substraatzijde. De resultaten van dit onderzoek geven een veelbelovend vooruitzicht. Niet alleen voor de inspectie, maar ook voor het optimaliseren van nanolithografische processen.

Summary

Integrated circuits on a global scale are produced with nanolithography. The industry is driven by maintaining Moore's law as chip feature sizes are decreasing from generation to generation. The downscaling may reduce the throughput of a lithographic system, i.e. the number of wafers produced per hour, which in turn determines the cost effectiveness of the production of integrated circuits. As a consequence, lithographic systems are optimized for high throughput, which can be increased in several ways. In optical lithography, for example, a more powerful source is one solution. In electron-beam lithography, an increase of the number of parallel beams is another solution. In either case, the use of a more sensitive chemically amplified resist (CAR) results in a reduction of the required illumination dose, and hence a reduction of the exposure time of the wafers. In order to maximize throughput, it is tempting to choose the most sensitive chemically amplified resist with the lowest possible illumination dose. In that limit, however, an increase of line edge roughness (LER), and hence an insufficient control of critical dimension (CD) is observed. This increase of LER is primarily caused by fundamental quantum noise (shotnoise) effects and becomes the dominant mechanism in the formation of LER. This, in turn, complicates the downscaling of transistor dimension. The subject of this PhD is to investigate the fundamental limits of quantum noise effects and strengthen the theoretical basis to improve the production of smaller integrated circuits. The achievements of this work can be summarized as follows,

1. Development of a novel measurement technique for LER from noisy two-dimensional scanning electron microscopy images (Chapter 2).

2. Development of a fast GPU accelerated Monte-Carlo electron-matter interaction simulator based on the most advanced models in solid state physics (Chapter 3, 4 and 6).

3. Application for metrology using the GPU simulator: roughness characterization of 3D features from simulated 2D SEM images (Chapter 5 and 6).

4. Development of a Monte-Carlo framework for the ab-initio calculation of quantum noise effects in electron beam direct-write lithography (Chapter 7).

The root cause for the emergence of quantum noise effects in lithographic systems is the number statistics of electrons (or photons in optical lithography) during exposure. The effects are best seen in consecutive acquisitions of patterns of lines and spaces obtained with decreasing (lithographic) exposure dose. The shotnoise effects are then measured from top-down 2D images acquired with an industry standard critical dimension scanning electron microscope (CD-SEM). From the images of lines and spaces, the LER is determined, which increases dramatically for decreasing exposure dose. Although such measurements are well known, the actual determination of LER is anything but trivial and the accuracy of existing measurements is questionable. To begin with, the determination of LER involves edge detection which in turn introduces a limit to the maximum allowable image noise. This means that images must be acquired with a higher electron dose or the CD-SEM image is filtered before edge detection. However, by increasing the imaging dose, there is the risk of resist shrinkage due to the impact of the electron beam, which in turn affects the quantity being measured. The most obvious solution is to reduce the image noise by applying a noise-suppressing filter to the image before edge detection. Although this is common practice in semiconductor industry, the use of such a filter complicates the determination of LER by introducing a bias to the measurement. Our first attempts are based on optimized elliptic filtering of noisy experimental SEM images, where we use threshold-based peak detection to determine the edge displacements. What is observed is that the effect of transversal and longitudinal filtering cannot be ignored, even when considering an optimized filter strength. We have addressed the issue of measurement by introducing a special method which avoids the use

of a noise-suppressing filter. We have exploited the signal response function of the CD-SEM in order to determine the edge displacements along the length of a line. The idea is to fit the SEM signal profile, which is obtained by integrating an experimental SEM image of line edges in the direction of the edges, against the raw (unfiltered) amplitude of edges in the SEM image. Although the edge displacements already give a direct measurement of the roughness, LER is best analyzed by the Power Spectral Density (PSD). The roughness is characterized by fitting the model of Palasantzas with four parameters (the LER, correlation length, roughness exponent and image noise) to the PSD of the edge displacements. With edge based fitting, LER can be determined more accurately from very noisy images without increasing the risk of resist shrinkage and without biasing due to image filtering. In a simulation study we show that the LER can still be determined from very noisy images with only two electrons per pixel on average $10\,\mu C/cm^2$. The PSDs are generally averaged over many line edges to reduce the statistical error. However, even a single image with 12 line edges, produces an estimation of the LER with a relative error of only 10%. Furthermore, increasing the dose beyond 20 electrons per pixel does not significantly improve the LER determination. Finally, we have applied the method to experimental CD-SEM images. We have accomplished an experiment where a pattern of lines and spaces is repeatedly imaged with the lowest possible settings of a CG4000 CD-SEM from Hitachi: A beam of 300 eV at 10 pA with a single integration frame. The resulting dose equals one electron per square nanometer. For each consecutive image, the net accumulated dose increases, which causes the resist to shrink. We have analyzed two different types of resist and used a multi-exponential decay model to characterize the shrinkage. What is observed is that most of the shrinkage happens in the very first few frames and is non-linear of nature. Furthermore, we have observed that the LER decreases for accumulating dose. We conclude that low voltage and low dose CD-SEM image analysis is a necessary prerequisite for reliable characterization of the CD and the LER.

The true size, shape and roughness characteristics of resist features are not fully investigated in the analysis of 2D SEM images. In reality, rough resist features are complex 3D structures. The characterization of roughness of resist features naturally extends to the analysis of sidewall roughness (SWR) which can be measured, for instance, by using an atomic force microscope (AFM). However, in view of the large volume of wafers being produced in a typical production line, the AFM is not considered as a suitable metrology tool. Another problem with the AFM involves the complexity of the measurement, in which the exact shape of the tip plays a crucial role. Another way in which SWR can be studied is by creating virtual rough samples of patterns of lines and spaces and simulate the image acquisition using a CD-SEM image simulator. Numerical studies on sidewall roughness are difficult to perform because the details of the roughness are defined at the (sub) nanometer scale. Studies involving the discrete modeling of the roughness on the sidewalls are therefore subject to time consuming simulations, and especially in the case of Monte-Carlo simulations. Because of the latter, detailed Monte-Carlo simulations are usually avoided and simulators with simplified physical models are used instead. Although simplified physical models do not have the performance issue, the validity and accuracy remains an open question. Our quest in reducing the computation time of SEM image simulations has led us to investigate the use of graphics processing units (GPUs). We have created a rigorous Monte-Carlo simulation program for CD-SEM images, which to our best knowledge, is the first and only simulator for CD-SEM images which runs entirely on a graphics processing unit (GPU). We have employed the following models in the simulator: Mott scattering for elastic scattering, dielectric function theory for inelastic bulk and surface scattering, acoustic/optical phonon scattering, trapping of electrons for insulators and interface effects such as quantum mechanical transmission/reflection. The GPU simulator is verified by making comparisons to experimental secondary/backscatter yields and EELS spectra. As a case study, we consider the determination of SWR from simulated 2D images of 3D rough features. This is a noto-

riously complex and extensive simulation as it requires (1) the acquisition of many images for statistics, (2) variation of many parameters (critical dimension, roughness parameters, feature height and beam energy) and (3) spatially resolved details of micrometer-sized features with roughness defined at the nanometer scale. At first we run the simulation on a GeForce GTX480 from NVIDIA. The very same simulation is duplicated on a CPU-based program, for which we have used an Intel Xeon X5650. We conclude that, apart from statistics in the simulation, no difference is found between the CPU and GPU simulated results. This, however, cannot be said about the total simulation time. We have determined that the GTX480 generates the images (depending on the primary electron energy) 387 to 894 times faster than a single threaded Intel X5650 CPU. The performance increase is achieved as follows. First of all, we have used a special implementation for the geometry. The geometry includes the vacuum, sample, electrons and shape of the detectors and is subdivided into a special 3D octree of cuboid cells. The shape of the detectors and the material boundary, which defines the interface between two different materials, is triangulated. Every cell in the octree has a list of precisely those triangles which overlap with this particular cell. This gives a significant reduction on the number of electron-triangle intersections during the simulation of a complex geometry. Second, electrons with similar events are grouped by using a parallel radix sorting method, which also runs on the GPU. By sorting the electrons by event, we reduce the risk of instruction divergence within a warp of 32 threads. Finally, we saturate the GPU by using sufficient electrons per iteration to track in parallel. In practice, the GTX480 does not have sufficient memory to process all electrons of the SEM image at once. Instead, a batch process is used to push primary electrons at regular intervals. The speedup enables the fast acquisition of simulated SEM images for metrology. This means, for example, that many parameters of a rough feature (critical dimension, roughness, height, and so on) can be simulated in a reasonable amount of time. Moreover, because of the speedup, statistics can be generated as well by simulating a multitude of SEM images.

In the final piece of the puzzle we investigate the formation of SWR. In nanolithography, a resist layer is spin coated on a wafer with a substrate. The resist layer is then exposed optically or with a beam of electrons. In any case, the effect of the exposure of the resist layer results in the release of acids. In the post-exposure baking phase, the acids start diffuse and break or cross-link molecular bonds which either become soluble or insoluble. The subject of computational lithography is to predict the effects in detail, such that the lithographic process can be optimized for throughput and minimizing the errors to within acceptable tolerances on feature dimensions for example. One of the major challenges is to investigate how quantum noise effects, starting from the initial exposure, affect the resulting feature. For studying the formation of shotnoise induced SWR, we have developed a fast 3D Monte-Carlo framework. The calculation outline is demonstrated by an exposure of a 100 nm thick layer of chemically amplified resist (CAR) on top of an infinitely thick silicon substrate. We use the GPU accelerated Monte-Carlo electron-matter interaction simulator for the purpose of lithography. A pattern of an isolated line is written into the resist layer by scanning a beam with 20 keV electrons over an area of 32 nm×1 µm (width times length). During the exposure, we use a spot-size of 20 nm, beam step-size of 4 nm and a Poisson distributed exposure dose of 80 µC/cm^2, 60 µC/cm^2 and 40 µC/cm^2. During the exposure of the sample, we record the locations of the inelastic events within the resist layer. The distribution of released acids is determined under the simplified assumption that every inelastic event corresponds to a release. We now construct a three dimensional image of the (in)solubility of the resist layer within a cuboid of 128 nm(256px) wide, 800 nm(1024px) in length and 100 nm(128px) in height. The (in)solubility is obtained by summing the contribution of all acids to every voxel in the 3D image, where we have used a 3D Gaussian with $\sigma_{x,y,z} = r_D = 5\,\text{nm}$ for the diffusion of the acid. The boundary between exposed and unexposed resist is determined by a threshold. The resulting image of the (in)solubility is analyzed in different ways by considering slices and three dimensional views of the boundary.

The average LER is obtained by calculating the standard deviation of the left and right boundary from xy-slices. By considering all slices, ranging from the top of the resist layer to the bottom of the substrate, the average LER as a function of the depth from the top surface of the resist layer is obtained. Shotnoise effects are observed as we decrease the exposure dose. An increased effect of shotnoise is observed near the vacuum and substrate interface. One contribution relates to the actual number of acids, which due to the scattering is less near the interface than away from the interface. Another contribution stems from the fact that no acids are found on the vacuum side nor on the substrate side. The results of this study provide a promising prospect. Not only for the inspection, but also for the optimization of nano-lithographic processes.

Chapter 1

Introduction

Integrated circuits on a global scale are produced with nanolithography. The industry is driven by maintaining Moore's law as chip feature sizes are decreasing from generation to generation. The downscaling reduces the throughput of a lithographic system, i.e. the number of wafers produced per hour, which in turn determines the cost effectiveness of the production of integrated circuits. As a consequence, lithographic systems are optimized for high throughput, which can be increased in several ways. In optical lithography, for example, a more powerful source is one solution. In electron-beam lithography, an increase of the number of parallel beams is another solution. In either case, the use of a more sensitive chemically amplified resist (CAR) results in a reduction of the required illumination dose, and hence a reduction of the exposure time of the wafers. In order to maximize throughput, it is tempting to choose the most sensitive CAR with the lowest possible illumination dose. In that limit, however, an increase of line edge roughness (LER), and hence an insufficient control of critical dimension (CD) is observed.[1] This increase of LER is primarily caused by fundamental quantum noise (shotnoise) effects and becomes the dominant mechanism in the formation of LER.[2, 3, 4, 5, 6, 7, 8] This, in turn, complicates the downscaling of transistor dimension. The subject of this study is to investigate the fundamental limits of quantum noise effects and strengthen the theoretical basis to improve the production of smaller integrated circuits.

Quantum noise effects are best seen in patterns of lines and spaces, obtained from exposing resist with decreasing dose in consecutive experiments. The shotnoise effects are then measured from top-down two-dimensional images acquired with an industry standard critical dimension scanning electron microscope (CD-SEM). From the images of lines and spaces, the LER is determined, which increases dramatically for decreasing exposure dose. Although such measurements are well known,[9, 10, 11] the actual determination of LER is anything but trivial. The accuracy of existing measurements is questionable. Typically in LER analysis, the fluctuations in edge displacements are determined using a threshold based peak detector, a Canny-edge detection filter or by a homemade edge detection algorithm, see for example Ref. 12. Although the edge displacements already give a direct measurement of the roughness, LER is best analyzed by the Power Spectral Density (PSD).[13, 14, 15, 16, 17] There are a couple of problems related to this type of determination. First of all, there are statistical and systematic errors because the actual PSD is approximated by sampling the edge displacements of a pattern with a finite number of measurement intervals. The statistical errors are for example described in Refs. 15, 16. The systematic errors have recently been studied in Ref. 17. There is also the problem of shrinkage, where the act of measuring the edge displacements by irradiating with an e-beam induces changes in the pattern, see for example Refs. 18, 19, 16, 20. Finally, there is the problem of biasing due to the use of a filter prior to the edge detection in noisy CD-SEM images, see for example Refs. 12, 18, 21, 15, 22, 23.

Another problem is that the true size, shape and roughness characteristics of resist features are not fully investigated in the analysis of two-dimensional scanning electron microscopy images. In reality, rough resist features are complex three-dimensional structures. The characterization of roughness of resist features naturally extends to the analysis of sidewall roughness (SWR) which can be measured, for instance, by using an atomic force microscope. However, in view of the large volume of wafers being produced in a typical production line, the atomic force microscope (AFM) is not considered as a suitable metrology tool. Another problem

with the AFM involves the complexity of the measurement, in which the exact shape of the tip plays a crucial role. Another possibility is to create virtual rough samples of patterns of lines and spaces and simulate the image acquisition using a CD-SEM image simulator. Numerical studies on SWR are difficult to perform because the details of the roughness are defined at the (sub) nanometer scale. Studies involving the discrete modeling of the roughness on the sidewalls are therefore subject to time consuming simulations, and especially in the case of Monte-Carlo simulations. Because of the latter, detailed Monte-Carlo simulations are avoided and simulators with simplified physical models are used instead. Although simplified physical models do not have the performance issue, the validity and accuracy remains an open question. The studies that we found in literature indicate that the true SWR is larger than the measured LER in a top-down CD-SEM image.[24, 25] We have a few remarks with respect to these studies. In the study of Li et al., the focus is on pure poly crystalline silicon lines with a Gaussian roughness model for the sidewalls.[24] The roughness of a resist feature, however, is typically characterized by more than just the standard deviation of a Gaussian-like distribution and involves additional parameters, such as correlation length and a roughness exponent.[26, 16, 27, 28] What happens, for example, to the measured LER when the correlation length of the SWR changes? The study of Lawson et al. is different for two reasons.[25] In the first place, Lawson et al. have used a more sophisticated model for the roughness of the sidewalls.[29] Unfortunately, the relation of the roughness parameters to correlation length and roughness exponent are not explicitly mentioned nor are the roughness parameters varied to study its influence on the LER. Second, the lines used in the study of Lawson et al. are made of pure poly-methyl methacrylate (PMMA) coated on a pure silicon substrate. Not only is PMMA a different material, it also comes, contrary to pure silicon lines, with a risk of resist shrinkage caused by the electron beam.[16] Although it is not addressed in the work of Lawson et al., this risk can be reduced, for example, by lowering the beam voltage and total electron dose, i.e. reduce the number of integration frames.[16, 28]

The final piece of the puzzle is to determine the formation of SWR. To that order, let us briefly discuss the process of nanolithography. In nanolithography, a resist layer is spin coated on a wafer with a substrate. The resist layer is then exposed optically or with a beam of electrons. In the case of a CAR, the exposure results in the release of acids of in the resist layer. In the post-exposure baking (PEB) phase, the acids start to diffuse and break or cross-link molecular bonds which either become soluble or insoluble. The subject of computational lithography is to predict the effects in detail, such that the lithographic process, for example, can be optimized for throughput while minimizing the errors on feature dimensions to acceptable values. One of the major challenges of this study is to investigate how quantum noise effects, starting from the initial exposure, affect the resulting feature.

This thesis is organized into seven chapters, which all can be read independently, as follows,

- Development of a novel measurement technique for quantum noise effects from two-dimensional scanning electron microscopy images (Chapter 2).
- Development of a fast GPU Monte-Carlo electron-matter interaction simulator based on the most advanced models in solid state physics (Chapter 3 and 4).
- Investigation of roughness characterization of 3D features from simulated 2D SEM images (Chapter 5 and 6).
- Development of a Monte-Carlo framework for the ab-initio calculation of quantum noise effects in e-beam direct write lithography (Chapter 7).

Every chapter of the thesis includes a brief description of the content and related publications, followed by an introduction and ends with a conclusion.

Chapter 2

Dimensional metrology of line edges

The root cause for the emergence of quantum noise effects in lithographic systems is the number statistics of electrons (or photons in optical lithography) during exposure. The effects are best seen in consecutive acquisitions of patterns of lines and spaces obtained with decreasing (lithographic) exposure dose. The shotnoise effects are then measured from top-down two-dimensional images acquired with an industry standard critical dimension scanning electron microscope (CD-SEM). From the images of lines and spaces, the line edge roughness (LER) is determined, which increases dramatically for decreasing exposure dose. Although such measurements are well known, the actual determination of LER is anything but trivial and the accuracy of existing measurements is questionable. To begin with, the determination of LER involves edge detection which in turn introduces a limit to the maximum allowable image noise. This means that images must be acquired with a higher electron dose or the CD-SEM image is filtered before edge detection. However, by increasing the imaging dose, there is the risk of resist shrinkage due to the impact of the electron beam, which in turn affects the quantity being measured. The most obvious solution is to reduce the image noise by applying a noise-suppressing filter to the image before edge detection. Although this is common practice in semiconductor

industry, the use of such a filter complicates the determination of LER by introducing a bias to the measurement.

In this chapter we address the issue of measurement by introducing a special method which avoids the use of a noise-suppressing filter. The approach is unique as the signal response function of the CD-SEM is exploited in order to determine the edge displacements along the length of a line. As a result, LER can be determined more accurately from very noisy images without increasing the risk of resist shrinkage and without biasing due to image filtering.

The content of this chapter is an updated version of the article we have published in the Journal of Micro/Nanolithography, MEMS, and MOEMS: *Verduin, T., Kruit, P., and Hagen, C. W., "Determination of line edge roughness in low dose top-down scanning electron microscopy images," Journal of Micro/Nanolithography, MEMS,and MOEMS 13, 033009 (2014).* This work was presented at the SPIE 2014 conference in San Jose and was rewarded with the Karel Urbanek best student paper award.

2.1 Introduction

The determination of Line Edge Roughness (LER) becomes increasingly important as the semiconductor devices decrease in dimensions.[9, 10, 11] This results in smaller tolerances on LER determination and as a consequence, the metrology becomes more critical. There are two classes of LER metrology. There is on-line metrology, which is typically performed in SEMs combined with dedicated (proprietary) software for LER analysis. Another class is off-line metrology and only deals with the image analysis. The latter is used for instance for resist characterization. Typically in off-line LER analysis, the fluctuations in edge displacements are determined using a threshold based peak detector, a Canny-edge detection filter or by a home-made edge detection algorithm, see for example Ref. 12. Although the edge displacements already give a direct measurement of the roughness, LER is best analyzed by the Power Spectral Density (PSD).[13, 14, 15, 16, 17] There are a couple of problems related to this type of determination. First of all,

there are statistical and systematic errors because the actual PSD is approximated by sampling the edge displacements of a pattern with a finite number of measurement intervals. The statistical errors are for example described in Refs. 15, 16. The systematic errors have recently been studied in Ref. 17. There is also the problem of shrinkage, where the act of measuring the edge displacements by irradiating with an e-beam induces changes in the pattern, see for example Refs. 18, 19, 16, 20. Finally, there is the problem of biasing due to the use of a filter prior to the edge detection in noisy SEM images, see for example Refs. 12, 18, 21, 15, 22, 23. In this chapter we focus on off-line metrology and investigate how much image noise is acceptable by decreasing the dose in simulated top-down SEM-like images. The determination of LER is a delicate issue and we ask the following questions: Can we reduce the influence of a filter to negligible levels by carefully optimizing the filter strength? Is it possible to avoid filtering in low dose images at all? How many low dose images of the line edges are required for estimating LER?

2.2 Line edge determination

Let us begin with a typical example of a top-down SEM image of line edges, which is given in Fig. 2.1. This image was recorded by J. Jussot from CNRS-LTM/CEA-LETI in 2012 using a CD-SEM from Hitachi. The properties of the resist are unknown to us due to disclosure restrictions. These resist properties, however, are of no further interest to us. In Fig. 2.2a we show the amplitude of a single horizontal scan-line taken from the center of Fig. 2.1. The noise in Fig. 2.2a is due to the pixel noise and corresponds to 16 integration frames. The number of integration frames should be kept as low as possible to reduce the effect of resist shrinkage.[18, 19, 16, 20] However, detection algorithms, such as threshold based detectors or Canny-edge detection filters, often do not find the edge or find too many edges in such noisy data. Working with low noise images has two problems: they take a long time to accumulate and there is a risk of resist shrinkage. An obvious way to reduce the noise is to apply a filter to the recorded

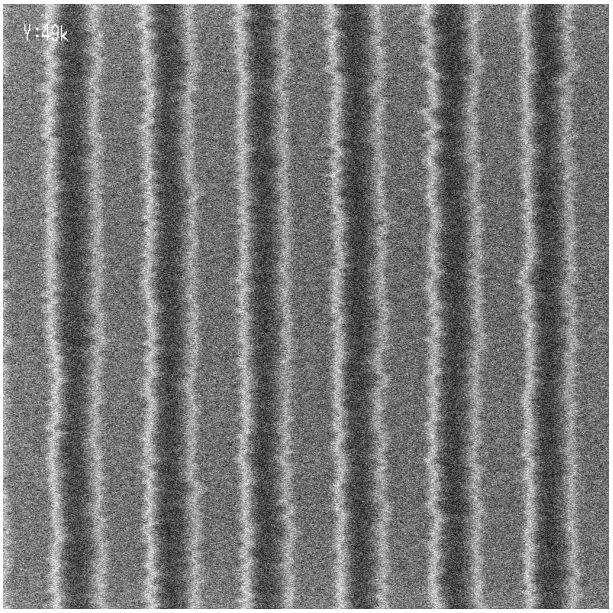


Figure 2.1: A typical top-down SEM image of line edges, which was recorded by J. Jussot from CNRS-LTM/CEA-LETI in 2012 using a CD-SEM from Hitachi. This image is recorded with a probe current of 6.0 pA at an acceleration voltage of 500 V. The number of integration frames equals 16. The magnification is such that the field size is 2.76 µm in length (1024px) and 450 nm wide (1024px). The resulting pixel size is approximately 2.7 nm × 0.44 nm.

image. The result of applying a symmetric Gaussian filter on the amplitude is shown in Fig. 2.2b. The edge displacements can be determined by a threshold detector, if the noise in the amplitude is sufficiently reduced.

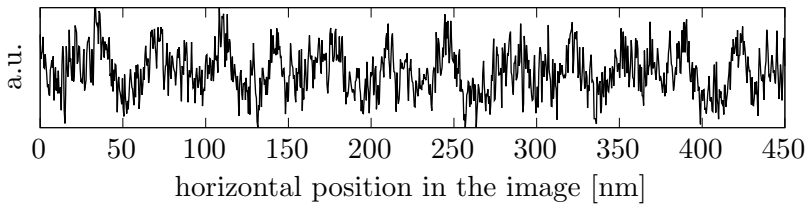
The discrete PSD of the edge displacements is obtained by calculating the complex amplitude of the Fourier coefficients,

$$P_n = \frac{L}{2\pi} |F_n|^2 \tag{2.1}$$

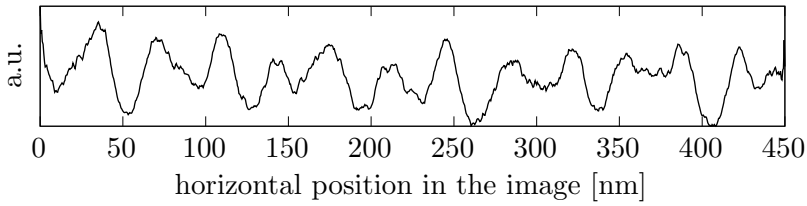
where the discrete Fourier transform is determined as,

$$F_n = \frac{1}{N} \sum_{j=0}^{N-1} (x_j - \langle x \rangle_N) \exp(-ik_n j \Delta y) \tag{2.2}$$

where N is the number of sampled edge displacements, x_j is the displacement of the j^{th} edge position, $\langle x \rangle_N$ is the mean position of the edge, Δy is



(a)



(b)

Figure 2.2: The amplitude of a single horizontal scan-line taken from the center of Fig. 2.1. The signal in Fig. 2.2a represents the amplitude directly from the image without any post-processing. This raw signal is too noisy for threshold based edge detection. However, the noise of this signal can be reduced by applying a symmetric Gaussian filter to the image and the result on the amplitude is demonstrated in Fig. 2.2b.

the measurement interval and k_n is the discrete wave number,

$$k_n = \frac{2\pi n}{L} \tag{2.3}$$

where $n = 0, 1, 2, \ldots, N-1$ and L is the length of the edge. The variance of the edge displacements is related to the PSD by,

$$\sigma^2 = \frac{2\pi}{L} \sum_{n=0}^{N-1} P_n \tag{2.4}$$

which is Parseval's relation. The discrete PSD given by Eq. 2.1 is only an approximation to the actual spectrum of a quasi-infinite long line and the finite line length L is a source of statistical noise, see for example Ref. 15, 16. The statistical noise in the discrete PSD can be reduced by averaging over many line edges,

$$P_n^\star = \frac{L}{2\pi} \left\langle |F_n|^2 \right\rangle_{N^\star} \tag{2.5}$$

where $N^\star$ counts the number of line edges over which the PSD is averaged. This averaging is possible if the length of the line edge is much longer than the correlation length of the edge fluctuations.

We calculate the discrete PSD of the line edges shown in Fig. 2.1 and, in order to reduce the noise in the power spectrum, we average the PSDs of the 50 recorded SEM images of line edges of the same kind. Every individual SEM image is at first convoluted with an elliptic Gaussian filter,

$$\widetilde{I}(x,y) = I(x,y) \star G(x,y) = I(x,y) \star \frac{\exp\left(-\frac{x^2}{2\sigma_{\mathrm{T}}^2} - \frac{y^2}{2\sigma_{\mathrm{L}}^2}\right)}{2\pi\sigma_{\mathrm{T}}\sigma_{\mathrm{L}}} \tag{2.6}$$

where $\widetilde{I}$ is the filtered SEM image, I the original SEM image and G the elliptic filter. The strength of the elliptic filter in the transverse direction to the edges is σ_{T} and the strength in the longitudinal direction is σ_{L}. The edge displacements are then determined using a threshold based edge detection algorithm. In this method, we identify an edge whenever the normalized amplitude of a peak is larger than 0.3 with respect to the noise level. This threshold is obtained empirically by analyzing many images. The choice of a particular threshold level is subjective and we will discuss the effect of this on our results later. The directional strengths of the filter are optimized on a per image basis and is such that every scan-line of a SEM image produces the real number of edges in the image. In other words, if we decrease the strength of the elliptic filter even further, then we will count more or fewer edges than physically present in the pattern. This happens because there are peaks within the signal (see Fig. 2.2a) which are not related to the edges but to the image noise. As we decrease the strength of the elliptic filter, the amplitudes of the peaks from the image noise increase and at some point we can no longer distinguish the peak due to an edge from the noise. It is at this point that the threshold method of determination becomes indecisive: it can no longer decide which peak is an edge and which peak is not. Suppose, for a particular SEM image of line edges, that we check for a range of longitudinal and transversal filter strengths whether the edge detection method produces the correct number of edges for all scan-lines. There exists a subdomain where the edge detection method produces the

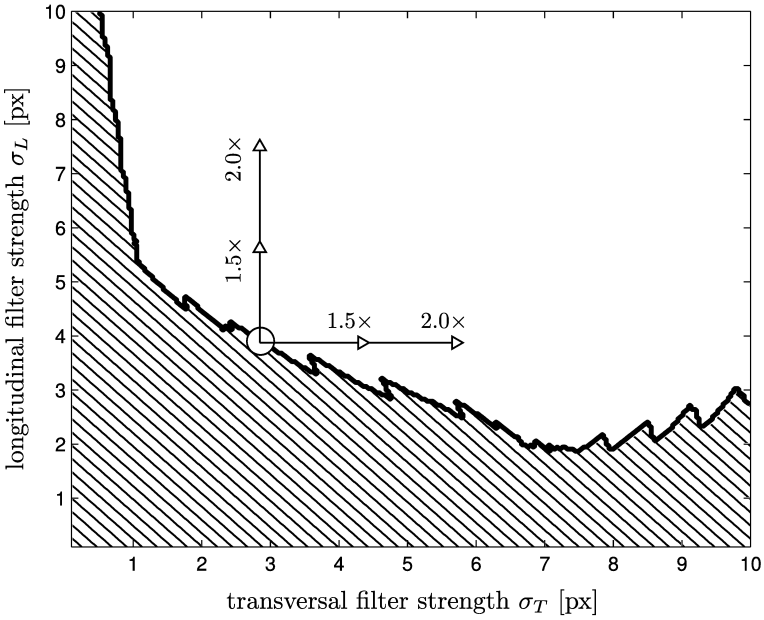


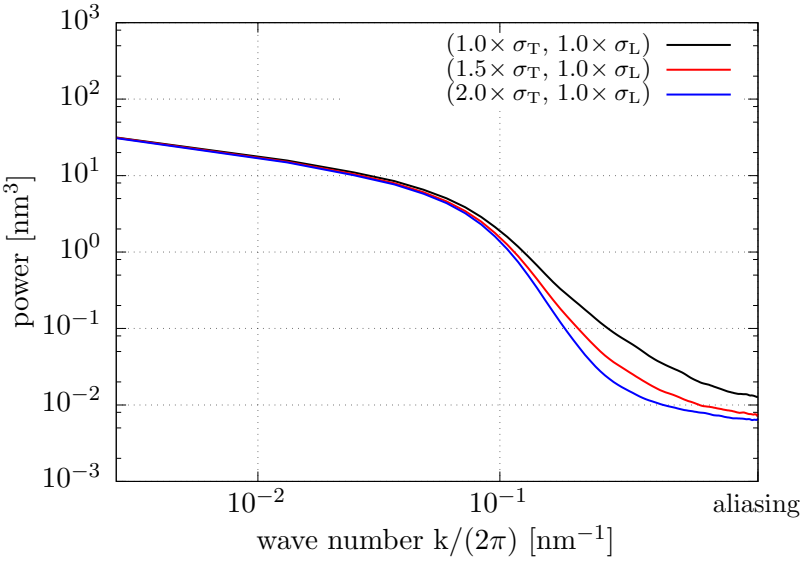
Figure 2.3: The effect of filtering on the number of detected edges in Fig. 2.1. The longitudinal filter strength σ_L is shown on the vertical axis and the transversal filter strength σ_T is shown on the horizontal axis. The marked region corresponds to a domain where the number of detected edges does not match the real number of edges (12). The open marker on the boundary represents one of the possible optimized filter settings and the arrows indicate how the directional filter strength is adjusted.

correct number of edges and a complementary domain where it does not produce the correct number of edges. This idea is illustrated in Fig. 2.3, where the marked region corresponds to the domain where the number of detected edges does not match (at least for one scan-line) the real number of edges. We define a filter to be optimized whenever it is on this boundary.

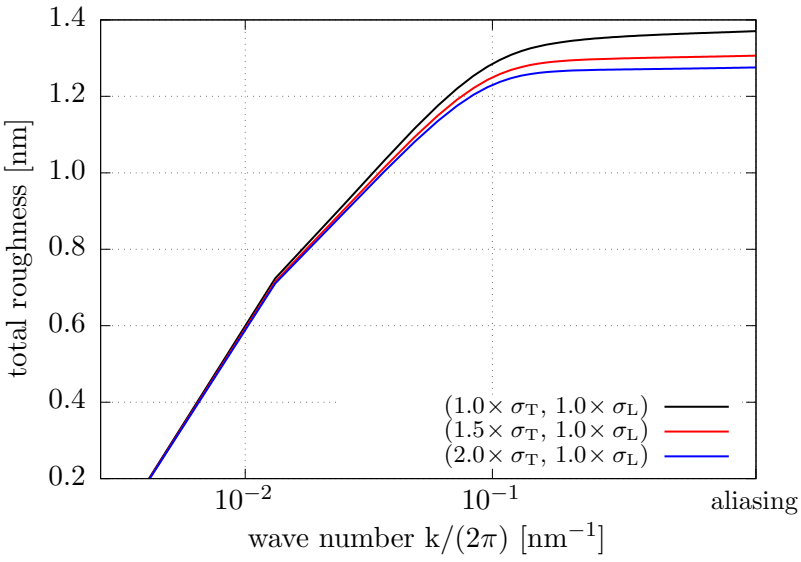
We note that an optimized filter is preferred, because a filter not only reduces image noise but also destroys frequency content related to the actual edge displacements. The idea of an optimized filter is to minimize this effect. A point on the boundary of Fig. 2.3 can be found, for example, by bisective optimization. The idea is to start with a predefined minimal and maximal filter strength and split the interval in half as long as the number of edges is correct. This procedure (bisection) is repeated until the resulting interval convergences to a point on the boundary. This way we have found,

for the 50 recorded SEM images, the following optimized directional filter strengths: $\sigma_{\mathrm{T}} = 2.5\mathrm{px} \pm 0.87\mathrm{px}$ and $\sigma_{\mathrm{L}} = 3.6\mathrm{px} \pm 0.69\mathrm{px}$. We emphasize that this particular filter strength relates to just one arbitrary point on the boundary of Fig. 2.3. However, we attempt to demonstrate that the PSD is affected, even though we optimize our filter strength. In this respect, any point on the boundary is acceptable as long as we are consistent in our method of determination. In Fig. 2.4a we increase the strength of the filter in the transverse direction, while the longitudinal strength remains fixed at the optimized value. Similarly, in Fig. 2.5a we only increase the strength of the filter in the longitudinal direction. The arrows in Fig. 2.3 show how the directional filter strength is adjusted. In Fig. 2.4b and Fig. 2.5b we plot the square root of the cumulative sum of the PSD from the lowest wave number towards the highest wave number. This cumulative sum gives the variance via Parseval's Eq. 2.4 and by taking the square root we can see how the standard deviation develops as a function of increasing wave numbers. In Figs. 2.4a and 2.5a we observe an undesired effect of the filtering on the PSD: besides suppression of the high frequencies (which is required to reduce the image noise), the center frequencies (near $k/(2\pi) = 10^{-1}$) are affected as well. This conclusion (filtering biases the result) is drawn by various authors as well, see for example Refs. 12, 18, 21, 15, 22, 23. Here we show that this effect, for both transversal and longitudinal filtering, cannot be ignored, even when considering an optimized filter strength. This is best seen in the cumulative sums in Figs. 2.4b and 2.5b, where the total roughness is shown for different filter strengths. We have summarized the effect on LER in Table 2.1 for varying directional filter strengths. Several attempts have been made to correct for this biasing, see for example Refs. 30, 18, 19, 31, 16, 23. The problem is, however, that these attempts do not capture the effect of filtering, instead they capture the effect of image pixel noise.

We now ask the following question: How much further do we need to reduce the strength of the filter (beyond the optimized filter), such that the filter dependency on the measured LER becomes insignificant? We failed in reducing the strength of the filter, because then the edges cannot be

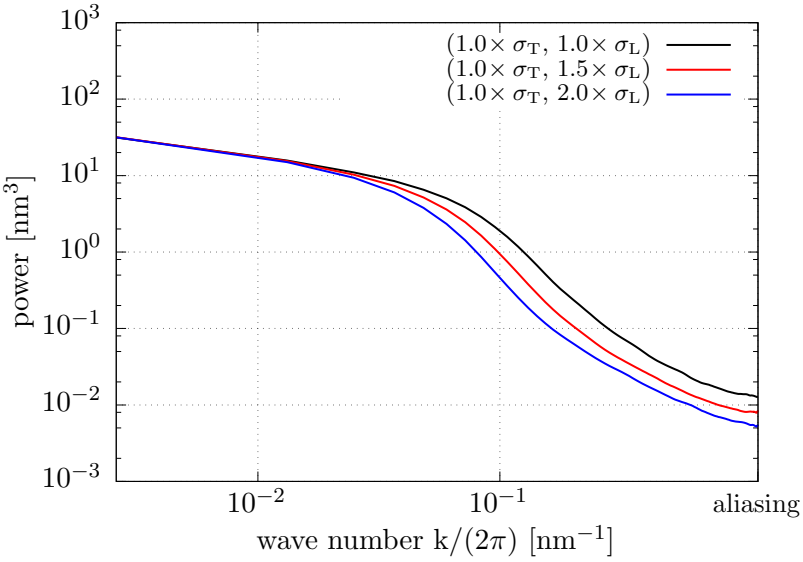


(a)

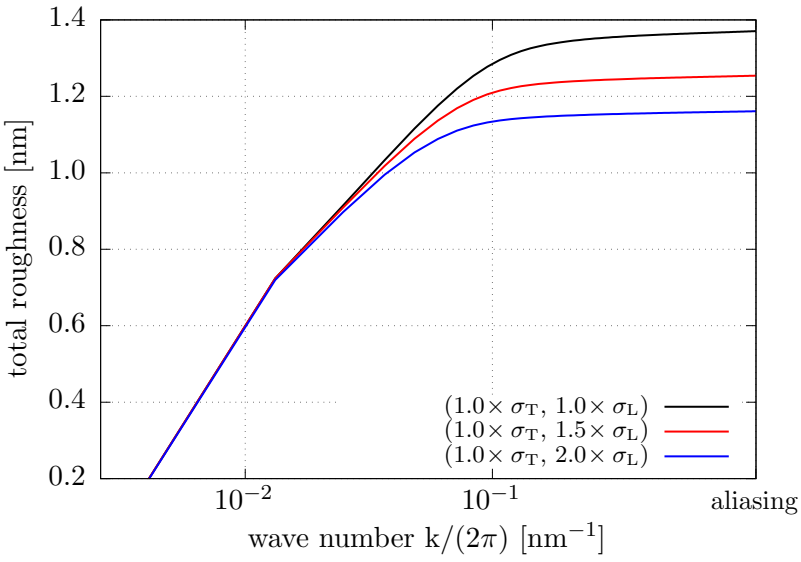


(b)

Figure 2.4: The effect of transverse filtering on the discrete PSD and the total roughness is shown. The results are obtained from analyzing the set of 50 experimental SEM images recorded by J. Jussot. The PSD is given in Fig. 2.4a and the total roughness is given in Fig. 2.4b, which is obtained by taking the square root of the cumulative sum of the PSD. The markers correspond to different transversal strengths with respect to the optimized elliptic filter.



(a)



(b)

Figure 2.5: The effect of longitudinal filtering on the discrete PSD and the total roughness is shown. The results are obtained from analyzing the set of 50 experimental SEM images recorded by J. Jussot. The PSD is given in Fig. 2.5a and the total roughness is given in Fig. 2.5b, which is obtained by taking the square root of the cumulative sum of the PSD. The markers correspond to different longitudinal strengths with respect to the optimized elliptic filter.

Filter strength	Transversal	Longitudinal
Optimized filter	1.37 nm	1.37 nm
Intermediate filter	1.30 nm	1.25 nm
Strong filter	1.28 nm	1.16 nm

Table 2.1: LER when using an intermediate or strong filter in respectively the transverse or longitudinal direction. The optimized elliptic filter, obtained by analyzing 50 recorded SEM images of line edges, corresponds to a transversal strength of $\sigma_{\mathrm{T}} = 2.5\mathrm{px} \pm 0.87\mathrm{px}$ and a longitudinal strength of $\sigma_{\mathrm{L}} = 3.6\mathrm{px} \pm 0.69\mathrm{px}$. The intermediate filter has $1.5\times$ the strength of the optimized filter in either the transversal or longitudinal direction. Similarly, the strong filter has $2.0\times$ the strength of the optimized filter.

detected anymore by our threshold based peak detection. The interested reader might argue that we could use the threshold level to reduce the strength of the optimized filter even further. It is true indeed that the strength of the optimized filter depends on the particular choice of the threshold level. The threshold level balances the amplitude of the image noise and the amplitude of a peak due to an edge. This must be seen in the context of filtering, where we would like to reduce the filter strength as much as possible. This means that the threshold level should be as low as possible. In addition, the threshold level could be optimized dynamically per line edge which would probably produce a weaker filter than the optimized filter we have used. In other words, maybe we could reduce the influence of the filter by clever optimization tricks, but how do we know if and when the influence has become insignificant? We conclude that the best solution would be to avoid a filter at all.

We recently developed a method to detect edge displacements without the use of a filter. In fact, this method works without using any post-processing of the SEM image. Let us explain this method in detail. We approximate the signal profile of the SEM by integrating a recorded SEM image in the direction of the line edges. For example, the approximate SEM signal profile of Fig. 2.1 is given in Fig. 2.6. We emphasize that this is only an approximation, because by integrating in the direction of the line edges, the actual shape of this profile becomes a function of the roughness,

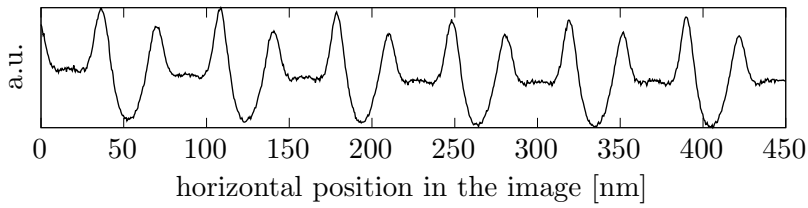


Figure 2.6: The SEM signal profile for twelve line edges as obtained by integrating the SEM image of Fig. 2.1 in the direction of the line edges.

which is present in the edges. In principle, the roughness of the edges must be corrected for by counter displacing the rows. For the moment we assume that this roughness dependency can be neglected and that the profile function can be seen as a the SEM signal response to a straight edge. We will come back to this issue later. The SEM signal profile of an isolated

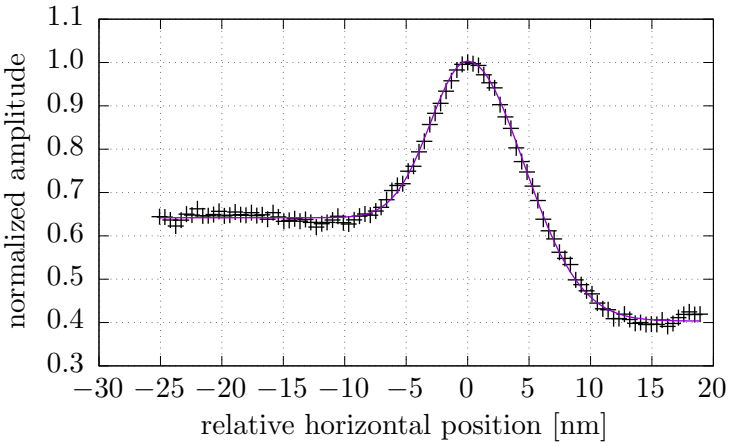


Figure 2.7: The SEM signal profile of an isolated edge taken from Fig. 2.6. The solid line is the best fit of the model given by Eq. 2.7.

edge is shown in Fig. 2.7, which is just one of the twelve peaks shown in Fig. 2.6. We model the SEM signal profile of an isolated edge by matching two vertically shifted and normalized Gaussians at the center, which has the following mathematical representation,

$$P(x) = \begin{cases} b_L + (1 - b_L) \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_L^2}\right) & x < \mu \\ b_R + (1 - b_R) \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_R^2}\right) & x \geq \mu \end{cases} \tag{2.7}$$

where b_L and b_R defines the base level found at the far left and right in Fig. 2.7, σ_L and σ_R defines the standard deviation to the left and right of the center μ. The left base level is defined such that $b_L = \lim_{x\to-\infty} P(x)$ and the right base is defined such that $b_R = \lim_{x\to+\infty} P(x)$. Note that if $b_L = b_R = 0$ (the Gaussians are not shifted upwards), and $\sigma_L = \sigma_R$ (same standard deviation), then Eq. 2.7 reduces to the standard definition of a Gaussian distribution, up to a normalization factor. The parameters b_L, b_R, σ_L and σ_R are fitted against the integrated SEM signal profile. The solid line in Fig. 2.7 is the best fit of this model against a single isolated experimental profile obtained by integrating Fig. 2.1 in the direction of the line edges. The method of fitting a Gaussian to an integrated SEM image is also found in Ref. 32, where it is used to estimate line widths. This study, however, considers a more complicated fit for the determination of LER.

The idea is now to match this double Gaussian model to every sampled row of a single edge using parameter optimization. We introduce the following degrees of freedom to the model for the SEM signal profile,

$$s \cdot P(x - \Delta x) \tag{2.8}$$

where s scales the profile amplitude P and Δx is the transversal displacement of the profile. The parameters s and Δx are determined by using an interior trust-region-reflective minimization algorithm. The interested reader is referred to the article of Coleman[33] for details on the minimization procedure. In Fig. 2.8 we demonstrate one of the matching results using the raw unfiltered signal, which is taken directly from the SEM image. Clearly for such noisy data, a threshold based edge detection method could not have found the position of the edge. Now that we can detect the edge displacements without using a filter, we reconsider the set of recorded SEM images from J. Jussot. However, this time we use our SEM profile based edge detection. The PSD without using a filter is given in Fig. 2.9a together with the optimized filter found in Figs. 2.4a and 2.5a. In Fig. 2.9a we observe that the pixel noise really starts to contribute after the marker 'noise limit'. In Fig. 2.9b we observe that the cumulative sum after the marker accounts for approximately $1.6\,\mathrm{nm} - 1.4\,\mathrm{nm} = 0.2\,\mathrm{nm}$. Therefore, a crude

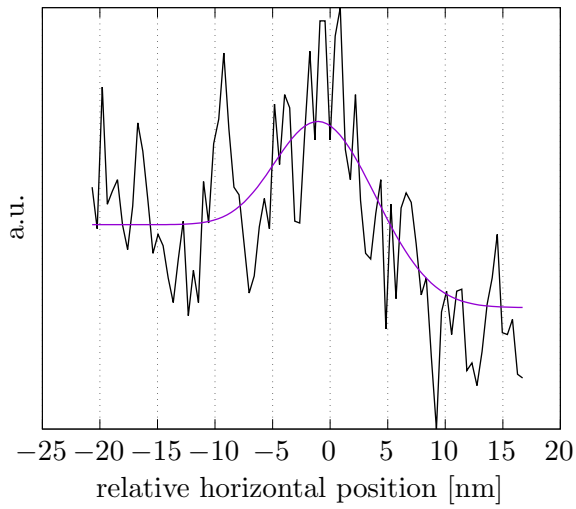
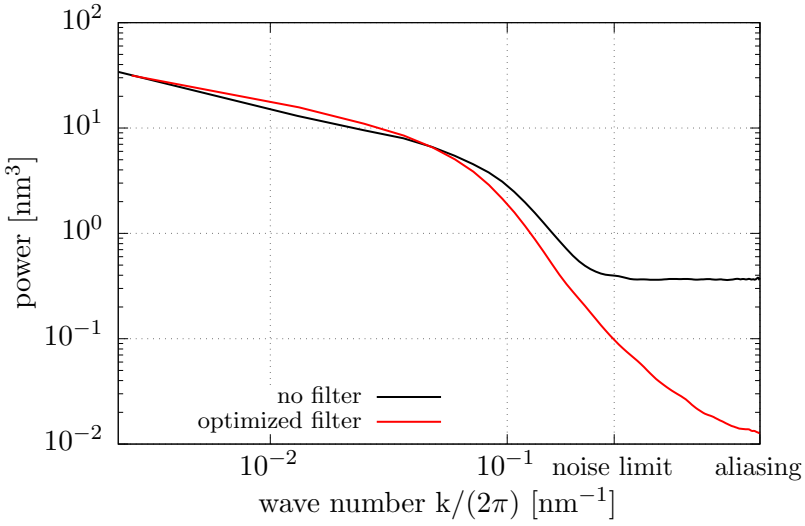
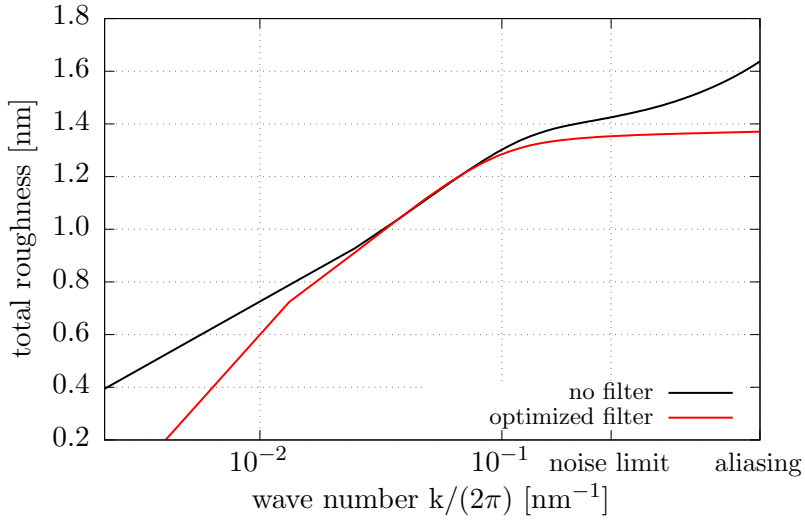


Figure 2.8: The matching of the model for the SEM signal profile in the raw (unfiltered) amplitude of an edge in a SEM image is illustrated. The profile is matched by varying the transversal displacement and amplitude of the profile function.

estimate for the actual LER is 1.4 nm. This is to be compared to the estimate of 1.3 nm for the optimized filter (square markers) found in Fig. 2.9b. Apparently, the optimized filter that we have applied in our first attempt is not that far off. What about the fact that we neglected the roughness in the edges when determining the SEM signal profile? At the moment, the roughness of the edges is effectively absorbed into the parameters (specifically σ_L and σ_R) of the approximated SEM signal profile. We expect that the noise level in the PSD (see Fig. 2.9a) could be lowered by correcting for the roughness in the SEM signal profile. This ultimately means that we have not yet established the Cramér-Rao Lower Bound (CRLB). At this point LER no longer depends on post-processing (such as filtering), but has become a function of the pixel noise and, because our estimator is not the CRLB, additional noise due to the specific method of edge determination. This, however, is not a problem because the total noise level can be taken into account[16] and will be our next point of attention.



(a)



(b)

Figure 2.9: A comparison is shown of the PSD and total roughness of filter-free SEM profile based detection to optimized elliptic filtered edge detection. These results are obtained from analyzing the set of SEM images from J. Jussot (see Fig. 2.1 for an example of one of the images). The PSD is given in Fig. 2.9a and the total roughness is given in Fig. 2.9b, which is obtained by taking the square root of the cumulative sum of the PSD. Note that the optimized filter from Figs. 2.4a and 2.5a is used for comparison.

2.3 Image noise analysis

In estimating LER (Fig. 2.9a) we used all available images (50 in total) to reduce the uncertainty in the discrete PSD. We now improve this LER determination by considering a model for the PSD and question how much noise is acceptable, such that we still can determine LER? In fact there are two questions: How much noise can we allow in a single image such as Fig. 2.1 and how many line edges do we need in total? We begin our investigation on the effect of image noise on the determination of LER by generating rough edges at random using the model of Palasantzas[26] with known parameters,

$$\mathrm{PSD}(k) = \frac{\sqrt{\pi}}{2\pi} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)} \frac{2\sigma^2\xi}{(1 + k^2\xi^2)^{\alpha+\frac{1}{2}}} \tag{2.9}$$

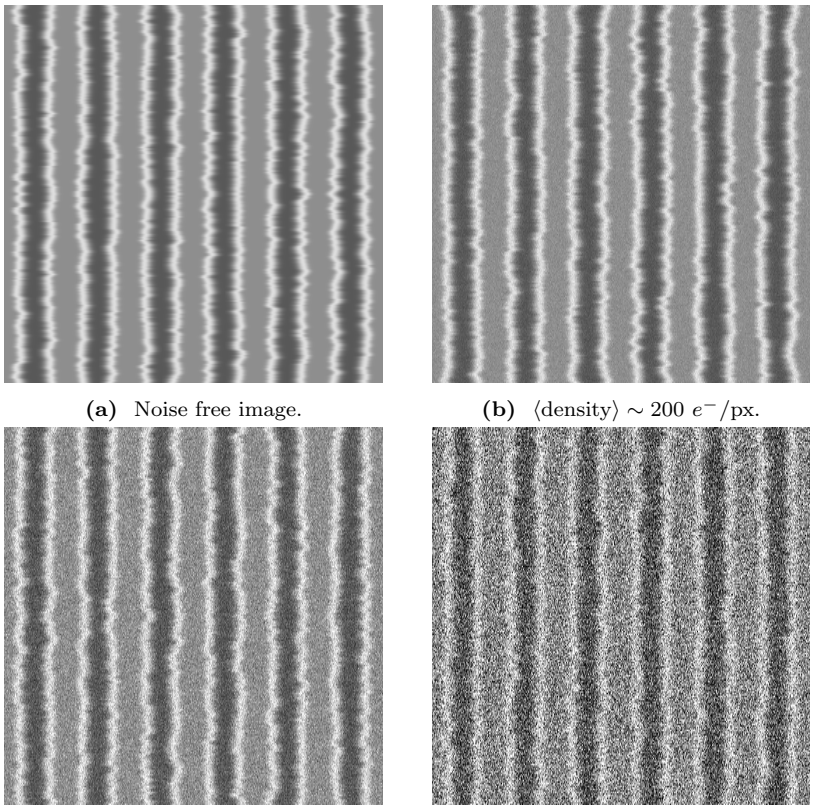
This PSD defines an infinitely long line with σ as the LER, ξ the correlation length and α the roughness exponent. It can be verified that the integral of this PSD equals the variance,

$$\sigma^2 = \int_{-\infty}^{+\infty} \mathrm{PSD}(k)\,\mathrm{d}k \tag{2.10}$$

The random displacements can be generated via the algorithm of Thorsos, which is explained in Ref. 27. The algorithm of Thorsos produces random edge displacements that, in the limit of large averages, converges towards the PSD of Palasantzas up to a bias in the standard deviation. This bias is explained in Ref. 27 as well and can be compensated for by multiplying the edge displacements with a constant factor.

We generate top-down SEM-like images by using the model of the SEM signal profile obtained earlier (Fig. 2.7). A SEM-like image is obtained by displacing the SEM signal profile at every row in accordance to the random generated displacements, which in turn satisfy the PSD given by Eq. 2.9. If we collect a number of randomly generated lines next to each other in one image, we obtain the result of Fig. 2.10a. This is a simulated result of a noise-free top-down SEM-like image of randomly generated line edges.

The image is further processed by adding Poisson distributed noise to every pixel of the noise-free image. This Poisson distributed noise is added in a similar way as in Ref. 23. We emphasize that, in contrast to Ref. 23, we did not include Gaussian distributed noise. Examples of Poisson noise generated SEM-like images are given in Figs. 2.10b, 2.10c and 2.10d. In each of the images, the average electron density is set to (in respective order) 200, 20 and 2 electrons per pixel. The corresponding average charge density equals (in respective order) $1000\,\mu C/cm^2$, $100\,\mu C/cm^2$ and $10\,\mu C/cm^2$.



(a) Noise free image. **(b)** $\langle \text{density} \rangle \sim 200\ e^-/\text{px}$. **(c)** $\langle \text{density} \rangle \sim 20\ e^-/\text{px}$. **(d)** $\langle \text{density} \rangle \sim 2\ e^-/\text{px}$.

Figure 2.10: Randomly generated top-down SEM-like images with line edges. The field size is 2 µm in length (256px) and 450 nm wide (1024px). The pixel size is about 7 nm × 0.5 nm.

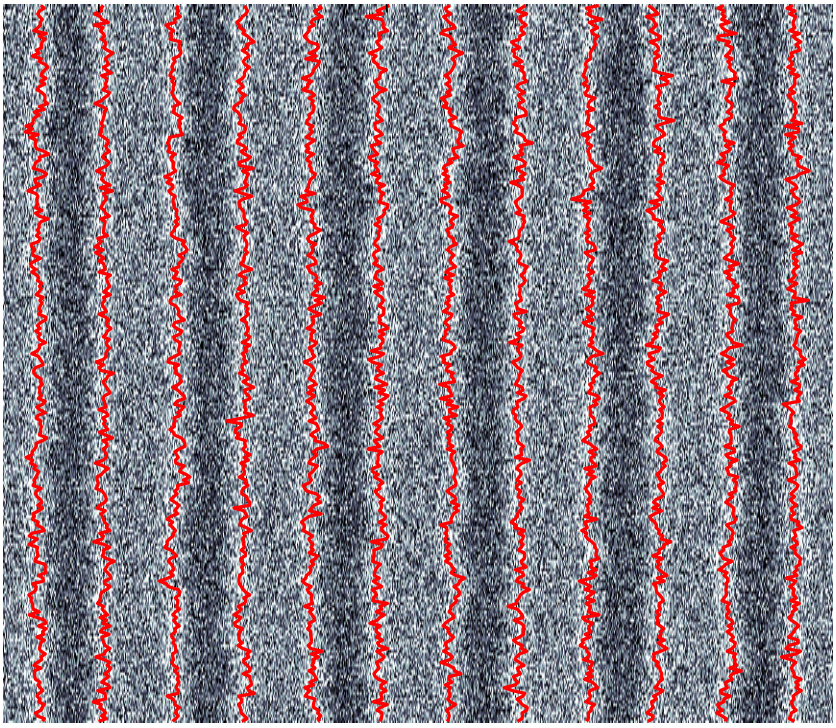


Figure 2.11: Demonstration of filter-free SEM profile based detection applied to a randomly generated top-down SEM-like image with line edges. The field size is 2 µm in length (256px) and 450 nm wide (1024px). The average electron density of this image is about 2 electrons per pixel and corresponds to an average charge density of about 10 µC/cm^2.

The edges of the random generated images are determined using filter-free SEM profile based detection as described before. An example of displacement detection applied to a randomly generated image is given in Fig. 2.11. The difference between the randomly generated displacements and the detected displacements after adding pixel noise tells us how pixel noise translates to noise in the edge displacements. This is illustrated in Fig. 2.12, where the black line corresponds to the randomly generated displacements and the red line corresponds to the detected displacements after adding pixel noise. We conclude from running many simulations that this pixel noise translates to a noise distribution in the edge displacements which is uniform (flat) in the PSD. In other words, this means that pixel noise translates to white noise in the edge displacements. Now that we have classified

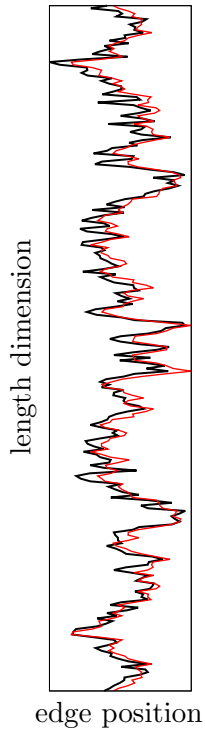


Figure 2.12: The randomly generated displacements (black) are compared with the detected displacements (red) after adding Poisson distributed pixel noise. The difference between the generated and detected displacements gives the noise distribution in edge detection.

the total noise (constituting of pixel noise and additional noise due to our specific method of edge determination), we are free to add this noise term to the Palasantzas model (see also Refs. 16, 17),

$$\mathrm{PSD}_{\mathrm{w/noise}}(k) = \mathrm{PSD}_{\mathrm{w/o\ noise}}(k) + \sigma_N^2 \frac{\Delta y}{2\pi} \tag{2.11}$$

where σ_N is the noise level and Δy the measurement interval. The PSDs obtained by detecting the edge displacements in simulated SEM-like images with an electron density of 2, 20 and 200 electrons per pixel are given in Fig.2.13. In these images we can also see that pixel noise translates to white noise: The power of the high frequencies flattens out to a straight line as we decrease the electron density per pixel and thus introduce more noise. The

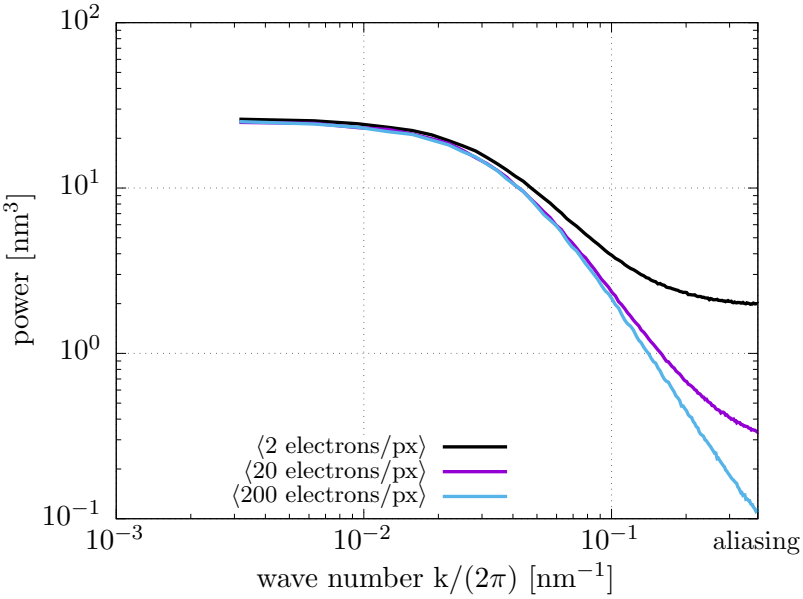
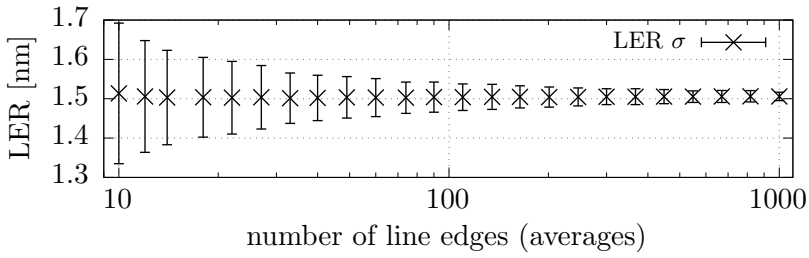


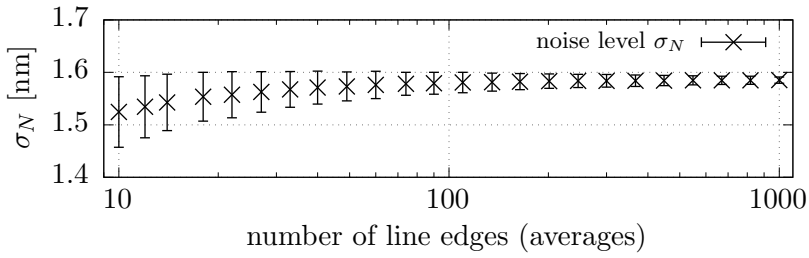
Figure 2.13: The effect of adding a noise term to the Palasantzas model is shown. The PSDs in this figure are obtained from averaging many images for different noise settings.

idea is now to fit the simulated discrete PSD as a function of the number of line edges and analyze the convergence of the parameters (σ, ξ and α) of the Palasantzas model, Eq. 2.11, extended with a term to capture the total noise (σ_N). We remark that by fitting Eq. 2.11, we neglect the systematic errors described by Ref. 17. We neglect the systematic errors because in our study we have a larger number of sampled edge displacements N and the power of the noise level renders the effects due to aliasing or spectral leakage to negligible levels. Our simulation is set up as follows. We generate random line edges with a length of 2 µm (256px) using the modeled SEM signal profile of Fig. 2.7. We choose a roughness (LER) of 1.5 nm, correlation length ξ of 25 nm and a roughness exponent α of 0.75 in the Palasantzas model. These are typical values for experimentally measured edges.[16, 17] Now we consider the worst case in image noise of the densities given in Fig. 2.10, which corresponds to an electron density of about 2 electrons per pixel on average (charge density $\approx 10\,\mu C/cm^2$). We run many simulations and determine the distributions of the outcome values of the parameters

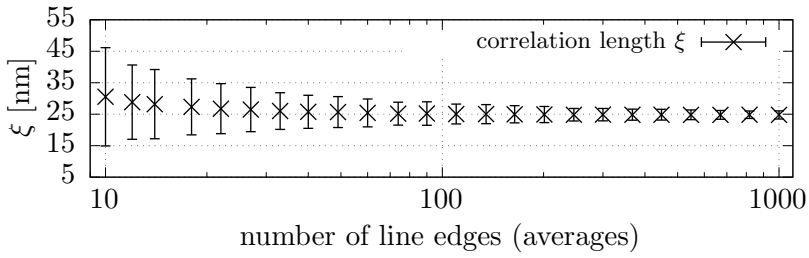
of the Palasantzas model (LER σ, noise level σ_N, correlation length ξ and roughness exponent α). The results of this low dose (high noise) simulation are given in Fig. 2.14. At first we observe that all estimated parameters of the Palasantzas model are converging towards the predefined parameters. The convergence, however, seems to be asymptotic, which is best seen by the noise term σ_N in Fig. 2.14b and the correlation length ξ in Fig. 2.14c. There seems to be a very small bias in the LER, which is best seen in Fig. 2.14a. One of the possible causes for this bias is related to the intrinsic bias in Thorsos algorithm. Although we have corrected this bias in Thorsos algorithm numerically, it is with limited precision (about 3 digits). The errorbars roughly decrease as the square root of the number of averages, which is to be expected based on averaging principles. The relative errors (size of the errorbars divided by the value of the predefined parameter) are given in Fig. 2.15.



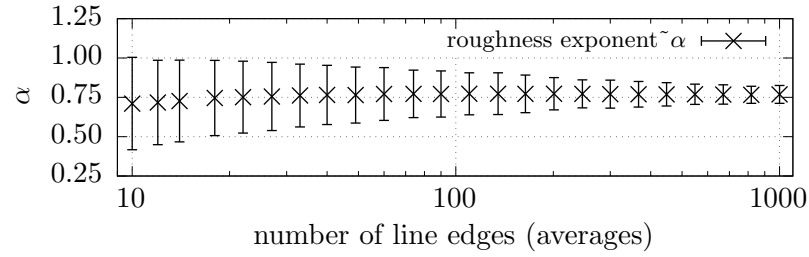
(a)



(b)



(c)



(d)

Figure 2.14: The simulation results for fitting the parameters of the Palasantzas PSD model to random generated lines after adding Poisson distributed pixel noise are shown. The random lines are generated with a length of 2 µm (256px) using the modeled SEM signal profile of Fig. 2.7. The generated lines have a roughness (LER) of 1.5 nm, correlation length ξ of 25 nm and a roughness exponent α equal to 0.75. The simulated electron density is about 2 electrons per pixel on average. This corresponds to a charge density of approximately 10 µC/cm^2.

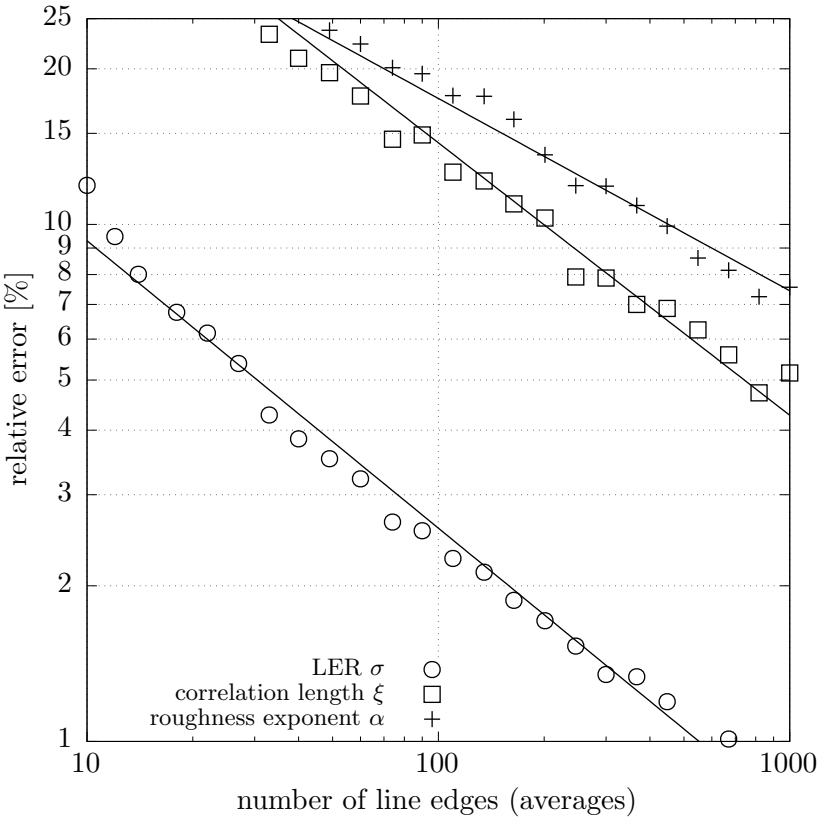


Figure 2.15: The relative error (one-sigma) in parameter estimation with respect to the Palasantzas model for simulated top-down SEM-like images with a density of about 2 electrons per pixel on average is shown. The result of this figure is based on the errorbars in Fig. 2.14.

We observe in Fig. 2.15 that the correlation length ξ and roughness exponent α are harder to estimate, i.e. it takes more averages to produce the same relative error as for LER σ and noise level σ_N. The same conclusion can be found in Ref. 23. The most interesting parameter for industry is LER σ, because that is the parameter against which process performance is evaluated. When the number of averages is low, the intrinsic noise in the discrete PSD is significant, as can be seen in Fig. 2.16. It is remarkable that under these conditions LER can still be estimated with a relative error (one-sigma) of about 10%. In other words, it only takes one single image, Fig. 2.10d, with 2 electrons per pixel (charge density $\approx$ 10 µC/cm^2) to estimate LER as 1.5 nm $\pm$ 10%. We now question how the parameters converge

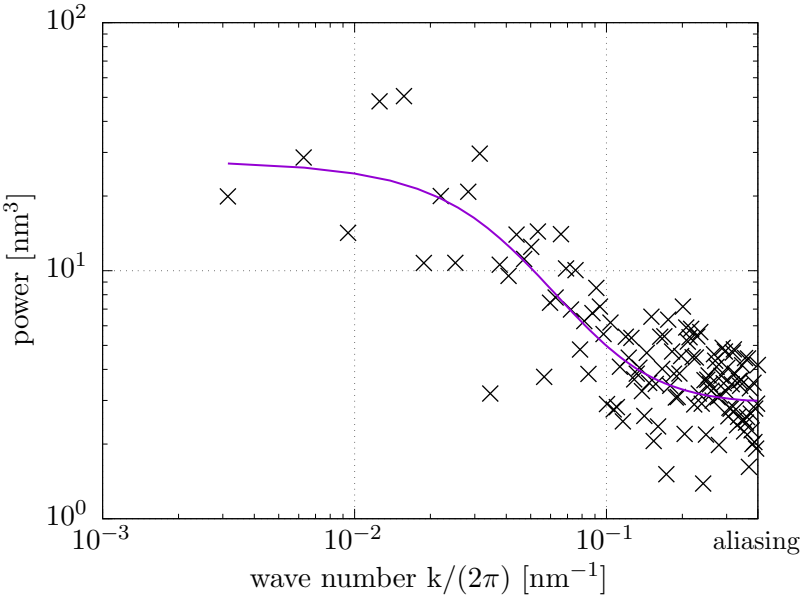


Figure 2.16: The intrinsic noise in the discrete PSD after averaging only 10 edges is shown. The purple line corresponds to the limit of many averages. The data is obtained from a simulation corresponding to a density of about 2 electrons per pixel, which corresponds to a charge density of about 10 µC/cm^2.

when we increase the electron density per pixel. The effect of this on LER σ is given in Fig. 2.17, where we plot the relative error as a function of the number of line edges for densities of 2, 20 and 200 electrons per pixel. The result of Fig. 2.17 shows that it hardly makes any difference if we increase the density from 20 electrons per pixel (charge density $\approx$ 100 µC/cm^2) to 200 electrons per pixel (charge density $\approx$ 1000 µC/cm^2). The explanation is as follows. We identify that the relative error has two contributions: Pixel noise and variance because of limited edge length. This can also be seen by integrating the Palasantzas model given by Eq. 2.11,

$$\sigma^2_{\text{w/noise}} = \sigma^2_{\text{w/o noise}} + \sigma^2_N \tag{2.12}$$

The relative error is then determined as,

$$\text{error}\left(\sigma^2_{\text{w/noise}}\right) = \text{error}\left(\sigma^2_{\text{w/o noise}}\right) + \text{error}\left(\sigma^2_N\right) \tag{2.13}$$

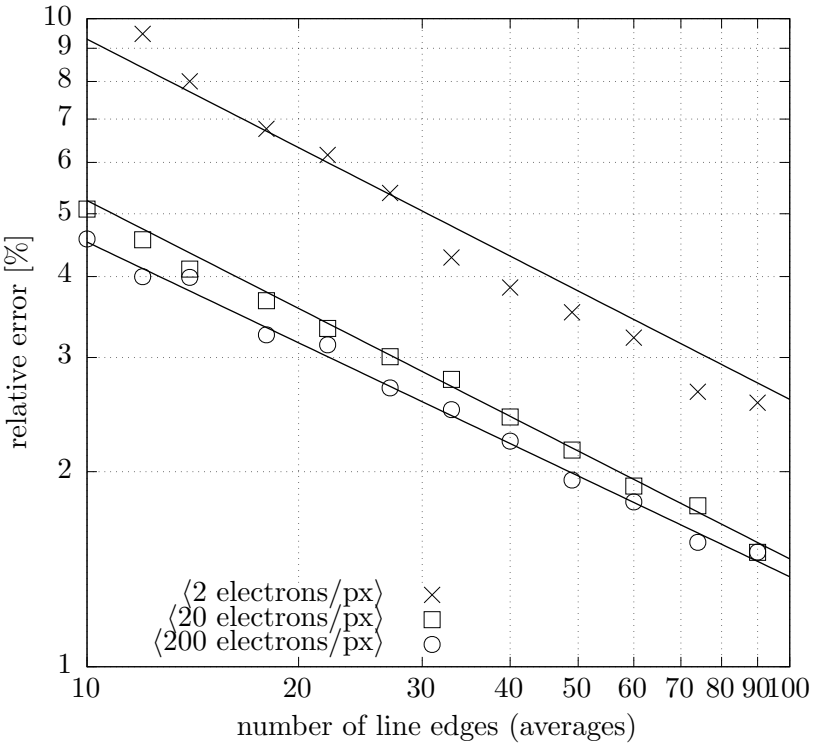


Figure 2.17: The relative error in LER determination for simulated top-down SEM-like images with a density of 2, 20 and 200 electrons per pixel on average is shown.

In the simulation of an average density of 2 electrons per pixel we have that both terms in Eq. 2.13 contribute to the total error. The error due to pixel noise, which is the second term in Eq. 2.13, decreases as we increase the electron density per pixel. If we increase the electron density even further, the variance caused by limited edge length, which is the first term in Eq. 2.13, becomes significantly larger than the error contribution due to pixel noise. When we reach this point, the total relative error is primarily determined by the variance caused by limited edge length. This means that there is no point in further increasing the dose as this does not improve LER determination. In Fig. 2.17 we see that this already occurs at an averaged density of 20 electrons per pixel.

2.4 Resist shrinkage

So far we have only applied our method of determination to artificially generated images. Our intention for this section is to apply the method to images acquired with a *real* CD-SEM. We consider the acquisition of images in the limit of image shot noise using a CG4000 CD-SEM from Hitachi. In order to achieve maximum image noise, and hence the lowest electron beam impact, we choose a beam voltage of 300 eV, a probe current of 10 pA and a single integration frame. The latter corresponds to the lowest possible default configuration.[1]

In collaboration with CNRS-LTM/CEA-LETI, we have accomplished the following experiment. At first a wafer is exposed using a pattern generator from VISTEC with a repeated pattern of lines and spaces with fixed spacing and fixed line width.[2] After the development, a sequence of images is recorded at a fixed position on the wafer using only a single integration frame per image. An example of an isolated image in that sequence is shown in Fig. 2.18. The consequence of 'hammering' the same pattern is that the net dose accumulates for consecutive images in the sequence. The essential idea is to extract the effect of shrinkage by comparing the average critical dimension (CD) per image in the sequence. Care was taken to be sure that the first recorded image was really the first deposited imaging dose (the focus and optimization of the CD-SEM was performed on a nearby pattern). For statistics, we repeat this procedure at different locations on the wafer. We have recorded a total of 25 independent sequences with 50 images per sequence. In other words, we have 25 independent measurements for each frame with a certain net accumulated dose,

$$D = \frac{\text{probe current} \times \text{scan rate}}{\text{scan area}} \times \text{frame number} \qquad (2.14)$$

For the CG4000, we have a probe current of 60 pA and a scan rate of 50 Hz (TV scan mode). The scan area equals 2.76 µm×450 nm. The resulting

[1] In reality, the CD-SEM can be pushed further in service mode, but then requires manual fine tuning which we would like to avoid for additional complications.

[2] The resist that was used by CNRS-LTM/CEA-LETI is unknown to us due to disclosure restrictions.

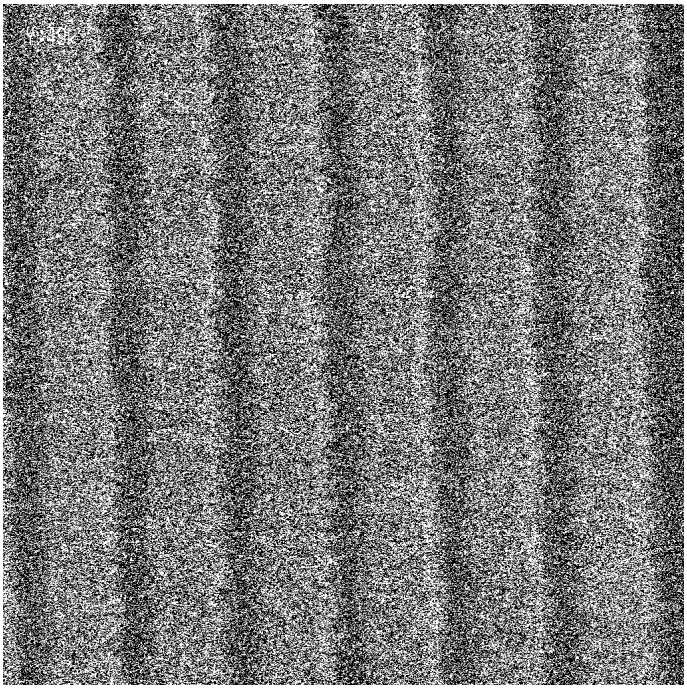


Figure 2.18: An image of line edges acquired using the lowest possible default configuration of a CG4000 CD-SEM from Hitachi is shown. This image which was recorded by I. Servin from CNRS-LTM/CEA-LETI in 2016. The lowest possible configuration corresponds to a beam voltage of 300 eV, probe current of 10.0 pA and only a single integration frame. The magnification is such that the field size is 2.76 µm in length (1024px) and 450 nm wide (1024px). The resulting pixel size is approximately 2.7 nm × 0.44 nm.

dose equals approx. 16.13 µC/cm^2 per image. Please note that the latter boils down to a single electron per square nanometer per image. The interested reader is invited to compare the experimental image (Fig. 2.18 with 1 electron/nm^2) to the randomly generated image (Fig. 2.11 with 2 electrons/nm^2).

We have determined the change in CD per image as follows. At first we detect the edges using our profile based edge detection algorithm. In reality, care must be taken in how to *define* the line width from the edge positions in a two dimensional image. We, however, are only interested in how the CD *changes* per image. This means in practice, that we can choose any definition as long as we are consistent. For each line, we therefore simply determine the CD by subtracting the position of the right edge from

the left edge. The CD for each image is then obtained by calculating the average CD per line in that image. The change in CD for consecutive images is then obtained as follows,

$$\Delta\mathrm{CD} = \mathrm{CD}_i - \mathrm{CD}_0 \tag{2.15}$$

where CD_i is the average CD for image number $i \geq 1$ and CD_0 is the extrapolated CD for zero exposure dose. The CD at zero exposure dose is determined by invoking a model based on independent multi-exponential decay,[34, 35, 36]

$$\mathrm{CD}(D) = \sum_i \alpha_i \exp \frac{D}{\beta_i} \tag{2.16}$$

where D is the net accumulated dose, α is the amplitude for the exponential decay and β the corresponding decay constant. The summation accounts for the fact that different processes take place on different time scales. For example, there could be (1) a very fast process (small β) which only affects the outer rim of the material, (2) a slower process (larger β) for the main bulk of the material and (3) a very slow process (largest β) accounting for the decay of the entire material.[34] In the end, a Levenberg-Marquardt non-linear least square fit of the decay model is applied to the measured CD per image with respect to the net accumulated dose. The final result of our measurement is shown in Fig. 2.19. We found, by trial and error, that a model with two terms for the decay works best for Fig. 2.19. The corresponding parameters of the exponential decay model are given by Table 2.2.

magnitude (1)	2.434 nm (12.92 %)
decay constant	4.129 electrons/nm^2
magnitude (2)	16.41 nm (87.08 %)
decay constant	691.5 electrons/nm^2

Table 2.2: The fitted parameters in the double exponential decay model for resist 'S13' is shown.

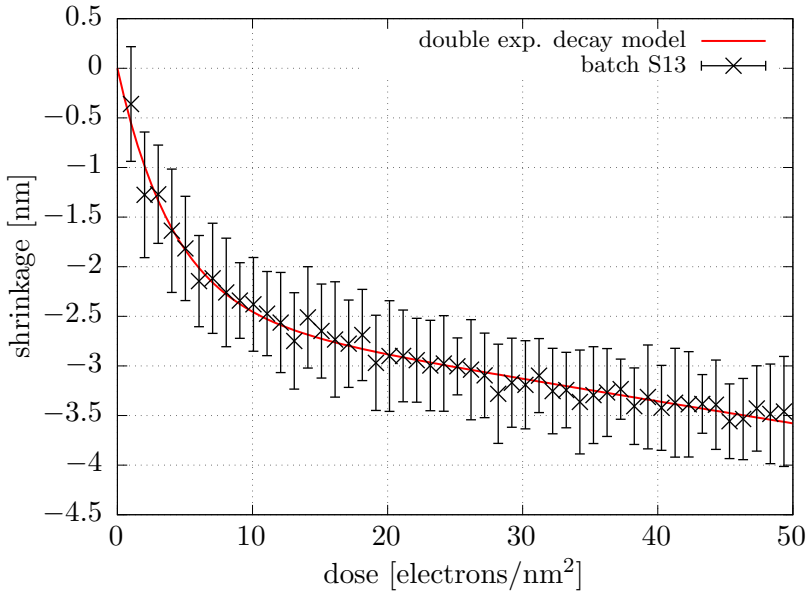


Figure 2.19: The amount of shrinkage of resist 'S13' as a function of accumulated dose is shown. The red line corresponds to a fitted double exponential decay model. The extrapolated CD to zero dose is found to be 18.8 nm ± 0.351 nm. Details on the acquisition and analysis are explained in the main text.

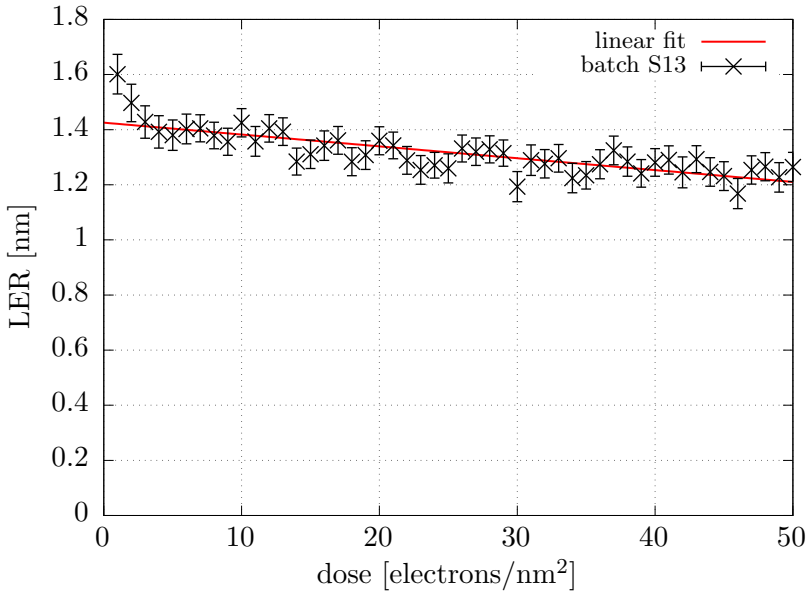


Figure 2.20: The LER of resist 'S13' as a function of accumulated dose is shown.

Notice that in Fig. 2.19, there is a strong non-linear effect of shrinkage for a dose lower than 10 electrons/nm^2 (approx. 161 µC/cm^2). In fact, most of the shrinkage occurs in this regime. The typical amount of integration frames used for the analysis of CD is somewhere between 8 and 32, see for example Ref. 16. We would like to point out that by the time we have accumulated 8 frames, the CD has shrunk already by more than 2 nm! The shrinkage beyond 10 electrons/nm^2 is dominated by the decay of the bulk, which accounts for approx. 87%. Notice that the corresponding decay constant renders the reduction of the CD to approximately linear. In addition, we have also determined the LER as a function of accumulated dose, see Fig. 2.20. What is observed is a decreasing trend in LER as the dose accumulates.

We have repeated the shrinkage analysis by using *another* wafer with a different type of resist. The result of that analysis for the CD is shown in Fig. 2.21, and for the LER in Fig. 2.22. For this particular resist (which

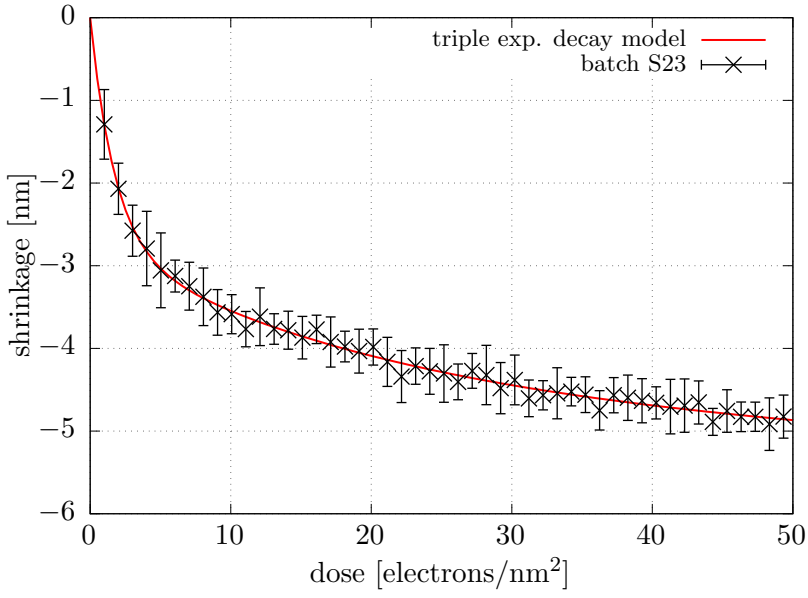


Figure 2.21: The amount of shrinkage of resist 'S23' as a function of accumulated dose is shown. The red line corresponds to a fitted triple exponential decay model. The extrapolated CD to zero dose is found to be 20.4 nm ± 5.27 nm. Details on the acquisition and analysis are explained in the main text.

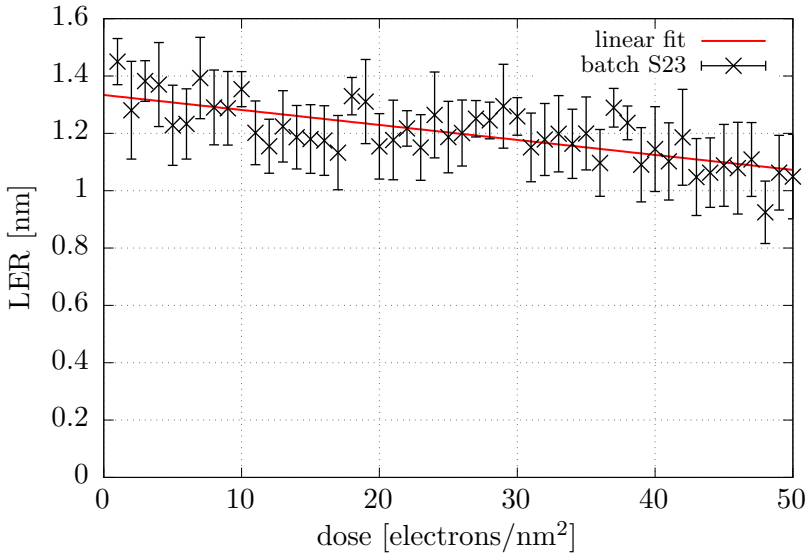


Figure 2.22: The LER of resist 'S23' as a function of accumulated dose is shown.

is also unknown to us due to disclosure restrictions) we found that a model with three terms for the decay works best. The corresponding parameters of the exponential decay model are given by Table 2.3. The latter includes, in comparison to Table 2.2, a decay process with an intermediate rate.

magnitude (1)	2.719 nm (13.36 %)
decay constant	1.740 electrons/nm^2
magnitude (2)	1.923 nm (9.450 %)
decay constant	19.97 electrons/nm^2
magnitude (3)	15.71 nm (77.19 %)
decay constant	2031 electrons/nm^2

Table 2.3: The fitted parameters in the triple exponential decay model for resist 'S23' are shown.

Inevitably, what we see is that the amount of shrinkage of the CD strongly depends on the type of resist and that LER decreases as a function of accumulated dose. The resist shown by Fig. 2.21 has a stronger decay (the shrink at 8 integration frames is more than 3 nm). This demonstrates that the risk of resist shrinkage is *real* and has a significant impact.

2.5 Conclusion

The off-line determination of LER in top-down SEM image requires low noise images. This means that we need a higher dose or we filter the SEM image before edge detection. However, by increasing the dose we run the risk of resist shrinkage. We have shown that the effect of transversal and longitudinal filtering cannot be ignored, even when considering an optimized filter strength. We conclude that the best solution is to avoid a filter at all.

We developed a different method in which we match the approximated SEM signal profile against the raw (unfiltered) amplitude of edges in a SEM image. The SEM signal profile is obtained by integrating an experimental top-down SEM image of line edges in the direction of the edges. With this method it is possible to detect the edge displacements in very noisy images without using a filter. In fact, this method works without using any post-processing of SEM images.

In a simulation study we show that LER can still be determined from very noisy images with only 2 electrons per pixel on average ($\approx 10\,\mu C/cm^2$). The PSDs are generally averaged over many line edges to reduce the statistical error. However, even a single image with 12 line edges, produces an estimation of the LER with a relative error of only 10%. Furthermore, increasing the dose beyond 20 electrons per pixel does not significantly improve the LER determination.

Finally, we have applied the method to experimental CD-SEM images. We have accomplished an experiment where a pattern of lines and spaces is repeatedly imaged with the lowest possible settings of a CG4000 CD-SEM from Hitachi. For each consecutive image, the net accumulated dose increases, which causes the resist to shrink. We have analyzed two different types of resist and used multi-exponential decay model to characterize the shrinkage. What is observed is that most of the shrinkage happens in the very first few frames and is non-linear of nature. In addition, we have observed that LER decreases for accumulating dose.

We conclude that low voltage and low dose CD-SEM image analysis is a necessary prerequisite for reliable characterization of CD and LER.

Chapter 3

The physics of electron scattering in matter

The purpose of this chapter and the upcoming chapters is to provide a detailed description of our electron-matter interaction simulator. There are two separate topics to be mentioned: we have (1) physical models and (2) the details of the implementation. In this chapter we only consider the physical models. In general, the physics of electron scattering can become very complicated depending on what level of detail must be included. Therefore, it is very important that we frame our intentions more precisely. Let us first discuss what we would like to simulate. We would like to (1) calculate SEM images of a given arbitrary complex solid geometry, and (2) determine the spatial distribution of electron interactions inside resist. The calculation of SEM images will be used for advanced studies concerning the dimensional metrology of patterns of lines and spaces (chapter 5). The determination of the spatial electron distribution is useful for computational lithography (chapter 6). In either case, the contents of this chapter is designed in such a way as to provide the interested reader with a consistent and self-contained theoretical description of electron scattering in matter.

Let us briefly discuss how this chapter is organized. At first we introduce and discuss the very first principles of electron scattering, such as Beer-Lambert absorption, the mean free path, (differential) scattering

cross-sections and sampling of distributions. The physics of scattering is then divided into two separate sections: elastic and inelastic scattering events. Special attention is given to the realm of low energy physics, in which the scattering of electrons with acoustic phonons plays a crucial role. We then proceed by introducing the physical effects occurring at an interface. The latter includes the interaction with surface plasmons, transmission, reflection and refraction of electrons. Finally we will discuss the limitations of the present state of the art.

Throughout this chapter we will use three different unit systems: International system of units , Gaussian units (G.S.) and natural units (A.U.). To avoid confusion, system of units other than S.I. are expressed explicitly in the right margin of each equation.

3.1 The basics of scattering

We begin our journey with the concept of electron scattering. One particular scattering event is illustrated in Fig. 3.1, where we have an electron moving in a straight line in the direction of a target, which is located inside the circle. Inside the circle, there is an interaction between the electron and the target. The net result of this event, where the target is the frame of reference, is the deflection of the electron by an angle θ.

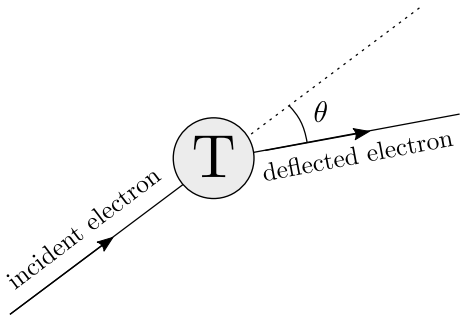


Figure 3.1: The deflection of an electron caused by the interaction with a target is shown.

In general, there are a few possible types of scattering events for the electron to be considered,

$$e^- + T \longrightarrow \begin{cases} e^- + T & \text{elastic event} \quad (3.1a) \\ e^- + T & \text{inelastic event} \quad (3.1b) \\ e^- + T^\star + \cdots & \text{inelastic event} \quad (3.1c) \\ T^\star & \text{absorption event} \quad (3.1d) \end{cases}$$

where T is the target and $T^\star$ implies that the target has changed in the interaction process. In an elastic event, the energy of the electron is conserved (Eq. 3.1a) and the electron is deflected by an angle. On the other hand, inelastic events are characterized by energy-momentum transfer. Actually, there are two subcases for the inelastic event. The first case (Eq. 3.1b) is a purely dissipative one and the second case (Eq. 3.1c) involves additional products $(\cdots)$ following from the interaction process. The absorption event (Eq. 3.1d) occurs when an electron is absorbed by a target and the electron is therefore not found after the interaction process. There is, in fact, another event which is kind of special: the electron and the target can be annihilated in a recombination event. In any case, the target T must be seen as generic and could be, for example, a molecule, atom, electron, or collective excitations, such as a plasmon or a phonon. Throughout this chapter we will discuss the physics and models of the individual interactions in detail. For the moment, however, we will be concerned with a far more pragmatic question: what is the distance an electron travels before an event occurs? All that we need to answer this question is a very intuitive description about the number of scattered electrons inside matter. Suppose that we have N initial electrons with energy E moving parallel through a material in a particular direction. We expect that the number of electrons that scatter within a slice Δs is proportional to the number of electrons going through that slice, i.e.,

$$\Delta N = -\mu(E)\Delta s N \tag{3.2}$$

where $\Delta N = N_{\text{after}} - N_{\text{before}}$ relates to the number of scattered electrons,

N is the initial number of electrons and $\mu(E)\Delta s$ is the fraction of the number of electrons with energy E that scatter within a slice Δs. There is an important assumption that we would like to mention. The scattering likelihood inside the interaction volume, for which $\mu(E)$ applies, is assumed to be homogeneous. In other words, any local change in the scattering likelihood is a violation of homogeneity. As a consequence, the initial electrons do not influence the scattering likelihood, nor is the scattering likelihood influenced by the result of the scattering event. Neither, however, are really true in reality and will be discussed further in the upcoming sections. Nevertheless, let us proceed and note that N actually is the number of electrons that have not scattered and are thus left unperturbed. In that view, let us rewrite Eq. 3.2 as follows,

$$\frac{\Delta N}{\Delta s} = \frac{N(E, s + \Delta s) - N(E, s)}{\Delta s} = -\mu(E)N(E, s) \tag{3.3}$$

By taking the limit of $\Delta s \to 0$, we obtain a first-order differential equation for the number of unperturbed electrons N. The solution to that equation is,

$$\frac{N(E, s)}{N_0} = \exp\left(-\mu(E) \cdot s\right) \tag{3.4}$$

which we have written in such a way that it gives the ratio of the number of unperturbed electrons $N(E, s)$ to the number of initial electrons N_0. The latter becomes a probability function upon normalization,

$$p(E, s) = \frac{\exp\left(-\mu(E) \cdot s\right)}{\int_0^\infty \exp\left(-\mu(E) \cdot s'\right)\mathrm{d}s'} = \mu(E) \exp\left(-\mu(E) \cdot s\right) \tag{3.5}$$

This is the celebrated Beer-Lambert law and μ is known as the attenuation coefficient of the material. The Beer-Lambert law gives us, in this case, the probability that an electron with energy E propagates a distance s through a medium without scattering. Now that we have a probability function, let us determine the mean free path (MFP) by calculating the expected value,

$$\lambda(E) = \mu(E) \int_0^\infty s \exp\left(-\mu(E) \cdot s\right) \mathrm{d}s = \frac{1}{\mu(E)} \tag{3.6}$$

Although this basically answers our question, the distance before an event occurs, however, is a random distribution. A random distance is obtained via the cumulative probability,

$$p(E, S \leq s) = \frac{1}{\lambda(E)} \int_0^s \exp\left(-\frac{s'}{\lambda(E)}\right) \mathrm{d}s' = 1 - \exp\left(-\frac{s}{\lambda(E)}\right) \tag{3.7}$$

By inverse sampling of the cumulative probability, we can draw a random distance before a particular event (associated with a certain MFP) occurs,

$$s = -\lambda(E) \ln U \tag{3.8}$$

where U is a random sample from the uniform distribution. In the latter expression we have used that the distribution of $1 - U$ is the same as U.

So far we have only considered a single type of scattering event with a characteristic MFP defined by λ. Let us now discuss how the Beer-Lambert law is affected by a variety of different possible scattering events. The probability that an electron has *not* experienced *any* event within a given distance s is determined by the product of the individual probabilities,

$$p(E, s) = \frac{\prod_i \exp\left(-\frac{s}{\lambda_i(E)}\right)}{\int_0^\infty \prod_i \exp\left(-\frac{s'}{\lambda_i(E)}\right) \mathrm{d}s'} = \left(\sum_i \frac{1}{\lambda_i(E)}\right) \exp\left(-s \sum_i \frac{1}{\lambda_i(E)}\right) \tag{3.9}$$

Note that the latter is identical to Eq. 3.5 if, and only if we define the total MFP as follows,

$$\frac{1}{\lambda(E)} = \sum_i \frac{1}{\lambda_i(E)} \tag{3.10}$$

Is this rule of addition of inverse MFPs applicable in general? The answer is no. The addition rule relies on the assumption that all of the scattering events considered in the summation are independent of each other. Another case where the addition rule becomes questionable will be discussed later when we consider the determination of the scattering cross-section of a compound using its atomic constituents. Nevertheless, one might point out that Eq. 3.10 actually corresponds with the addition rule of (differential) scattering cross-sections. Indeed, the addition rule of scattering cross-sections

is an intrinsic property of the Beer-Lambert law. To see this, let us go back to Eq. 3.2 and assume that we can write the attenuation coefficient μ as follows,

$$\mu(E) = \sigma(E)\rho_n = \sigma(E)\rho_m \frac{N_A}{M} \tag{3.11}$$

where σ is the scattering cross-section, ρ_n is the number density, i.e. the number of elements per volume, ρ_m is the mass density, N_A is Avogadro's number and M is the molar mass. Since the attenuation coefficient μ relates to the likelihood of scattering, then so does the scattering cross-section. We already know that $\mu = \lambda^{-1}$ has units of inverse length and ρ_n has units of inverse length to the third power. This means that σ must have units of length squared. Indeed, the scattering cross-section can be interpreted as the effective surface area for collision as seen by the electrons. The addition rule shown in Eq. 3.10 can now be expressed in terms of scattering cross-sections as follows,

$$\frac{1}{\lambda(E)} = \sum_i \frac{1}{\lambda_i(E)} = \sum_i \sigma_i(E)\rho_{n,i} \tag{3.12}$$

For completeness we mention the discrete probability p_i for a particular event, which is determined by,

$$p_i(E) = \frac{\sigma_i(E)\rho_{n,i}}{\sum_j \sigma_j(E)\rho_{n,j}} = \frac{\lambda(E)}{\lambda_i(E)} \tag{3.13}$$

We now introduce the concept of differential scattering cross-sections. We are free to define a probability density, from which we not only deduce the likelihood of scattering in general but also, for example, in which direction. The latter can be expressed as follows,

$$\sigma_T(E) = \int_0^{2\pi} \int_0^{\pi} \frac{\partial\sigma(E)}{\partial\Omega} \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi \tag{3.14}$$

where σ_T is the total scattering cross-section and $\partial\sigma/\partial\Omega$ is the angular differential scattering cross-section with units of length squared per steradian. The normalized angular probability density function for scattering is then

given by,

$$p(E,\theta) = \frac{2\pi}{\sigma_T(E)} \frac{\partial \sigma(E)}{\partial \theta} \tag{3.15}$$

where we have assumed that the scattering is azimuthally symmetric. A random scattering angle is then obtained, similar to what is done with Eq. 3.7, by solving the integral equation,

$$U = \int_0^{\theta} p\big(E,\theta'\big)\mathrm{d}\theta' \tag{3.16}$$

for the scattering angle θ, where U is a random number with uniform distribution.

We would like to formulate the details and remaining theory on electron scattering by considering the following experiment. Suppose that we have a half-infinite sample of a perfectly conducting material, which is homogeneous with respect to the scattering likelihood of electrons. We expose the sample with a constant beam of primary electrons with energy E. At first, we will assume that the primary electron beam is perfectly monochromatic and infinitely sharp. We also assume that the material-vacuum interface plays no role in the interaction. We would like to study the electrons escaping the material again. To that order, let us consider an analytical model based on the Beer-Lambert law at first, which gives us the probability density that an electron with energy E at depth z travels to the interface and escapes,

$$p(E,z) \sim \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{z}{\lambda(E)\cos\theta}\right) \sin\theta \,\mathrm{d}\theta \mathrm{d}\phi \tag{3.17}$$

where we have expressed the probability density function conveniently in an angular-based coordinate system, see Fig. 3.2. Notice that expression Eq. 3.17 assumes that the emission of electrons at depth z is isotropic. So far we only considered the escape of the electrons from the sample, but what about the distribution of electrons inside the sample? We can create a more general version by introducing a probability density function to the integrand. It is explained in words as follows. Upon exposure with a primary electron beam of energy E, the electrons start to scatter around inside the

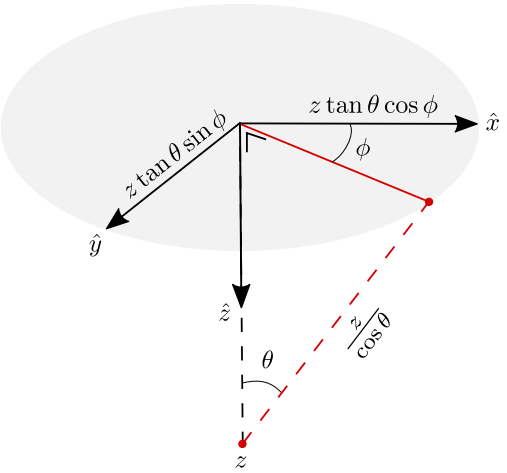


Figure 3.2: The angular coordinate system for the scattering of electrons towards the vacuum interface is shown. The probability that an electron at depth z escapes to the surface in the direction of (θ, ϕ) is calculated by using the Beer-Lambert law for the effective distance $z/\cos\theta$.

material. Due to the inelastic collisions, we expect to find electrons with an energy less than E within the sample. The general distribution is therefore a multivariate function which not only depends on the primary energy E and depth z, but also on the energy of the electrons inside the sample E' and the direction (θ, ϕ) in which the electrons are moving towards the interface. A more general case is therefore written as follows,

$$p(E,z) \sim \int_0^E \int_0^{\frac{\pi}{2}} S\big(\theta, z, E', E\big) \exp\left(-\frac{z}{\lambda(E')\cos\theta}\right) \sin\theta \,\mathrm{d}\theta \mathrm{d}E' \qquad (3.18)$$

where we have, without loss of generality, applied azimuthal symmetry for a homogeneous material. The first term $S(\theta, z, E', E)$ in the innermost integrand gives us the probability to find an electron at depth z with energy E' moving in the direction of θ. The second term (exponential) gives the probability that this particular electron reaches the interface and escapes. We must emphasize that Eq. 3.18 has severe limitations in view of reality. First of all, electrons can only move in a straight line to the interface and escape the material without any intermediate collisions. Second, energy-momentum transfers between the electron and the material are not considered. There are no interface effects included (absorption, reflection,

transmission, interactions with surface plasmons, ...). In view of these limits, we will restrict ourselves for the moment to elastic backscattered electrons only. The elastic backscattered electrons are the (primary) electrons that have penetrated the material and then directly scattered elastically out of the material again without losing energy. Furthermore, we assume that the backscattering of electrons occurs without preference for any particular direction. This implies that the distribution function simplifies to

$$S\big(\theta, z, E', E\big) \sim \delta\big(E' - E\big) \exp\left(-\frac{z}{\lambda(E)}\right) \tag{3.19}$$

which actually is the Beer-Lambert law for the probability that a primary electron with energy E reaches a depth of z. By substitution, we arrive at the following integral,

$$p(E, z) \sim \int_0^{\frac{\pi}{2}} \exp\left(-\frac{z}{\lambda(E)} \frac{\cos\theta + 1}{\cos\theta}\right) \sin\theta \, \mathrm{d}\theta \tag{3.20}$$

We now determine the cumulative probability to quantify the likelihood that backscatters escape from within a certain depth,

$$p(E, Z \leq z) = \frac{\int_0^z p(E, z') \, \mathrm{d}z'}{\int_0^\infty p(E, z') \, \mathrm{d}z'} = \frac{\int_0^z p(E, z') \, \mathrm{d}z'}{\lambda(E)(1 - \ln 2)} \tag{3.21}$$

The cumulative probability function is shown in Fig. 3.3 in the dimensionless variable $z/\lambda(E)$. As a rule of thumb, we can see that approximately 95% of the backscattered electrons escape from within a distance of $\lambda(E)$. Now that we have a rule of thumb for elastic backscatters, let us consider all the other electrons emitted from the sample. In order to proceed, we must somehow determine the function $S(\theta, z, E', E)$ in Eq. 3.18 for a more general case. We have already seen that $\lambda(E)$ can be determined from the differential scattering cross-sections. In addition, it turns out that the function S can also be determined from the differential scattering cross-sections by using a method which we will introduce in a moment. In fact, we will show that some of the aforementioned shortcomings related to Eq. 3.18 can be resolved.

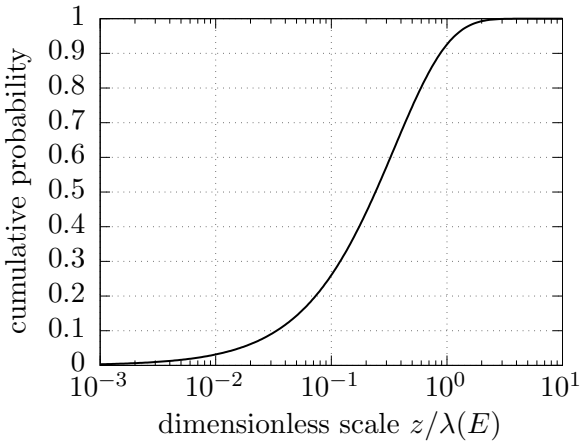


Figure 3.3: The cumulative probability for backscattered electrons to escape from a half-infinite sample is shown.

This brings us to describe the rigorous Monte-Carlo method for electron transport. The idea is to take the Beer-Lambert law on a per event basis for every electron. The moment an electron enters the material, we will sample the scattering probability and draw a random distance. The trajectory of the electron is updated accordingly. In the case of inelastic events, we will create new particles which will be tracked as well. The new particles could, in turn, create other particles. The net result of that is a cascading process, where we start with a single primary electron and end with a nested tree of descendants. Because of the latter, the rigorous Monte-Carlo method, where each descendant is tracked individually, is known to be notoriously slow. A detailed discussion on performance (and a solution to lengthy computation times) is given in chapter 4. Nevertheless, additional physical models can be included whenever an electron reaches the surface prior to escaping the material. For example, a quantum mechanical transmission and reflection model will be introduced in Section 3.5. The function S simply follows from the distribution of electrons inside the material as the rigorous Monte-Carlo method is applied. It is important to note that the rigorous Monte-Carlo method requires the differential scattering cross-sections for every relevant scattering event. The upcoming sections are therefore devoted to the determination of the differential scattering cross-sections.

3.2 Elastic electron scattering

In an elastic event, the electron is deflected by an angle θ without changing the energy. In reality, there can be a certain transfer of energy between the electron and a target. However, we neglect the recoil energy as the mass of the electron is orders of magnitude smaller than that of a typical target. We will concern ourselves with the determination of the elastic angular differential scattering cross-sections shown in Eq. 3.14. The elastic differential scattering cross-sections can be calculated ab-initio for all of the atomic elements of the periodic table, see for example Ref. 37. Let us outline the details of the calculation, which involves the elastic scattering of an electron wave against a central scattering potential. The latter is a textbook example of partial wave analysis (PWA), which is typically based on the Schrödinger equation and hence does not account for relativistic effects. Interestingly, and as counter intuitive as it may seem, relativistic effects are also important at lower energies. The most important effect is the spin-orbit coupling. Relativistic effects are accounted for in the Dirac equation, which in the system of natural units, i.e. $\hbar = c = m_e = 1$, reads,

$$[\boldsymbol{\alpha} \cdot \boldsymbol{p} + \beta + V(r)]\varphi = W\varphi \qquad (3.22) \quad \text{(A.U.)}$$

where W is the total energy of the electron in units of mc^2, the distance r is measured in units of $\hbar/mc$,

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & -\mathrm{I} \end{pmatrix} \qquad (3.23) \quad \text{(A.U.)}$$

In the latter, I is the 2×2 identity matrix and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli matrices,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad (3.24) \quad \text{(A.U.)}$$

It is not our intention to duplicate literature, but for the sake of completeness, let us summarize the method of PWA for the relativistic Dirac

equation. At first, the Dirac Eq. 3.22 is reduced to a single first-order differential equation,[37, 38, 39, 40]

$$\frac{\mathrm{d}\phi_l^{\pm}(r)}{\mathrm{d}r} = \frac{k^{\pm}}{r}\sin\left(2\phi_l^{\pm}(r)\right) - \cos\left(2\phi_l^{\pm}(r)\right) + W - V(r) \qquad (3.25) \quad \text{(A.U.)}$$

where $\phi_l^{\pm}$ is a function related to the phase shifts and the plus- and minus-sign represent the spin-up and spin-down case respectively. In particular, $k^{+} = -l-1 = -\left(j+\frac{1}{2}\right)$ and $k^{-} = l = j+\frac{1}{2}$. For each quantum number l, the phase shifts for spin-up and spin-down are calculated as follows,

$$\tan\delta_l^{\pm} = \frac{\kappa j_{l+1}(\kappa r) - j_l(\kappa r)\left[(W+1)\tan\phi_l^{\pm} + (1+l+k^{\pm})/r\right]}{\kappa n_{l+1}(\kappa r) - n_l(\kappa r)\left[(W+1)\tan\phi_l^{\pm} + (1+l+k^{\pm})/r\right]} \qquad (3.26) \quad \text{(A.U.)}$$

where $j_l(\kappa r)$ is the spherical Bessel function, $n_l(\kappa r)$ is the spherical Neumann function, $\kappa^2 = W^2 - 1$ and $\phi_l^{\pm}$ is the limit of $\phi_l^{\pm}(r)$ as $r \to \infty$. Notice that κ is the relativistic kinetic energy. The differential scattering cross-section for an unpolarized beam of electrons is now calculated as follows,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |f(\theta)|^2 + |g(\theta)|^2 \qquad (3.27) \quad \text{(A.U.)}$$

where the direct scattering amplitude $f(\theta)$ and spin-flip scattering amplitude $g(\theta)$ are determined by,

$$f(\theta) = \frac{1}{2i\kappa}\sum_{l=0}^{\infty}\left[(l+1)\left(\exp\left(2i\delta_l^{+}\right)-1\right) + l\left(\exp\left(2i\delta_l^{-}\right)-1\right)\right]P_l(\cos\theta) \qquad \text{(A.U.)} \quad (3.28)$$

$$g(\theta) = \frac{1}{2i\kappa}\sum_{l=1}^{\infty}\left[\exp\left(2i\delta_l^{-}\right) - \exp\left(2i\delta_l^{+}\right)\right]P_l(\cos\theta) \qquad (3.29) \quad \text{(A.U.)}$$

In practice, the details of the calculation must be found in the choices that can be made for the interaction potential $V(r)$.

We could calculate the elastic scattering cross-sections ourselves, provided that we have the interaction potential at hand. Fortunately, a program called ELSEPA exists which does this exactly for us.[40] In addition, the ELSEPA program includes sophisticated physical models for the inter-

action potential. We will use the ELSEPA program in the remainder of this thesis for the calculation of elastic scattering cross-sections. To provide more insight, we would like to discuss the optical-model potential which is used by that program,

$$V(r) = V_{\mathrm{st}}(r) + V_{\mathrm{ex}}(r) + V_{\mathrm{cp}}(r) - iW_{\mathrm{abs}}(r) \qquad (3.30) \quad \text{(A.U.)}$$

Let us briefly mention each of the terms in the optical-model potential,[40]

- The electrostatic potential, $V_{\mathrm{st}}(r)$, accounts for the interaction with the electron and nuclear charge distributions of the atom. The electrostatic potential is defined as

$$V_{\mathrm{st}}(r) = -e\left[\varphi_n(r) + \varphi_e(r)\right] \qquad (3.31) \quad \text{(A.U.)}$$

  where φ_n is the contribution from the nucleus and φ_e is the contribution from the electron cloud. The ELSEPA program uses the Fermi distribution for the nuclear charge and numerical Dirac-Fock densities obtained from density functional calculations by default.

- The exchange potential, $V_{\mathrm{ex}}(r)$, accounts for the indistinguishability of the incident electron from the electrons of the atom. The ELSEPA program uses the Furness-McCarthy exchange potential by default.

- The correlation-polarization potential, $V_{\mathrm{cp}}(r)$, accounts for the polarization of the atomic charge as the electron approaches the atom. The net result is an electric field due to an induced dipole moment which attracts the electron. We use the implementation of the elaborate local density approximation (LDA) correlation-polarization potential in the ELSEPA program.

- The absorption potential, $W_{\mathrm{abs}}(r)$, accounts for a loss of particles from the elastic to the inelastic channels above the first excitation threshold. It is modeled by means of LDA, where it is assumed that the electron interacts with the electron cloud as if it is moving through a homogeneous electron gas.

We would like to emphasize that the optical-model potential given by Eq. 3.30 applies to free atoms only. The scattering of electrons by an atom bound to a crystal structure can effectively be described by the muffin-tin model. The electrostatic potential is defined as follows,

$$V_{\mathrm{st,mt}}(r) = V_{\mathrm{st}}(r) + V_{\mathrm{st}}(2r_{\mathrm{mt}} - r) \qquad (3.32) \quad \text{(A.U.)}$$

where $2r_{\mathrm{mt}}$ is the average distance between two neighboring atoms. What is assumed, is that the electrons of the neutral atom are confined to within a sphere of radius r_{mt} centered at the nucleus. This implies, for example, that for distances beyond r_{mt}, the electrostatic potential of a bound atom is significantly smaller than that of a free atom, see Fig. 3.4.

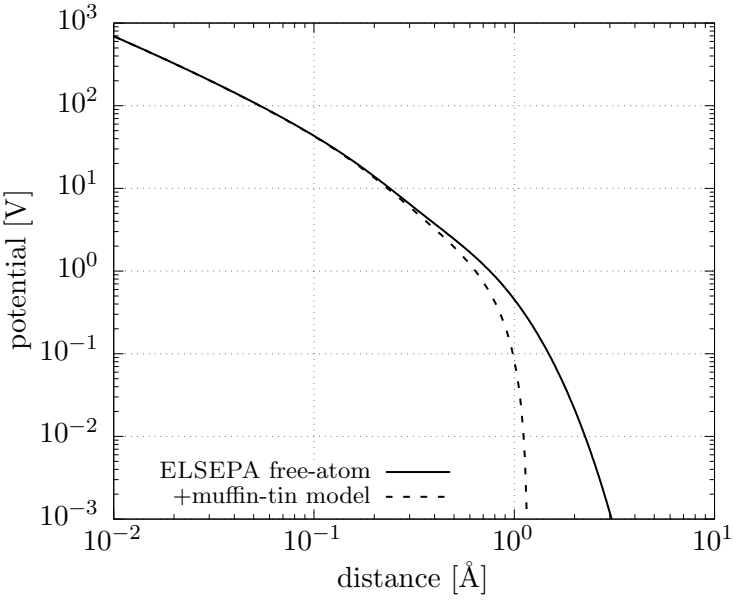


Figure 3.4: The electrostatic potential of silicon as used in the ELSEPA program is shown. The muffin-tin radius r_{mt} equals 1.175 Å. Note that for distances beyond r_{mt}, the electrostatic potential of a bound atom (muffin-tin model) is smaller than that of a free atom.

As a consequence, the differential scattering cross-section at smaller angles are lower than for a free atom. The optical-model potential for the muffin-tin approximation is defined as follows,

$$V_{\mathrm{mt}}(r) = \begin{cases} V_{\mathrm{st,mt}}(r) + V_{\mathrm{ex}}(r) + V_{\mathrm{cp}}(r) - iW_{\mathrm{abs}}(r) & \text{if } r \leq r_{\mathrm{mt}} \\ V_{\mathrm{st,mt}}(r_{\mathrm{mt}}) + V_{\mathrm{ex}}(r_{\mathrm{mt}}) + V_{\mathrm{cp}}(r) & \text{if } r > r_{\mathrm{mt}} \end{cases} \quad (3.33) \quad \text{(A.U.)}$$

Although this provides us a first step towards the scattering of electrons in a solid, the fact that the potential function is defined as rotational symmetric, and hence only depends on the scalar r, in Eq. 3.22 sets a limit to what can be achieved. For the purpose of demonstration, we have calculated the elastic MFP for silicon using the optical-model potential function, which is shown in Fig. 3.5. The specific models used for the optical-model potential are the default of ELSEPA, except for the correlation-polarization potential, where we have used the LDA model. For each curve we have assumed a number density of approx. $5 \cdot 10^{22}$ silicon atoms per cubic centimeter. The refinements shown in Fig. 3.5 are cumulative from top to bottom in

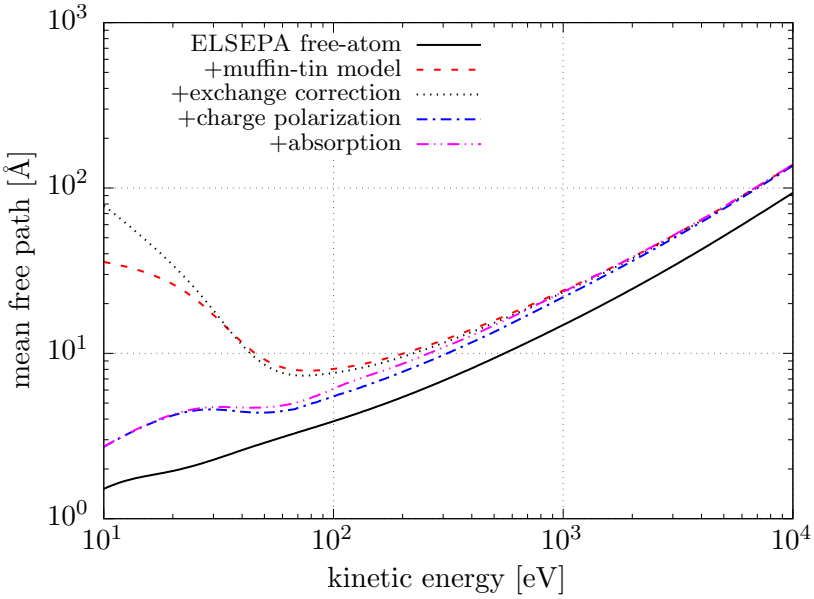


Figure 3.5: The elastic MFP of silicon for increasing level of detail to the free-atom potential is shown. Notice the sensitivity of the MFP to the potential as the kinetic energy decreases.

the legend. This means, for example, that the curve for charge polarization includes the exchange correction and the muffin-tin approximation as well. What we are trying to show here is that the elastic MFP becomes sensitive to the details of the potential as we approach lower electron energies. Below an energy of approx. 100 eV, the additional terms in the optical-model potential have an enormous impact on the elastic MFP. This, however, is opposite for higher electron energies, where we can see that none (except for muffin-tin) really have a significant effect on the elastic MFP. The effect on the angular probability of scattering can be seen in Fig. 3.6, where we show the differential scattering cross-section at an energy of 100 eV for various models of the optical-model potential. Here we can see that the

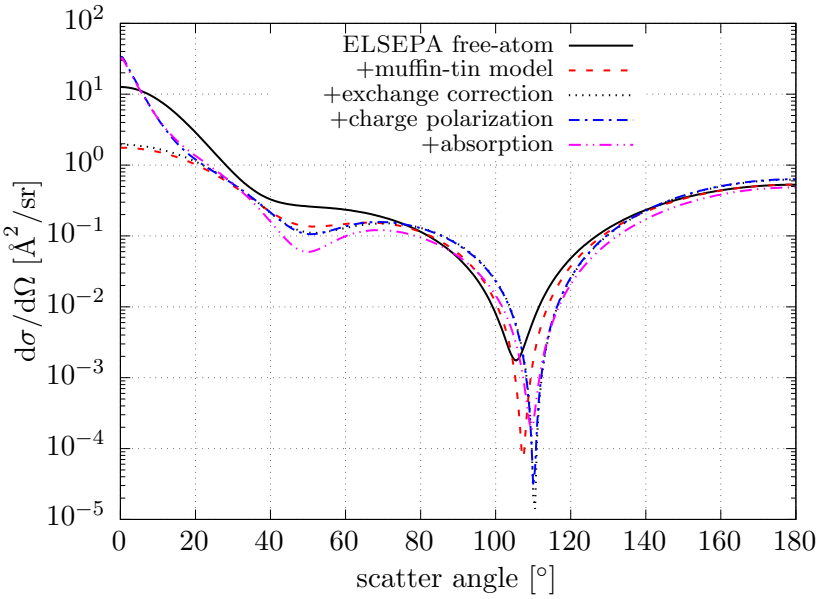


Figure 3.6: The elastic differential scattering cross-section of silicon for $E = 100$ eV is shown. The scattering angle is taken with respect to the incident direction of the electron.

differences associated with the terms of the optical-model potential are relatively small perturbations, but integrate out to larger differences in the elastic MFP. We conclude that the optical-model muffin-tin potential given by Eq. 3.33 fails to accurately describe the interaction of electrons with the atoms bound to a crystal lattice at lower energies. A more accurate

result is obtained by a full three-dimensional function for the nuclear and electron charge density. In view of the latter, we identify the following two additional effects originating from the periodicity of the lattice:

- Coherent scattering resulting into constructive and/or destructive interference of the scattered electron wave (Bragg diffraction).
- The electron wavefunction is expected to behave like a Bloch-electron with the periodicity of the lattice.

Unfortunately, the potential in Eq. 3.22 only depends on the scalar r and thus it is not possible to have a full three-dimensional analysis in the present calculation. What we can do however, is predict the effect of coherent scattering with free atoms (with radial potential function) in a lattice by introducing to Eq. 3.27 a term for a spatial dependent phase shift,[40]

$$F(\theta) = \sum_i \exp\left(i\boldsymbol{q}\cdot\boldsymbol{r_i}\right) f_i(\theta) \tag{3.34}$$ (A.U.)

$$G(\theta) = \sum_i \exp\left(i\boldsymbol{q}\cdot\boldsymbol{r_i}\right) g_i(\theta) \tag{3.35}$$ (A.U.)

where $\boldsymbol{q}$ relates to the momentum transfer, $\boldsymbol{r_i}$ is the position vector of the nucleus of the i-th atom and $f_i(\theta)$ and $g_i(\theta)$ are the scattering amplitudes for the free atom. To see the difference, we calculate the differential scattering cross-sections by using the lattice (diamond) structure of silicon (see Fig. 3.7) and vary the number of unit-cells. We consider a single unit cell, three unit cells per dimension (27 unit cells in total) and five unit cells per dimension (125 unit cells in total). The result of that in units Angstrom per steradian per silicon atom is shown in Fig. 3.8, and the effect on the elastic MFP is shown in Fig. 3.9. Let us discuss the results of the aforementioned figures. First of all, note that the differential scattering cross-section for coherent scattering oscillates around the curve for the free-atom cross-section. This comes as no surprise: the curves for coherent scattering are based on scattering with free atoms located at the lattice sites of the unit cell shown in Fig. 3.7. The oscillatory behavior comes from constructive and destructive interference of the scattered electronwave. Furthermore,

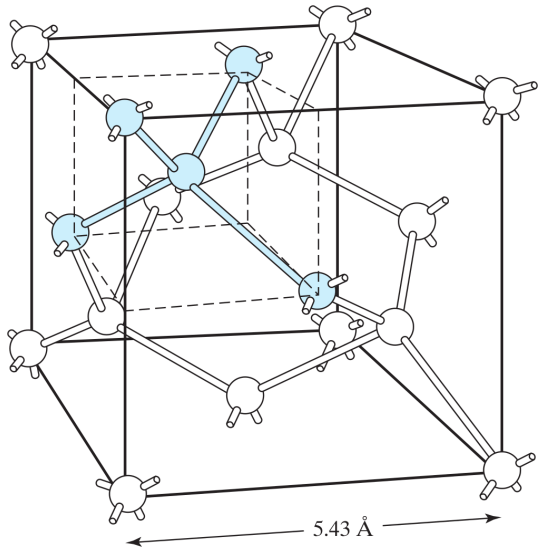


Figure 3.7: The diamond structure of a silicon unit cell is shown. The atoms are tetrahedrally bonded by sharing valence electrons. The lattice constant is 5.43 Å. The number density is approx. $5 \cdot 10^{22}$ silicon atoms per cubic centimeter. Image taken from the book of Shockley.[41]

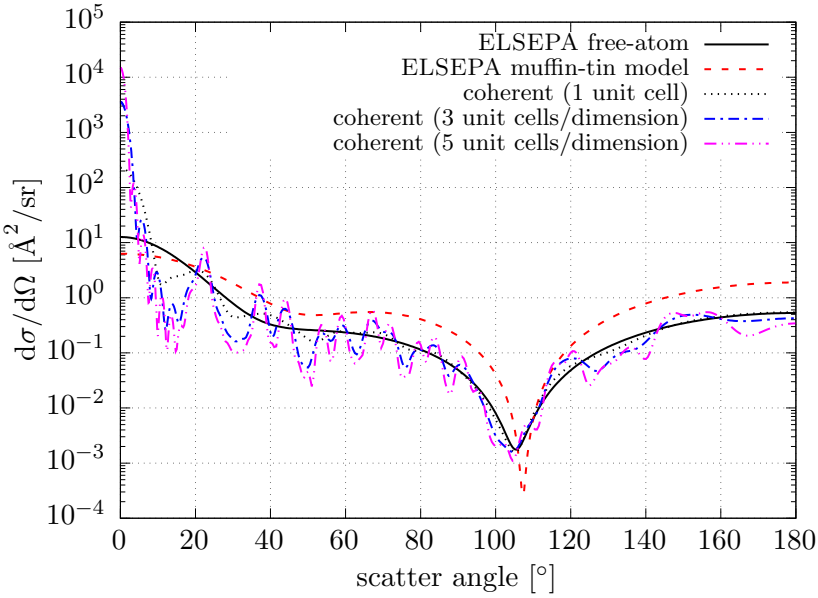


Figure 3.8: The effect of coherent scattering in silicon on the differential scattering cross-section is shown. The electron has an incident energy of $E = 100\,\mathrm{eV}$.

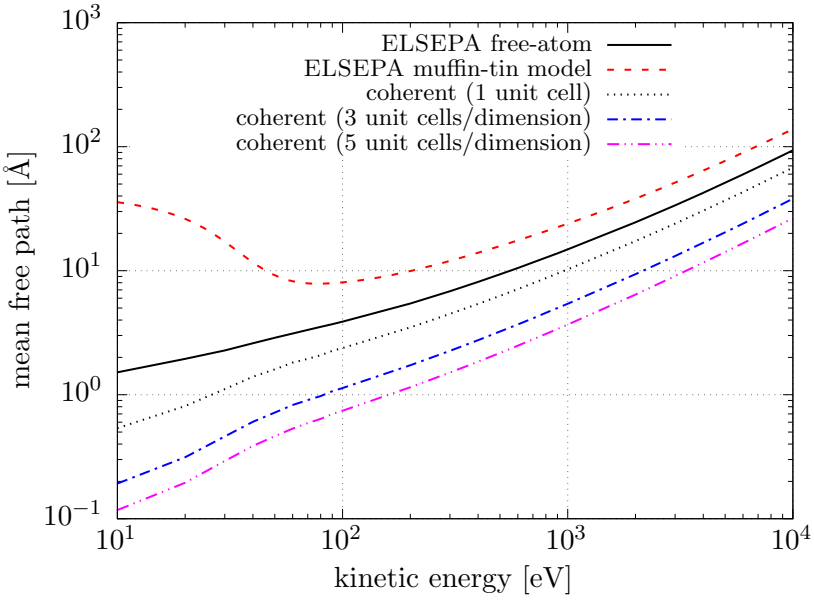


Figure 3.9: The effect of coherent scattering in silicon on the elastic MFP.

note that the elastic MFPs in Fig. 3.9 show a peculiar behavior: the elastic MFP is decreasing significantly as we increase the number of unit cells and does not seem to convergence. We suspect this is an artificial effect attributable to the simplistic view of scattering with free atoms located at the lattices sites. We are therefore tempted to follow the advice given in the ELSEPA manual and accept the muffin-tin model as a first order approximation to solid state effects. Nevertheless, we cannot ignore the fact that the real lattice structure is not accounted for in the muffin-tin model. We argue that the most acceptable solution is to include both effects: Bragg diffraction and a modified electron density function to account for the aggregation. We will come back to this issue at a later stage in the thesis.[1] In any case, we should, for the sake of the validity of single-atom scattering, only consider electron energies where the electron wavelength $\ll$ the lattice constant. The electron wavelength can be estimated by using the formula

[1] To be more precise, we will make the connection to experiments once we have a fully operational electron-matter interaction simulator.

of de Broglie,

$$\lambda_{e^-} = \frac{h}{p} = \frac{h}{\sqrt{2mE\left(1 + \frac{E}{2mc^2}\right)}} \approx \frac{h}{\sqrt{2mE}} \tag{3.36}$$

where E is the kinetic energy of the electron and the approximate result applies to the non-relativistic case. The electron wavelength as a function of kinetic energy is shown in Fig. 3.10. We have added a typical lattice

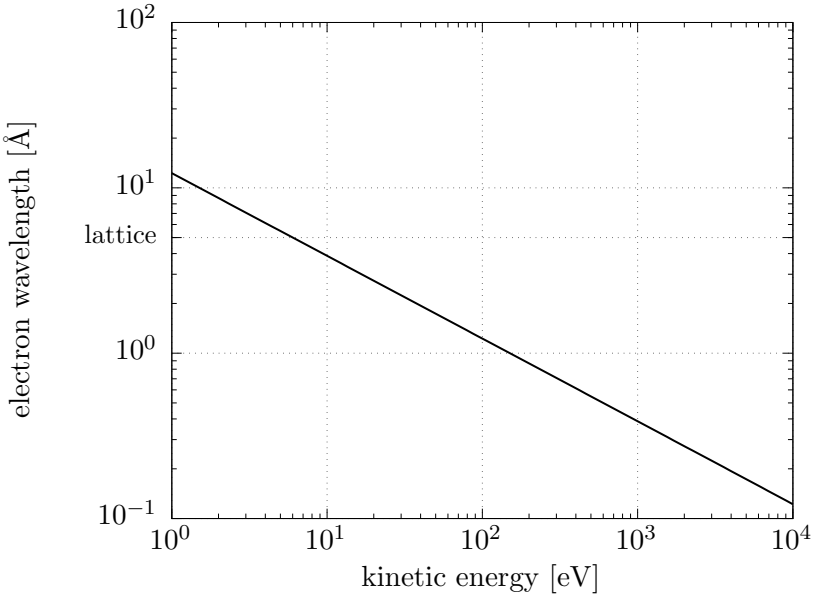


Figure 3.10: The electron wavelength as a function of kinetic energy is shown. We have included an additional tic to the y-axis for a typical lattice constant of 5 Å. Notice that for low kinetic energies, the electron wavelength approaches the lattice constant.

constant of approx. 5 Å to the vertical axis. We conclude that, for energies smaller than 10 eV, the electron wavelength becomes on the order of the lattice constant. This means that the models of elastic scattering presented in this section should no longer be used for that regime. There is a special Section 3.4 in this chapter devoted to the regime of low-energy scattering.

We would like to conclude this section with a graph (Fig. 3.11) of the elastic MFP for various materials: aluminum (metal), silicon (semiconductor), gold (metal) and silica (insulator). The MFP of each curve is de-

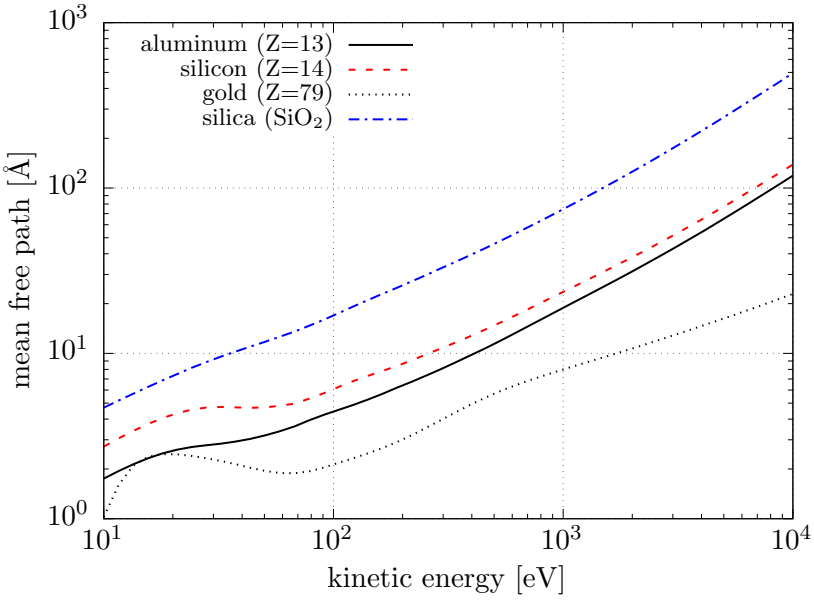


Figure 3.11: The elastic MFP for various materials is shown. Each curve is obtained by using the *full* options of ELSEPA. Details for the compound silica are given in the main text.

termined using the *full* options of ELSEPA. For the compound silica, we have used the (weighted) addition rule ($SiO_2 \to \frac{1}{3}Si + \frac{2}{3}O$) to approximate the total effective elastic scattering cross-section.[2] For the MFP we have used the actual number density associated with silica, rather than weighted average number density of the constituents.

3.3 Inelastic electron scattering

In this section, we concern ourselves with inelastic scattering events. In contrast to the theory of elastic events from the previous section, the inelastic event is associated with energy-momentum transfer. We will pursue the idea of a dielectric in the sense that the presence of an electron perturbs the equilibrium charge distribution of the material. Consequently, the material becomes polarized and induces an irreversible electromagnetic stopping force on the electron. Starting from first principles, we derive the

[2]There is no muffin-tin radius r_{mt} for oxygen. Instead, we have used the muffin-tin radius of silicon for the scattering cross-sections of oxygen.

relevant dielectric function expressions for electron energy losses in matter. We would like to mention that the derivation closely follows the elaborate work of Francesc Salvat-Pujol.[42] However, instead of a more complicated generic case, we present a derivation specific for bulk losses. The idea is that we would like to provide sufficient theoretical background for dielectric losses to at least understand the key assumptions and physical principles.

Suppose that we have an electron inside a homogeneous and isotropic medium of infinite size. The electric field is determined by Maxwell's equations, which we choose to express in the Gaussian system of units. In addition, we choose to work in the Coulomb gauge, which implies that the vector potential $\boldsymbol{A}(\boldsymbol{r},t)$ is transverse. The time-dependent electric field $\boldsymbol{\mathcal{E}}(\boldsymbol{r},t)$ is then expressed in terms of a scalar potential $\varphi(\boldsymbol{r},t)$ and vector potential $\boldsymbol{A}(\boldsymbol{r},t)$ as follows,

$$\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) = -\nabla\varphi(\boldsymbol{r},t) - \frac{1}{c}\frac{\partial \boldsymbol{A}(\boldsymbol{r},t)}{\partial t} \tag{3.37}$$ (G.S.)

We assume that the material responds linearly to the motion of the electron as it traverses through the material. In other words, the electric displacement $\boldsymbol{\mathcal{D}}(\boldsymbol{r},t)$ relates to the electric field $\boldsymbol{\mathcal{E}}(\boldsymbol{r},t)$ in the following way,

$$\boldsymbol{\mathcal{D}}(\boldsymbol{r},t) = \int_{-\infty}^{\infty}\int_{\mathbb{R}^3} \epsilon\big(\boldsymbol{r}-\boldsymbol{r}',t-t'\big)\boldsymbol{\mathcal{E}}\big(\boldsymbol{r}',t'\big)\mathrm{d}^3\boldsymbol{r}'\mathrm{d}t' \tag{3.38}$$ (G.S.)

We recognize the integral as a convolution and therefore, by the theorem of convolution, transforms to a plain multiplication in Fourier space,

$$\boldsymbol{\mathcal{D}}(\boldsymbol{q},\omega) = \epsilon(\boldsymbol{q},\omega)\boldsymbol{\mathcal{E}}(\boldsymbol{q},\omega) \tag{3.39}$$ (G.S.)

where $\boldsymbol{q}$ and ω are the Fourier conjugate variables with respect to position and time respectively and $\epsilon(\boldsymbol{q},\omega)$ is the dielectric function of the material. We assume that the speed of the electron is much less than the speed of light such that relativistic effects can be neglected. This implies that the transverse components of the electromagnetic field can be ignored and thus

we set the vector potential equal to zero,

$$\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) \approx -\nabla\varphi(\boldsymbol{r},t) \tag{3.40}$$ (G.S.)

With the assumptions at hand, we determine our first quantity of interest: the stopping power (SP), i.e.,

$$SP(E) = -\frac{\mathrm{d}E}{\mathrm{d}s} = -\frac{1}{v}\frac{\mathrm{d}E}{\mathrm{d}t} \tag{3.41}$$ (G.S.)

where we have used that $\mathrm{d}s = v\mathrm{d}t$. In order to account for the influence of the material, we must consider the *total* energy, viz. kinetic plus potential, of the electron,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}m\boldsymbol{v}\cdot\boldsymbol{v} - e\varphi_{\mathrm{ind}}(\boldsymbol{r},t)\right]_{\boldsymbol{r}=\boldsymbol{v}t} \tag{3.42}$$ (G.S.)

where $\varphi_{\mathrm{ind}}(\boldsymbol{r},t)$ is the induced potential on the electron. The time derivative can be worked out to give,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \boldsymbol{v}\cdot\boldsymbol{F}_{\mathrm{ind}}(\boldsymbol{r},t) - e\frac{\mathrm{d}\varphi_{\mathrm{ind}}(\boldsymbol{r},t)}{\mathrm{d}t}\bigg|_{\boldsymbol{r}=\boldsymbol{v}t} = -e\frac{\partial\varphi_{\mathrm{ind}}(\boldsymbol{r},t)}{\partial t}\bigg|_{\boldsymbol{r}=\boldsymbol{v}t} \tag{3.43}$$ (G.S.)

where $\boldsymbol{F}_{\mathrm{ind}}(\boldsymbol{r},t)$ is the induced force on the electron, which has canceled by virtue of applying the multivariable chain rule to the induced potential. It comes as no surprise that the change in energy of the electron is ultimately determined by the explicit time dependence of the scalar potential. We now need to figure out how the induced potential $\varphi_{\mathrm{ind}}(\boldsymbol{r},t)$ relates to the dielectric function $\epsilon(\boldsymbol{q},\omega)$ of the material. Evidently we find that the induced potential follows from Maxwell's equations for the dielectric displacement,

$$\nabla\cdot\boldsymbol{\mathcal{D}}(\boldsymbol{r},t) = 4\pi\rho(\boldsymbol{r},t) \tag{3.44}$$ (G.S.)

The charge density $\rho(\boldsymbol{r},t)$ on the RHS must account for the electron, which we model as a point charge using the Dirac delta,

$$\rho(\boldsymbol{r},t) = -e\delta(\boldsymbol{r}-\boldsymbol{v}t) \tag{3.45}$$ (G.S.)

The connection to the dielectric function is made by first transforming Eq. 3.44 to Fourier space,

$$i\boldsymbol{q}\cdot\boldsymbol{\mathcal{D}}(\boldsymbol{q},\omega) = -2e\delta(\boldsymbol{q}\cdot\boldsymbol{v}-\omega) \tag{3.46}$$ (G.S.)

where the Fourier transform[3] of the charge density of the point charge is evaluated as,

$$\rho(\boldsymbol{q},\omega) = -\frac{e}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{\mathbb{R}^3}\delta(\boldsymbol{r}-\boldsymbol{v}t)\exp\left[-i(\boldsymbol{q}\cdot\boldsymbol{r}-\omega t)\right]\mathrm{d}^3\boldsymbol{r}'\mathrm{d}t \tag{3.47}$$ (G.S.)

$$= -\frac{e}{(2\pi)^2}\int_{-\infty}^{\infty}\exp\left[-it(\boldsymbol{q}\cdot\boldsymbol{v}-\omega)\right]\mathrm{d}t \tag{3.48}$$ (G.S.)

$$= -\frac{e}{2\pi}\delta(\boldsymbol{q}\cdot\boldsymbol{v}-\omega) \tag{3.49}$$ (G.S.)

The expression for the electric field $\boldsymbol{\mathcal{E}}(\boldsymbol{q},\omega)$, in terms of the dielectric function $\epsilon(\boldsymbol{q},\omega)$, is found by substituting Eq. 3.39 into Eq. 3.46. The result of that yields the following expression for the electric field in Fourier space,

$$\boldsymbol{\mathcal{E}}(\boldsymbol{q},\omega) = i\frac{2\boldsymbol{q}e}{q^2}\frac{1}{\epsilon(\boldsymbol{q},\omega)}\delta(\boldsymbol{q}\cdot\boldsymbol{v}-\omega) \tag{3.50}$$ (G.S.)

What we actually have found is an expression for the *total* electric field $\boldsymbol{\mathcal{E}}(\boldsymbol{q},\omega)$. What we need, however, is the *induced* electric field $\boldsymbol{\mathcal{E}}_{\mathrm{ind}}(\boldsymbol{q},\omega)$. In order to obtain the induced electric field, the self-contribution of the electron must be subtracted from the total electric field,

$$\boldsymbol{\mathcal{E}}_{\mathrm{ind}}(\boldsymbol{q},\omega) = \boldsymbol{\mathcal{E}}(\boldsymbol{q},\omega) - \boldsymbol{\mathcal{E}}_0(\boldsymbol{q},\omega) = i\frac{2\boldsymbol{q}e}{q^2}\delta(\boldsymbol{q}\cdot\boldsymbol{v}-\omega)\left[\frac{1}{\epsilon(\boldsymbol{q},\omega)}-1\right] \tag{3.51}$$ (G.S.)

where $\boldsymbol{\mathcal{E}}_0(\boldsymbol{q},\omega)$ is the electric field when moving through vacuum. The expression for the induced potential is obtained via the Fourier transform of Eq. 3.40,

$$\boldsymbol{\mathcal{E}}_{\mathrm{ind}}(\boldsymbol{q},\omega) = -i\boldsymbol{q}\varphi_{\mathrm{ind}}(\boldsymbol{q},\omega) \tag{3.52}$$ (G.S.)

which gives us the following expression for the induced potential in Fourier

[3] We use symmetric normalization factors of $\frac{1}{2}\sqrt{2}$ for the forward and inverse Fourier transforms.

space,

$$\varphi_{\text{ind}}(\boldsymbol{q},\omega) = -\frac{2e}{q^2}\delta(\boldsymbol{q}\cdot\boldsymbol{v}-\omega)\left[\frac{1}{\epsilon(\boldsymbol{q},\omega)}-1\right] \tag{3.53}$$ (G.S.)

We are now ready to go back to our expression for the stopping power,

$$SP(E) = \frac{e}{v}\frac{\partial\varphi_{\text{ind}}(\boldsymbol{r},t)}{\partial t}\bigg|_{\boldsymbol{r}=\boldsymbol{v}t} \tag{3.54}$$ (G.S.)

and rewrite the term for the induced potential $\varphi_{\text{ind}}(\boldsymbol{r},t)$ as the inverse transform of $\varphi_{\text{ind}}(\boldsymbol{q},\omega)$, i.e.,

$$SP(E) = -i\frac{e}{v}\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\omega\int_{\mathbb{R}^3}\varphi_{\text{ind}}(\boldsymbol{q},\omega)\exp\left[it(\boldsymbol{q}\cdot\boldsymbol{v}-\omega)\right]\mathrm{d}^3\boldsymbol{q}\mathrm{d}\omega \tag{3.55}$$ (G.S.)

What remains to be done is: (1) substitute the expression for the induced potential, (2) transform the integral to spherical coordinates, (3) assume that $\epsilon(\boldsymbol{q},\omega) = \epsilon(q,\omega)$ and integrate out the azimuthal angle, (4) change of variables $\theta \to qv\cos\theta$ and work out the Dirac delta and finally (5) use that $\epsilon(q,\omega) = \epsilon^\star(-q,-\omega)$ to arrive at,

$$SP(E) = \int \omega\tau(E,\omega)\mathrm{d}\omega \tag{3.56}$$ (G.S.)

where $\tau(E,\omega)$ is proportional to the probability for an energy-loss between ω and $\omega+\mathrm{d}\omega$ per unit distance,

$$\tau(E,\omega) = \frac{e^2}{\pi E}\int\frac{\mathrm{d}q}{q}\mathrm{Im}\left[-\frac{1}{\epsilon(q,\omega)}\right] \tag{3.57}$$ (G.S.)

and the integration domains of both q and ω are determined by the kinematics of the inelastic collision. This is the celebrated expression for the SP of an electron from dielectric theory, see for example Refs. 43, 44, 42. Notice that we have used the non-relativistic expression to relate the velocity of the electron v to the kinetic energy E. In other words, the dispersion relation of a free electron is applied, which in turn defines the energy scale, see Fig. 3.12. Let us emphasize that the bending of the dispersion relation near the Brillouin zone boundary is neglected in the present model.

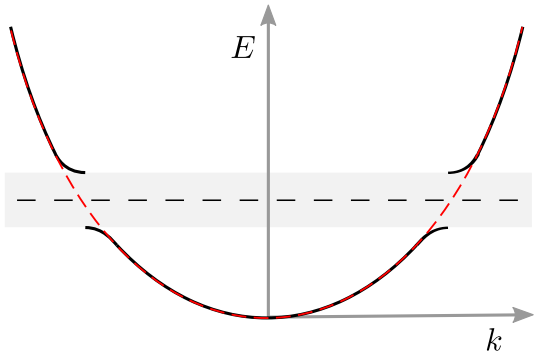


Figure 3.12: The schematic dispersion relation for electrons in a semiconductor is shown. The dashed red line corresponds to the free electron model, which is used in the dielectric function model. In reality, the dispersion relation bends near the Brillouin zone boundary (solid black line). The gray area in between is the forbidden energy gap. The dashed black line inside the bandgap represents the Fermi level.

A first order approximation for the latter is obtained by assuming that the bottom of the conduction band is purely parabolic,

$$E = E_{CB} + \frac{(\hbar k)^2}{2m^\star} \tag{3.58}$$

where E_{CB} is the energy at the bottom of the conduction band, $m^\star$ is the effective mass which determines the curvature of the band and $\hbar k$ defines the crystal momentum of the electron. What is assumed in the dielectric function model is that E_{CB} is zero and that the effective mass $m^\star$ is equal to the free electron mass m_e. This implies, for example, that Eq. 3.57 should only be used for energies E much larger than the Fermi level.

Another quantity of interest is the inelastic inverse MFP, which is obtained by directly integrating $\tau(E, \omega)$,

$$\frac{1}{\lambda(E)} = \int \tau(E, \omega) \mathrm{d}\omega \tag{3.59}$$ (G.S.)

Let us discuss the kinematics of the inelastic collision. It is more convenient to use natural units at this point, i.e. $\hbar = c = m_e = 1$, such that energy-loss ($\hbar\omega$) coincides with the Fourier variable ω and momentum transfer ($\hbar q$) coincides with the Fourier variable q. The momentum transfer $\boldsymbol{q}$ is deter-

mined as follows,

$$\boldsymbol{q} = \boldsymbol{p} - \boldsymbol{p}' \tag{3.60}$$ (A.U.)

where $\boldsymbol{p}$ is the wavevector before and $\boldsymbol{p}'$ is the wavevector after the collision. The lower and upper bound on the momentum transfer is determined by squaring the expression for momentum transfer,

$$q^2 = p^2 + p'^2 - 2p'p\cos\theta \tag{3.61}$$ (A.U.)

where θ is the polar scattering angle. In the non-relativistic limit, we obtain the following expression for the momentum transfer in terms of energy,

$$q^2 = 2E + 2(E-\omega) - 4\sqrt{E(E-\omega)}\cos\theta \tag{3.62}$$ (A.U.)

The maximum (q_+) and minimum (q_-) momentum transfer corresponds to a scattering angle of $\theta = \pi$ and $\theta = 0$ respectively and equals,

$$q_\pm = \sqrt{2E} \pm \sqrt{2(E-\omega)} \tag{3.63}$$ (A.U.)

Note that this last expression assumes that the energy-momentum transfer is the same as that for a free electron in vacuum.

Although we have defined the integration limits of Eq. 3.57, we still cannot calculate the inelastic scattering probability $\tau(E,\omega)$. The reason is that we need to know the "imaginary part of the inverse of the dielectric function". What is this complex function and how do we obtain it? The dielectric function can be written as a complex function in terms of a dispersive $\epsilon_1(q,\omega)$ and absorptive $\epsilon_2(q,\omega)$ term as follows,[45]

$$\epsilon(q,\omega) = \epsilon_1(q,\omega) + i\epsilon_2(q,\omega) \tag{3.64}$$ (A.U.)

The connection to observables is made through the optical limit $q \approx 0$,

$$\epsilon(0,\omega) = \epsilon(\omega) = (\eta(\omega) + i\kappa(\omega))^2 \tag{3.65}$$ (A.U.)

where the RHS is the complex index of refraction with the observables $\eta(\omega)$ as the refractive index and $\kappa(\omega)$ as the extinction coefficient. The observ-

ables are linked to the dielectric function in the optical limit by,

$$\epsilon_1(\omega) = \eta(\omega)^2 - \kappa(\omega)^2 \tag{3.66}$$ (A.U.)

$$\epsilon_2(\omega) = 2\eta(\omega)\kappa(\omega) \tag{3.67}$$ (A.U.)

The real (ϵ_1) and imaginary (ϵ_2) part of the complex dielectric function in the optical limit share a deep connection through the Kramers-Kronig relations,[46]

$$\epsilon_1(\omega) = \frac{2}{\pi}\mathcal{P}\int_0^\infty \frac{\omega'\epsilon_2(\omega')}{\omega'^2 - \omega^2}\mathrm{d}\omega' \tag{3.68}$$ (A.U.)

$$\epsilon_2(\omega) = -\frac{2\omega}{\pi}\mathcal{P}\int_0^\infty \frac{\epsilon_1(\omega') - 1}{\omega'^2 - \omega^2}\mathrm{d}\omega' \tag{3.69}$$ (A.U.)

where $\mathcal{P}$ is the Cauchy principal value. The Kramers-Kronig relations are based on the fundamental principle of causality applied to the optical response of a medium: effect cannot precede cause. The electron energy-loss function (ELF) shown in Eq. 3.57 is defined in terms of these observables as follows,

$$\mathrm{Im}\left[-\frac{1}{\epsilon(\omega)}\right] = \frac{\epsilon_2(\omega)}{\epsilon_1(\omega)^2 + \epsilon_2(\omega)^2} = \frac{2\eta(\omega)\kappa(\omega)}{\eta(\omega)^2 + \kappa(\omega)^2} \tag{3.70}$$ (A.U.)

The electron ELF can be determined, for example, from electron energy-loss spectroscopy (EELS). The idea is to measure the energy losses of the incident electrons after interaction with a particular target. The spectroscopy yields an electron spectrum in which the number of electrons as a function of energy-loss is given. Let us give a brief and generic description of a typical electron energy-loss spectrum. The first peak in an electron energy-loss spectrum is found at zero and is called the zero-loss peak. The zero-loss peak represents the fraction of transmitted electrons in TEELS or reflected (backscattered) electrons in REELS with negligible energy losses.[4] Next, a broader peak is found within approximately 50 eV away from the elastic-peak. This peak is the result of scattering with the outer-shell electrons and

[4]With negligible energy losses we mean no energy loss at all (pure elastic) or a few losses resulting from the interactions with phonons (quasi-elastic) for example.

scattering with plasmons. In practice, we may find two peaks for losses less than 50 eV: one for the bulk plasmons and one for the surface plasmons. For energy losses typically larger than 100 eV we find sharp edges which correspond to inner-shell electron excitations. Roughly speaking, each sharp edge corresponds with a measure of the binding energy of an inner-shell electron. In the acquisition of an electron energy-loss spectrum, we should only consider electrons with only one inelastic event in the interaction. In reality, multiple inelastic events cannot be ignored and hence form an intrinsic difficulty in the acquisition of an electron energy-loss spectrum. The electron energy-loss spectrum can be retrieved nevertheless by means of one or more deconvolution procedures. For example, see Refs. 47, 45 for a more thorough discussion on the acquisition of electron energy-loss spectra from measurements.

Let us discuss the electron ELF of silicon, which is shown in Fig. 3.13. How do we tell that this is the electron ELF of silicon? There are several characteristics in the electron ELF that reveal the identity of silicon. First

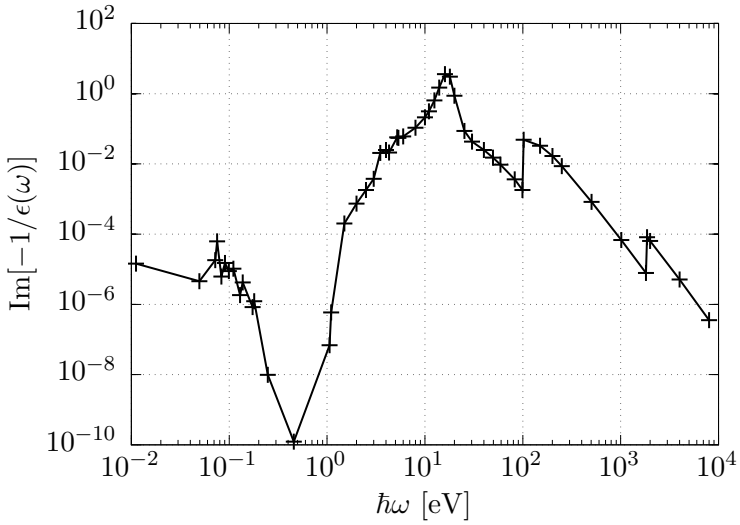


Figure 3.13: The electron ELF for silicon is shown. The markers are actual measurements, which are connected in the log-log figure by straight lines. The loss function is taken from Ref. 44.

of all, let us enumerate the binding energies of the electrons of silicon. The binding energies of the electrons for a free silicon atom can be found, for example, in the Livermore evaluated electron library (EEDL).[48] We have summarized, for the convenience of the reader, the binding energies of all 14 electrons of silicon from the Livermore database in Table 3.1. We would

electron shell	**occupancy**	**binding energy**
K (1s1/2)	2	1828.5 eV
L1 (2s1/2)	2	151.55 eV
L2 (2p1/2)	2	108.67 eV
L3 (2p3/2)	4	107.98 eV
M1 (3s1/2)	2	13.63 eV
M2 (3p1/2)	2/3	6.55 eV
M3 (3p3/2)	4/3	6.52 eV

Table 3.1: The binding energies for the electrons of each subshell of free silicon are shown. The values are taken from the Livermore evaluated electron library (EEDL).

like to emphasize once more that Table 3.1 applies to a free silicon atom only. This means that the lowest binding energies shown in Table 3.1 are of no interest to us because of solid state physics. Nevertheless, the binding energy of 1828.5 eV (associated to the two electrons in the K-shell) is identified by the rightmost sharp edge in Fig. 3.13. Unfortunately, the resolution of the electron ELF shown in Fig. 3.13 does not allow us to resolve the individual binding energies of the L-shell electrons. What is observed instead is that, in the vicinity of the binding energies of the electrons in the L-shell, the electron ELF decreases less rapidly when compared to the sharp edge of the K-shell. There is, in fact, another aspect in view of interaction with inner-shell electrons: not every subshell has an equal probability of electron-ionization. The scattering cross-section per subshell can also be found in the Livermore database, viz. Ref. 48. We have, once more for the convenience of the reader, made a graph of the electron-ionization scattering cross-sections. The corresponding scattering cross-sections of the K- and L-shell are shown in Fig. 3.14. Notice that the scattering cross-section, and hence the probability for electron-ionization, of the K-shell is about 2 to 3 orders less than the L-shell. Looking back at the electron ELF in Fig. 3.13, we see

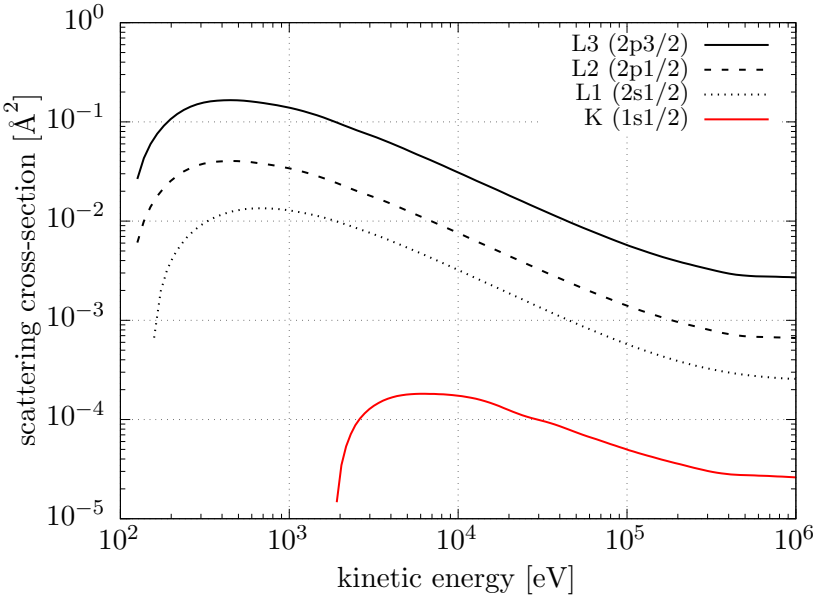


Figure 3.14: The electron-ionization scattering cross-section of the K- and L-shell for free silicon is shown. The curves are derived from the Livermore evaluated electron library (EEDL).

that the corresponding ratio between the K- and L-shell edge is of similar magnitude. We would like to point out the, rather satisfactory, agreement between the aforementioned characteristics of the *measured* electron ELF with the *ab-initio calculated* electron-ionization scattering cross-sections of the Livermore database! The reader is referred to Ref. 49 for intricate details on the calculation of electron-ionization scattering cross-sections. Another interesting feature of Fig. 3.13 is that near 1 eV, the electron ELF practically drops to zero. This drop to zero-loss relates to the band gap of silicon which is approx. 1.1 eV at room temperature. Below the band gap, however, we still see non-zero amplitudes in the electron ELF. What are these losses below the band gap? The answer is that below the band gap we have losses due to the interaction with phonons.[50] There are, in fact, two types of phonons: acoustic and optical phonons. The losses below the band gap in the electron ELF of Fig. 3.13 are due to optical phonons. The reason is that energy losses associated to acoustic phonons are on the order of meV and, hence, are not observed in the spectrum of Fig. 3.13. We will discuss

losses due to acoustic and optical phonons in more detail in an upcoming section on low-energy electron scattering. Another loss mechanism below the band gap is given by Ref. 51. In that article it is discussed that losses below the band gap could relate to transitions to exciton states[5] below the conduction band. This brings us to the following remark. The acquisition of the electron ELF, i.e. Fig. 3.13, covers quite a large domain of energy losses from various physical regimes. In fact, an electron ELF can range from a few tens of meV up to a MeV and beyond. It seems unlikely that an electron ELF for as much as nine orders is determined solely from EELS. The truth is indeed that, in practice, energy-loss measurements originating from different sources, such as optical measurements, EELS and atomic photo-absorption data resulting from X-ray interactions are combined to a single electron ELF, see for example Ref. 52. In this way, the determination of the electron ELF usually involves Kramers-Kronig analysis to connect a specific measurement of one observable to another optical observable.[45] A legitimate question rises: How do we know that the resulting energy-loss function is accurate when the different experiments are combined? The answer is that an energy-loss function is restricted by so called sum-rules. We will not discuss the sum-rules further in this thesis. Instead, we would like to refer to Refs. 43, 53, 54 for more discussion on the application of sum-rules to energy-loss functions.

The use of optical data gives us at least two advantages: (1) a connection to experiment through optical observables can be made, and (2) optical data for a variety of materials is readily available.[55, 56, 57, 58, 59] We would like to point out, however, that the connection to experiment is made through the optical limit, i.e. $q \approx 0$. What we need in view of Eq. 3.57, however, is a non-zero momentum-transfer energy-loss function. The connection to the optical limit for non-zero momentum transfers can be made in several ways. We will restrict ourselves to the connection made by Ashley,[43]

$$\mathrm{Im}\left[-\frac{1}{\epsilon(q,\omega)}\right] = \frac{1}{\omega}\int_0^\infty \omega' \,\mathrm{Im}\left[-\frac{1}{\epsilon(0,\omega')}\right]\delta\left(\omega - \left(\omega' - \frac{q^2}{2}\right)\right)\mathrm{d}\omega' \quad (3.71) \qquad \text{(A.U.)}$$

[5] The exciton is a quasi-particle: a bound state of an electron and electron-hole.

where a quadratic extension into the momentum-transfer plane is assumed and ω' is the zero-momentum energy-loss. Let us review the physics contained in the model of Ashley. The model of Ashley relates to the single-pole approximation of the Penn model[60], which in turn is based on the Lindhard dielectric function. The Lindhard dielectric function $\epsilon_L(q,t)$ describes the response of a degenerate free-electron gas to an external perturbation in terms of energy-momentum transfer. It includes the following two basic modes of energy absorption for the electron,

1. Electron-hole pair excitations.

2. Collective/plasmon excitations.

A shortcoming of the Lindhard dielectric function is that it cannot represent the finite width of the plasmon resonance, nor the associated lifetime of the plasmon. Nevertheless, let us proceed with the extension of Ashley. The Dirac delta allows us to integrate out the momentum-transfer q which results in the following expression,

$$\tau(E,\omega) = \frac{1}{2\pi E}\int_0^\infty \omega' \,\mathrm{Im}\left[-\frac{1}{\epsilon(0,\omega')}\right] F\big(E,\omega,\omega'\big)\mathrm{d}\omega' \tag{3.72}$$ (A.U.)

where,

$$F\big(E,\omega,\omega'\big) = \begin{cases} \frac{1}{\omega(\omega-\omega')} & \text{for } \omega' + \frac{q_-^2}{2} < \omega < \omega' + \frac{q_+^2}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.73}$$ (A.U.)

What happened to the e^2-term in front of the integral? Let us remind the reader that we have switched to atomic units in which the elementary charge e is set equal to one. We now express the inelastic inverse MFP as an integral over zero-momentum energy transfer. By using Fubini's theorem, we change the order of the ω and ω' integrals such that

$$\frac{1}{\lambda(E)} = \frac{1}{2\pi E}\int_0^{\frac{E}{2}} \mathrm{Im}\left[-\frac{1}{\epsilon(0,\omega')}\right] L\big(E,\omega'\big)\mathrm{d}\omega' \tag{3.74}$$ (A.U.)

which has an integral over zero-momentum energy transfer and where the

function $L(E,\omega')$ is defined as follows,

$$L\big(E,\omega'\big) = \omega' \int_{\omega_-}^{\omega_+} F\big(E,\omega,\omega'\big)\mathrm{d}\omega \tag{3.75}$$ (A.U.)

In a similar way, $SP(E)$ is calculated by multiplying the integrand of Eq. 3.75 with ω. The lower integration limit follows from momentum conservation,

$$\omega = \omega' + \frac{q_-^2}{2} = \omega' + \left(\sqrt{E} - \sqrt{E-\omega}\right)^2 \tag{3.76}$$ (A.U.)

Solving the latter for $\omega = \omega_-$ yields the following expression for the lower limit,

$$\omega_- = \frac{1}{2}E\left(1 + \frac{\omega'}{E} - \sqrt{1 - \frac{2\omega'}{E}}\right) \tag{3.77}$$ (A.U.)

The upper integration limit evaluates to,

$$\omega_+ = \frac{1}{2}E\left(1 + \frac{\omega'}{E}\right) \tag{3.78}$$ (A.U.)

In the work of Ashley, an additional exchange corrected version of Eq. 3.75 is given. We refer to Ref. 43 for the analytic solutions of both the regular and exchange corrected version.

It was observed by Kieft and Bosch from FEI company[44] that several refinements can be made to the quadratic extension of Ashley. The refinements can be summarized as follows.

1. Introduce a distinction between the interaction with (1) outer-shell electrons and plasmons, and (2) inner-shell electrons. It is assumed that outer-shell electrons and plasmon interactions apply to $\omega' < 50\,\mathrm{eV}$.

2. Neglect the exchange correction of Ashley for $\omega' < 50\,\mathrm{eV}$. The rationale is that the dominant loss mechanism in this regime is through the creation of a plasmon excitation, which then decays and subsequently creates a single electron-hole pair. Since plasmon excitations are of bosonic nature, exchange corrections should not be applied.

3. Restrict the energy loss such that a primary electron cannot end up with an energy below the Fermi level. This is achieved by using a modified upper limit, i.e. $\omega_+ \to \frac{1}{2}E\Big(1 - \frac{E_F}{E} + \frac{\omega'}{E}\Big)$, where E_F is the Fermi level.

4. Introduce a rescale, $L_c(E,\omega') = \frac{3}{2}L(E,\omega')$, for $\omega' < 50\,\mathrm{eV}$. The rescale is introduced to compensate the underestimation of Eq. 3.75 with respect to experimental and theoretical sources. The authors of Ref. 44 suspect that the limited validity is attributable to the single plasmon-pole approximation of Ashley. We emphasize that the authors have considered a constant factor of 3/2 for all materials.

5. Ignore momentum conservation for interactions with inner-shell electrons. Instead of momentum conservation, the following phenomenological model is used,

$$L\big(E,\omega'\big) = \int_{2\omega'}^{E+\omega'} \omega' F\big(E,\omega,\omega'\big)\mathrm{d}\omega = -\ln\frac{\omega'}{E} \tag{3.79}$$

   where the integrand is identical to Eq. 3.75, but with a modified domain of integration.

The consequences for the above considerations lead to the following expression,

$$L\big(E,\omega'\big) = \begin{cases} \frac{3}{2}\ln\left[\frac{2E}{\omega'}\left(1+\sqrt{1-\frac{2\omega'}{E}}\right)-1\right] + \frac{3}{2}\ln\frac{E-E_F-\omega'}{E-E_F+\omega'} & \text{for } \omega' < \frac{1}{2}E \\ -\ln\frac{\omega'}{E} & \text{for } \omega' > 50\,\mathrm{eV} \\ 0 & \text{otherwise} \end{cases} \quad \text{(A.U.)} \tag{3.80}$$

Let us quantify the consequences of the refinements of Kieft and Bosch. The resulting inelastic MFP is shown Fig. 3.15 and the SP is shown in Fig. 3.16. We have also included the inelastic MFP and SP from Ashley with and without the exchange correction applied. A few of the refinements can be observed directly from the inelastic MFP in Fig. 3.15. Let us first conclude that the effect of exchange correction for the inelastic MFP vanishes in the limit of higher electron energies. In that same limit, we observe

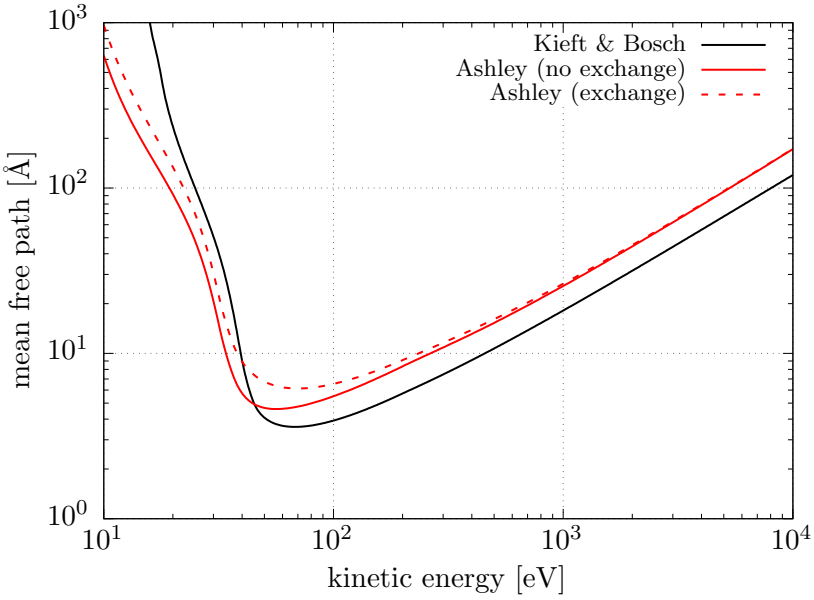


Figure 3.15: The inelastic MFP of silicon for different dielectric function models is shown.

that the inelastic MFP of Kieft and Bosch is about 1.5× lower than Ashley's version. This is a direct consequence of the rescale that was introduced (item number 4 in the refinement summary). Also notice that the inelastic MFP of Kieft and Bosch approaches 'infinity' at the left, i.e. where the inelastic MFP becomes very steep, at kinetic energy somewhat higher than Ashley's version without exchange correction. This observation relates to the decision to neglect the exchange correction (item number 2 in the refinement summary) and the restriction for energy losses below the Fermi level (item number 3 in the refinement summary). The latter implies that Kieft and Bosch start to run out of inelastic channels at an energy somewhat higher than Ashley. Similar effects are visible in the SP (Fig. 3.16). Notice that the loss per distance of Kieft and Bosch is about 1.5× larger than Ashley (without exchange correction) in the low energy regime. For completeness we have added data from the National Institute of Standards and Technology (NIST) to the SPs in Fig. 3.16. It should be noted that the SP from NIST is *calculated* by means of the Bethe model[61, 62] using a density-effect correction[63, 64]. That calculation, however, becomes 'ques-

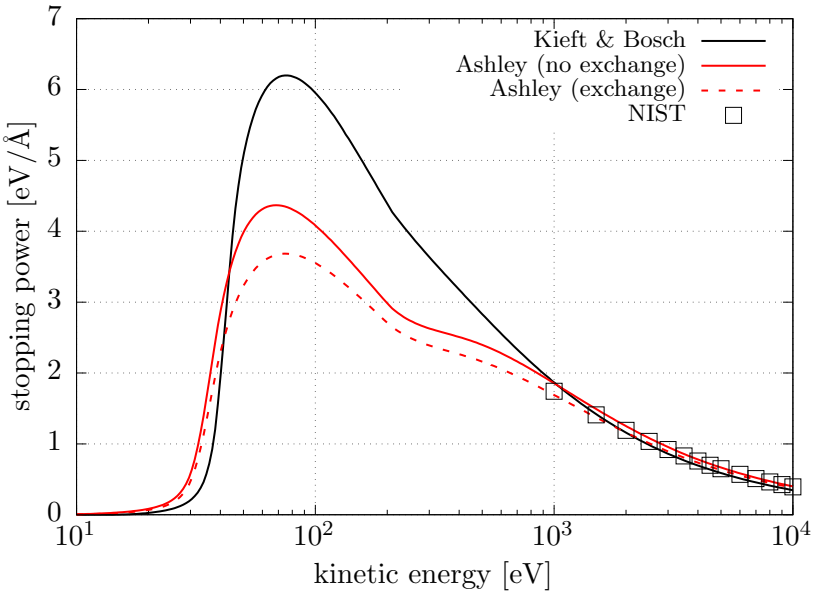


Figure 3.16: The SP of silicon is shown. This result is calculated using the dielectric function model of Ashley, including exchange correction.

tionable' at lower energies according to NIST, hence no data is given below 1 keV. The expected error at 1 keV is on the order of 10% and decreases for higher energies. We would like to refer to report no. 37 from the International Commission on Radiation Units and Measurements (ICRU) for more details.[65] Whether or not the SP of Kieft and Bosch for electron energies lower than 1 keV follows experiment more closely, is an open question we cannot address at the moment. The reason is that the measurement of SPs at (very) low electron energies is problematic. We will come back to validation of the models in another chapter.

For the moment, let us focus our attention back to the rigorous Monte-Carlo method. What we need to determine, in addition to the inverse MFP, is a way to find the energy loss for a particular inelastic event. The amount of energy loss in an inelastic event is probabilistic of nature and hence follows a particular distribution. What is this distribution and how do we sample a random energy loss? At first we remind that, by virtue of Eq. 3.11, the inelastic scattering cross-section is proportional to the inverse MFP. In other words, we may interpret the integrand of Eq. 3.74 as an inelastic

scattering probability density for the case of zero-momentum energy-loss (ω'). As a result, we identify from the integrand the following normalized probability density function,

$$p\big(E,\omega'\big) = \frac{\lambda(E)}{2\pi E}\mathrm{Im}\bigg[-\frac{1}{\epsilon(0,\omega')}\bigg]L\big(E,\omega'\big) \tag{3.81}$$ (A.U.)

A random zero-momentum energy-loss is obtained via the cumulative probability by solving the following integral equation for ω',

$$p\big(E,\Omega'\leq\omega'\big) = \frac{\lambda(E)}{2\pi E}\int_0^{\omega'}\mathrm{Im}\bigg[-\frac{1}{\epsilon(0,\omega')}\bigg]L\big(E,\omega'\big)\mathrm{d}\omega' = U \tag{3.82}$$ (A.U.)

where U is a random sample from the uniform distribution. What use do we have for a zero-momentum energy-loss (ω') anyway? Indeed, we need to sample yet *another* distribution from which the actual energy-loss (ω) can be determined, provided that ω' is given. The corresponding distribution is given by $L(E,\omega')$. To see this, let us go back to Eq. 3.74. We already discussed that the integrand can be seen as an inelastic scattering probability density. Notice that the function $L(E,\omega')$ is contained in the integrand, which in turn contains an integral over ω. In other words, the integrand of Eq. 3.74 actually is a probability density with respect to both ω' and ω,

$$p\big(E,\omega,\omega'\big) \sim \omega'\,\mathrm{Im}\bigg[-\frac{1}{\epsilon(0,\omega')}\bigg]F\big(E,\omega,\omega'\big) \tag{3.83}$$ (A.U.)

where the function F on the RHS is given by Eq. 3.73. The cumulative probability density function can be calculated analytically,

$$p\big(\Omega<\omega,\omega'\big) = \frac{\ln\left(1-\frac{\omega'}{\omega}\right)-\ln\left(1-\frac{\omega'}{\omega_-}\right)}{\ln\left(1-\frac{\omega'}{\omega_+}\right)-\ln\left(1-\frac{\omega'}{\omega_-}\right)} \tag{3.84}$$ (A.U.)

A random energy loss is obtained by equating the cumulative distribution function to a random sample from the uniform distribution and solve for ω.

The result of that gives us the following relation for a random energy loss,

$$\omega\big(U,\omega'\big) = \frac{\omega'}{1-\exp\left[(1-U)\ln\left(1-\frac{\omega'}{\omega_-}\right)+U\ln\left(1-\frac{\omega'}{\omega_+}\right)\right]} \qquad (3.85) \quad \text{(A.U.)}$$

provided that a random number U from the uniform distribution and a zero-momentum energy-loss ω' are given. Let us summarize the determination of a random energy loss in an inelastic scattering event:

1. Sample a random zero-momentum energy-loss by solving Eq. 3.82 using a random number from the uniform distribution. Preferably, the cumulative integral is precomputed such that a random zero-momentum energy-loss can be determined from a lookup table.

2. With ω' at hand, determine a random energy-loss (ω) through Eq. 3.85 by using another uniformly distributed random number.

We would like to consider what have we achieved so far by using a pragmatic example. Suppose that we have an electron with energy E which is about to scatter inelastically. We have explained that the distance to that event is determined randomly by using the total MFP, which includes the inelastic MFP. We can calculate the energy loss ω and hence determine that the electron will end up with an energy $E-\omega$ after the inelastic event. But where is the energy ω transferred to? Surely, in view of energy conservation, it must go somewhere! Indeed, a solid theory on inelastic scattering must include details of specific inelastic channels. We will restrict ourselves to the two basic inelastic channels which are captured by the model of Ashley: electron-hole pair excitations and collective/plasmon excitations. Let us first discuss the most elaborate one: electron-hole pair excitations. What happens physically is that an electron from a lower energy state is excited to a higher energy state. This is illustrated in the schematic band structure shown in Fig. 3.17. In reality, the band gap of a semiconductor can be classified into two categories: either it is a direct or an indirect band gap. The distinction is made as follows. In the band diagram of Fig. 3.17, each state is associated to a certain energy (vertical direction in the diagram) and crystal momentum (horizontal direction in the diagram). In an indi-

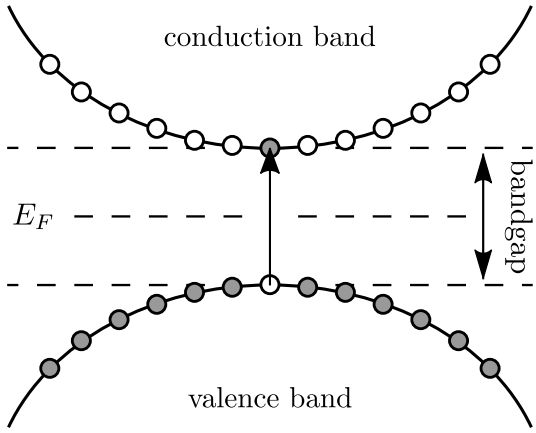


Figure 3.17: A schematic band structure for an intrinsic semiconductor is shown. At the top we have the conduction band with empty states (open circles). At the bottom we have a valence band with filled states (filled circles). The band gap is defined as the shortest distance between the highest state in the valence band and the lowest state in the conduction band. The dashed line halfway between the two bands represents the Fermi level (E_F). Notice that an electron is excited (arrow) across the band gap from the valence band into the conduction band, leaving behind an electron-hole in the valence band.

rect band gap material, the conduction band and valence band are shifted horizontally with respect to each other. This means that an excitation of an electron in an indirect band gap material from the valence band into the conduction band involves not only energy transfer (vertical change in the diagram), but also a change in crystal momentum (horizontal change).

Let us consider the band structure of a real material, such as silicon for example. The corresponding band structure is shown in Fig. 3.18. Indeed, silicon has an indirect band gap: the excitation of an electron from the highest energy state in the valence band (Γ_{25}) to the lowest state (X_1) across the band gap involves a change in crystal momentum.

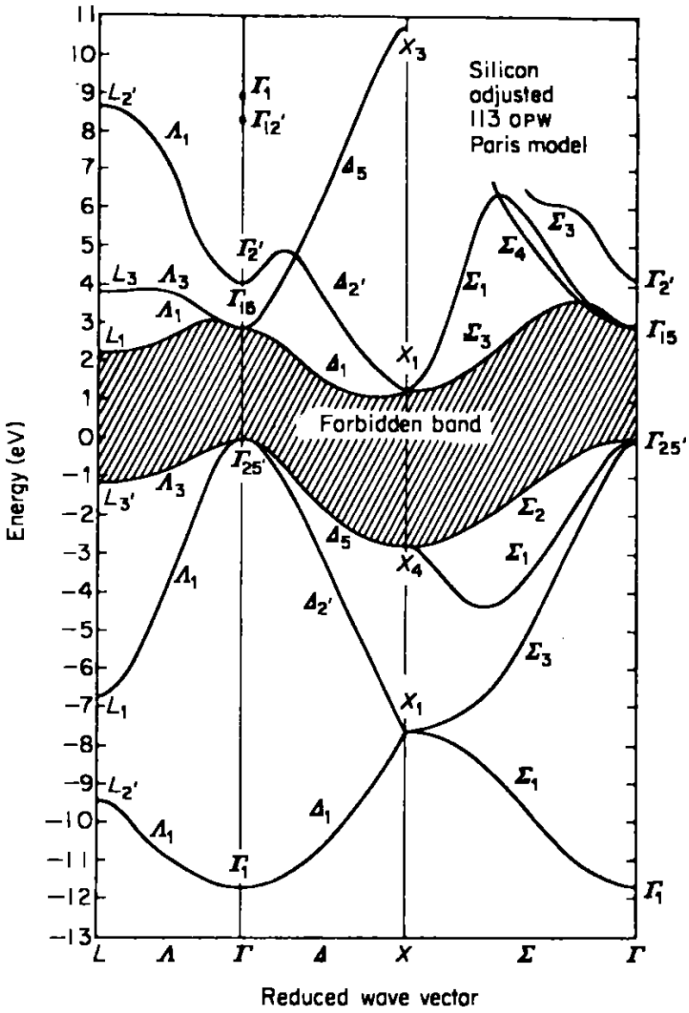


Figure 3.18: The band structure of silicon is shown. The symbols read $\Gamma \rightarrow k_x = k_y = k_z = 0$, $L \rightarrow k_x = 2\pi/a, k_y = k_z = 0$ and $X \rightarrow k_x = k_y = k_z = \pi/a$. The lattice constant a is 5.43 Å. Notice that silicon has an indirect band gap. The excitation of an electron from Γ_{25} in the valence band to X_1 across the band gap involves a change in crystal momentum. Image of the band structure was taken from Ref. 66.

Suppose that an energy transfer of ω from our primary electron with energy E is associated with the excitation of an electron-hole pair. What is the resulting energy of the secondary electron in the conduction band? First of all, we should be aware that, before the excitation, the secondary electron is in a bound state and hence has a certain binding energy. This binding energy must be overcome in order to excite the electron to a higher state in the conduction band. In other words, the expected energy of the secondary electron after the excitation equals,

$$E_{\text{se}} = E_F + \omega - E_B \qquad (3.86) \quad \text{(A.U.)}$$

where E_F is the Fermi level, ω is the amount of energy transfer and E_B is the binding energy of the secondary. Notice that, for reasons that will become clear later, we have defined the energy of the secondary electron with respect to the bottom of the band by including the Fermi level. Apparently we need to know the binding energy of the electron, but how do we determine it? This is where the sampling of the zero-momentum energy loss comes in handy. A random binding energy associated to an energy loss of ω' can be determined from the (relative) scattering cross-sections for electron-ionization, see Fig. 3.14. The only problem is that the scattering cross-sections from the Livermore database apply to free atoms. The use of electron-ionization cross-sections is therefore restricted to inner-shell excitations only. How do we deal with the outer-shell electrons? They need to be defined separately. In the simulator of Ref. 44, the following additional binding energies are used for silicon: 1.12 eV, 5 eV and 8.9 eV. Unfortunately, these binding energies are not provided with relative probabilities. Instead, the largest binding energy from that list, which is still smaller than ω', is considered. In any case, the band gap is defined as the lowest binding energy for an ω' above the band gap, but below the smallest available binding energy. The special case where ω' is lower than the band gap relates to optical phonons and will be dealt with in Section 3.4.

We briefly turn our attention to the electron-holes in the valence band. Are the electron-holes important for an electron-matter interaction simulator? At first we should mention the three key processes in which electron-holes play a role:

1. Radiative recombination: The electron-hole pair annihilates and a photon, with an energy equal to the difference between the electron and the electron-hole, is emitted consequently.

2. Auger recombination: A vacancy in an inner-shell is filled by an electron from a higher energy level of the same atom under the release of energy. This release of energy can be either a photon or another electron: the Auger-electron.

3. Trap-assisted recombination: This is similar to radiative recombination, but instead of direct, through intermediate energy states within the band gap. Such intermediate energy states are created by impurities for example.

Radiative recombination for materials with an indirect band gap involves a third (quasi) particle to account for the difference in crystal momentum. This particular type of recombination is, for a material such as silicon, very unlikely as conservation of crystal momentum requires the involvement of three (quasi) particles, i.e. the electron, the hole and a phonon for example. By looking at Fig. 3.13, we can see that inner-shell excitations have low probabilities. The amplitude of the L1-3 shell excitations is about two orders lower than the (bulk) plasmon excitation. The K-shell excitation is even more unlikely: it is about six orders lower. We therefore expect that the amount of inner-shell excitations, and hence the likelihood for Auger recombination, is of minor importance to our study. On the other hand, the trap-assisted recombination must be the most dominant recombination process for an indirect band gap material. The reason is that the trap inside the band gap allows the exchange of energy through lattice vibrations and hence can account for differences in crystal momentum between the electron-hole pair. Nevertheless, we expect that this process is of minor importance to us. The reason is that we are mainly interested in (hot)

electrons, i.e. the electrons with sufficient energy to (1) be ejected from the material and/or (2) initiate the chemical reactions in resists. We expect that the recombination rate of hot electrons through intermediate traps in the band gap is very low. Instead, they are most likely occupied by the electrons from the states close to the Fermi level. We therefore neglect the recombination of electron-hole pairs, for both indirect as well as direct band gap materials.

What remains to be discussed is the deflection angle of the primary electron and the resulting direction of the secondary electron. The corresponding scattering angles can be obtained by applying energy-momentum conservation. To that order, let us consider a collision of the primary electron (subscript index i) with relativistic energy $E + mc^2$ and a secondary electron (subscript index t). We choose to work in the frame of reference

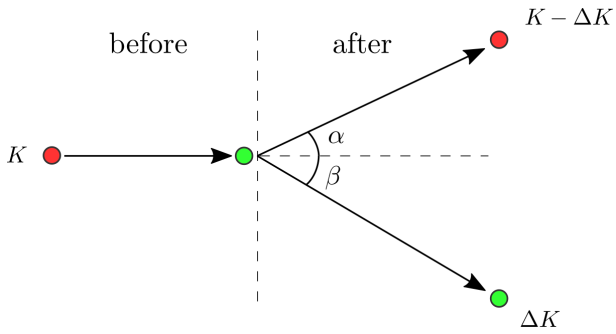


Figure 3.19: The inelastic collision of a primary electron (red) with a secondary electron (green) is shown. The frame of reference is with respect to the secondary electron (green). Before the collision, a primary electron (red) with energy E is traveling in the direction of the stationary secondary electron (green). After the collision, the primary electron (with energy $E - \Delta E$) and the secondary electron (with energy ΔE) are scattered in the directions of α and β respectively. The respective angles are determined from conservation of energy-momentum.

of the secondary electron, such that the secondary has zero kinetic energy. The kinetic energy of the primary electron before the collision is E. A total energy of ΔE is transferred to the secondary electron, such that the primary electron ends up with an energy of $E - \Delta E$, see Fig. 3.19. Conservation of momentum in the direction perpendicular to $\boldsymbol{p}_i$ gives us,

$$\sin\beta = \frac{p_i'}{p_t'}\sin\alpha \tag{3.87}$$

Similarly, conservation of momentum in the direction tangential to $\boldsymbol{p}_i$ gives us,

$$p_i = p_i' \cos\alpha + p_t' \cos\beta \tag{3.88}$$

$$= p_i' \cos\alpha + p_t' \sqrt{1 - \left(\frac{p_i'}{p_t'} \sin\alpha\right)^2} \tag{3.89}$$

where we have used Eq. 3.87 to eliminate the angular dependence (β) of the secondary electron. The scattering angle of the primary electron is obtained by considering,

$$\left(p_i - p_i' \cos\alpha\right)^2 = p_i^2 - 2p_i p_i' \cos\alpha + p_i'^2 \cos^2\alpha \tag{3.90}$$

$$= p_t'^2 - p_i' 2\sin^2\alpha \tag{3.91}$$

where the last equality follows from rewriting Eq. 3.89. By equating the RHS of both equations, we obtain the following expression,

$$2p_i p_i' \cos\alpha = p_i^2 + p_i'^2 \left(\cos^2\alpha + \sin^2\alpha\right) - p_t'^2 \tag{3.92}$$

The scattering angle of the primary electron as a function of the momenta becomes,

$$\cos\alpha = \frac{p_i^2 + p_i'^2 - p_t'^2}{2p_i p_i'} \tag{3.93}$$

The scattering angle of the secondary electron is obtained by eliminating the angular dependence (α) of the primary electron in Eq. 3.88. The net result of that gives us the following scattering angle,

$$\cos\beta = \frac{p_i^2 - p_i'^2 + p_t'^2}{2p_i p_t'} \tag{3.94}$$

We now determine the momenta in Eq. 3.93 and Eq. 3.94 in terms of the kinetic energy (E) and energy loss (ΔE) of the primary electron. Observe that the square of the energy-momentum relation reads,

$$\left(E + mc^2\right)^2 = (pc)^2 + \left(mc^2\right)^2 \tag{3.95}$$

This enables us to express the momenta in terms of kinetic energy and energy loss as follows,

$$(p_i c)^2 = E\big(E + 2mc^2\big) \tag{3.96}$$

$$\big(p_i' c\big)^2 = (E - \Delta E)\big(E - \Delta E + 2mc^2\big) \tag{3.97}$$

$$\big(p_t' c\big)^2 = \Delta E\big(\Delta E + 2mc^2\big) \tag{3.98}$$

We now substitute the momentum relations into Eq. 3.93 and Eq. 3.94. The scattering angle of the primary electron in terms of the kinetic energy and energy loss becomes,

$$\cos\alpha = \sqrt{\left(1 - \frac{\Delta E}{E}\right)\left(\frac{1 + \frac{E}{2mc^2}}{1 + \frac{E - \Delta E}{2mc^2}}\right)} \approx \sqrt{1 - \frac{\Delta E}{E}} \tag{3.99}$$

where the RHS is the non-relativistic approximation. Similarly, the resulting scattering angle of the secondary electron in terms of the same quantities (E and ΔE) equals,

$$\cos\beta = \sqrt{\frac{\Delta E}{E}\left(\frac{1 + \frac{E}{2mc^2}}{1 + \frac{\Delta E}{2mc^2}}\right)} \approx \sqrt{\frac{\Delta E}{E}} \tag{3.100}$$

Unfortunately, we cannot apply Eq. 3.99 and Eq. 3.100. The reason is that the secondary electron is in a bound state. To simplify matters, let us assume that we have an isolated system of an electron moving with charge $-e$ in a stable orbit around a positively charged nucleus with charge $+Q$. The classical Hamiltonian for this system reads,

$$H_{\text{se}} = E + P = \frac{1}{2}mv^2 - \frac{eQ}{4\pi\epsilon_0 r} \tag{3.101}$$

The kinetic energy of the bound electron can be determined in a classical way by equating the centrifugal force and electric force of the electron with respect to the positively charged nucleus,

$$F_c + F_e = \frac{mv^2}{r} - \frac{eQ}{4\pi\epsilon_0 r^2} = 0 \tag{3.102}$$

The 'stable' velocity for the electron for an orbit around the nucleus equals,

$$v = \sqrt{\frac{eQ}{4\pi\epsilon_0 mr}} \tag{3.103}$$

which can be substituted for the kinetic energy $E = \frac{1}{2}mv^2$ in the Hamiltonian:

$$E + P = \frac{1}{2}\frac{eQ}{4\pi\epsilon_0 r} - \frac{eQ}{4\pi\epsilon_0 r} = -\frac{1}{2}\frac{eQ}{4\pi\epsilon_0 r} = -B < 0 \tag{3.104}$$

Notice that the kinetic energy of the electron equals the binding energy and that the electric potential near the orbit is twice the binding energy. By transforming the kinetic energy $E \to E + 2B$, we obtain the Hamiltonian of a free electron, which is purely kinetic and equals the binding energy. The same transformation is applied to the primary electron, assuming that the electrostatic potential near the primary electron is also comparable to twice the binding energy. In addition, the Fermi level is subtracted from the primary energy to account for the fact that we are considering an isolated electron-atom system. The result of the transformation is that we can consider the primary and secondary electron as two 'free' scattering electrons. The energy of the primary electron is thus defined as $E \to E - E_F + 2B$ and the energy of the secondary is B. We cannot transform the 'initial' momentum, i.e. $p \sim \sqrt{2mB}$ of the secondary electron away because the direction is unknown. This means that we still cannot apply Eq. 3.99 and Eq. 3.100. Instead, we will pursue the procedure of Ref. 44, which in turn is based on the method of Ivanchenko in the Geant4 low energy ionization extension.[6] In that method the initial momentum of the secondary is neglected at first. Although no argument (neither by Kieft nor by Ivanchenko) is given for this reasoning, we suspect the following: If we assume that the initial momentum is completely random, then, for the *average* case, the net momentum must be zero. Despite the assumption for zero-momentum, the kinetic energy of the secondary is non-zero. As a consequence, they consider an inelastic process with an energy transfer of $(\omega - B) + 2B$ such that the resulting kinetic energy of the secondary

[6] See the source file '`G4LowEnergyIonization.cc`' of the Geant4 simulation toolkit.

electron includes the initial kinetic energy. The assumption of zero (net) momentum gives way to use Eq. 3.99 and Eq. 3.100. The relations between the kinetic energy, energy-transfer ΔE and energy-loss ω in the equations are,

$$E \to E - E_F + 2B \tag{3.105}$$ (A.U.)

$$\Delta E \to \omega + B \tag{3.106}$$ (A.U.)

The relative scatter direction of the secondary electron in the model of Ivanchenko, with respect to a primary electron moving in the $\hat{\boldsymbol{z}}$ direction, is modeled as follows,

$$\hat{\boldsymbol{p}}_t^{\star} = \sin\beta\cos(2\pi U)\hat{\boldsymbol{x}} + \sin\beta\sin(2\pi U)\hat{\boldsymbol{y}} + \cos\beta\hat{\boldsymbol{z}} \tag{3.107}$$

where U is a random sample from the uniform distribution. Note that the actual scatter direction $\hat{\boldsymbol{p}}_t'$ is obtained by alignment of the z-direction of Eq. 3.107 with the initial direction $\hat{\boldsymbol{p}}_i$ of the primary electron. In the end, a randomly oriented instantaneous momentum of the secondary electron is added,

$$\hat{\boldsymbol{p}}_t'' \sim \hat{\boldsymbol{p}}_t' + \sqrt{\frac{B}{\Delta E}}\Big(\sin\theta\cos(2\pi U_2)\,\hat{\boldsymbol{x}} + \sin\theta\sin(2\pi U_2)\,\hat{\boldsymbol{y}} + \cos\theta\,\hat{\boldsymbol{z}}\Big) \tag{3.108}$$

where $\cos\theta = 2U_1 - 1$ and both U_1 and U_2 are random samples from the uniform distribution. What about the direction of the primary electron? We cannot use Eq. 3.99 because we have added a randomly oriented instantaneous momentum to the direction of the secondary electron. We can still apply conservation of momentum nevertheless by equating the momentum before and after,

$$\boldsymbol{p}_i' = \boldsymbol{p}_i - \boldsymbol{p}_t'' \tag{3.109}$$

which can be recast to unit-directional vectors as follows,

$$\boldsymbol{p}_i' = \sqrt{E}\,\hat{\boldsymbol{p}}_i - \sqrt{\Delta E}\,\hat{\boldsymbol{p}}_t'' \tag{3.110}$$ (A.U.)

where we have used the non-relativistic energy-momentum relations on the

RHS. The resulting direction of the primary electron based on momentum conservation can be defined as follows,

$$\hat{\boldsymbol{p}}_i' \sim \frac{E}{\sqrt{E-\Delta E}}\,\hat{\boldsymbol{p}}_i - \frac{\Delta E}{\sqrt{E-\Delta E}}\,\hat{\boldsymbol{p}}_t'' = \frac{1}{\cos\alpha}\,\hat{\boldsymbol{p}}_i - \frac{\cos\beta}{\cos\alpha}\,\hat{\boldsymbol{p}}_t' \tag{3.111}$$

We would like to make a final remark about the inelastic scattering process. What we have assumed so far is that the secondary electron is released at the position of the primary electron. In reality, however, there can be a delocalization effect. The delocalization is described as follows. Energy of the primary electron is transferred by means of a (bulk) plasmon, which decays and subsequently releases a secondary electron at *another* position, which is some distance away from the primary. At present, delocalization is not included and is left for further discussion in the outlook.

We conclude this section with graphs of the SP (Fig. 3.20) and the inelastic MFP (Fig. 3.21) for various materials: aluminum (metal), silicon (semiconductor), gold (metal) and silica (insulator). We have included the refinements of Kieft and Bosch. Notice the peculiar behavior of the MFP

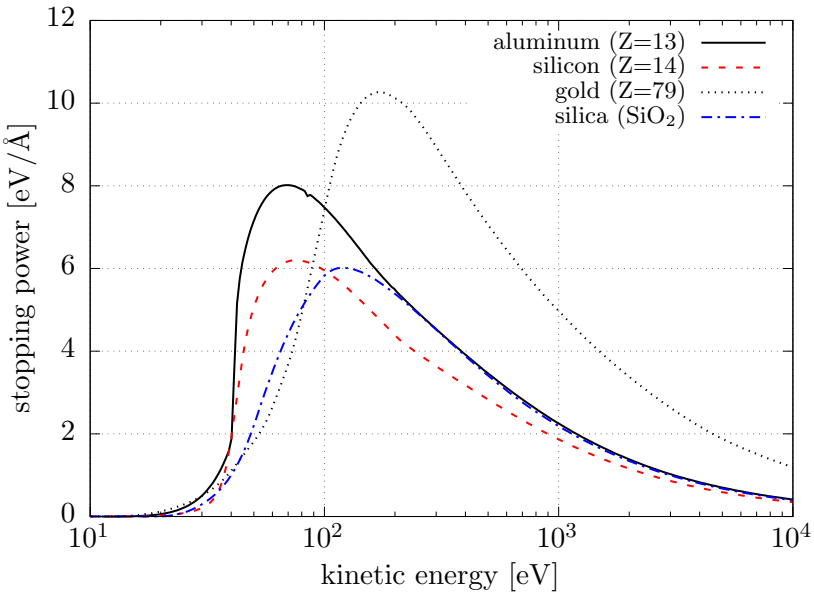


Figure 3.20: The SP for various materials is shown. The refinements of Kieft and Bosch are included.

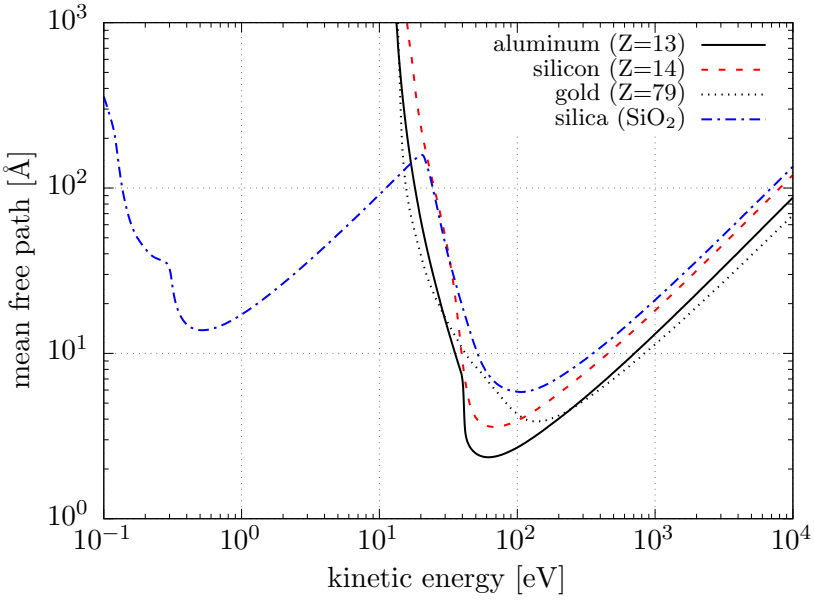


Figure 3.21: The inelastic MFP for various materials is shown. The refinements of Kieft and Bosch are included.

for silica in Fig. 3.21 for energies lower than 10 eV. Starting from the highest kinetic energy, the inelastic MFP of silica follows the same trend as the other materials. At approx. 100 eV the inelastic MFP reaches a minimum and then starts to rise. Then, at approx. 30 eV we see a deviation from the other curves: the inelastic MFP of silica starts to decrease again. What is happening there? What we actually see in this regime is the effect of longitudinal optical (LO) phonons. Let us have a closer look at the electron ELF of silica (Fig. 3.22) First of all, we can clearly see the large band gap of 9 eV. Observe that, similar to the electron ELF of silicon (Fig. 3.13), there are losses due to LO phonons below the band gap. If that is true, then why is the inelastic MFP of silica so different from silicon in the low energy range? The answer is found in the *amplitude* of the losses in the electron ELF. The losses due to LO phonons in the electron ELF of silica have, contrary to silicon, an amplitude comparable to the main (bulk) plasmon! We conclude that the coupling to LO phonons is much stronger for silica than for silicon. More evidence for this strong coupling to LO phonons is found, for example, in the work of Ref. 67. Our inelastic MFP for silica in the low en-

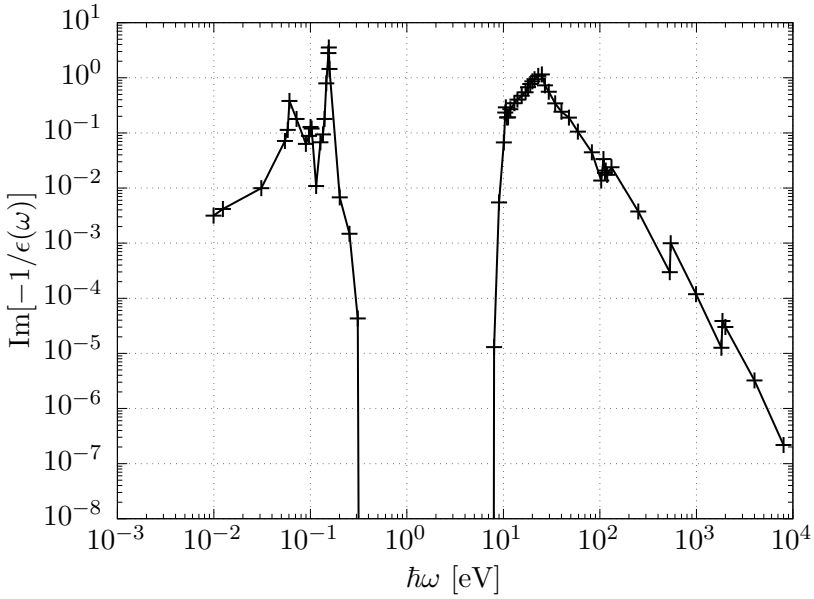


Figure 3.22: The electron ELF for silica (SiO_2) is shown. The markers are actual measurements, which are connected in the log-log figure by straight lines. The loss function is taken from Ref. 44.

ergy regime is consistent with the LO phonon model shown in Fig. 11 from Ref. 67. This suggests that phonon scattering potentially plays a crucial role. In the upcoming section we discuss the subject of phonon scattering in more detail.

3.4 Scattering at low electron energies

We have already discussed the elastic scattering of electrons by the interaction potential of atoms by using the Dirac equation using PWA. One problem with the latter is that it only considers the binary collision interaction which becomes questionable at energies lower than 100 eV.[68] We have argued that it is more likely that low energetic electrons behave like Bloch-electrons and interact with perturbations of the lattice (phonons) instead. There are two types of phonons to be considered: acoustic (AC) and longitudinal optical (LO) phonons. In the work of Ref. 67 (more precisely Fig. 11) it is shown that the dielectric function model applied to silica is

consistent with the Fröhlich theory of electron polar-optical phonon interactions. That very same conclusion is drawn in Ref. 50. We would like to point out that the consistency applies to the MFP. What about the angular scattering probabilities? Let us remember that the angular scattering distribution does not follow from the dielectric function model as presented in this study, and hence must be modeled separately. Fortunately, the angular scattering distribution associated to LO phonons is mainly forward scattered.[68] This means that the path of the electron effectively remains unperturbed after the interaction with an LO phonon. That very same assumption is employed in the program of Ref. 44, although it was never stated explicitly.[7] The AC phonons, however, must be modeled separately and there are at least two reasons why we need to do so.

Before we give our reasons, let us first obtain a rule of thumb for the typical amount of energy-loss associated to AC phonons. Throughout this section we will use S.I. units, unless stated otherwise. We consider the largest energy of the AC phonon in the first Brillouin zone, which is found at the boundary with $k = k_{BZ}$. Furthermore, we assume that the first Brillouin zone is spherically symmetric and that AC phonons have the following isotropic dispersion relation,

$$\omega_{AC}(k) = u_s k - \alpha k^2 \tag{3.112}$$

where u_s is the sound velocity and α relates to the bending of the dispersion relation towards the Brillouin zone boundary. The sound velocity is obtained from the dispersion relation by taking the zero limit of the group velocity,

$$\lim_{k \to 0} \frac{\mathrm{d}\omega_{AC}(k)}{\mathrm{d}k} = u_s \tag{3.113}$$

The expected energy of the acoustic phonons at the first Brillouin zone equals,

$$\hbar\omega_{AC}(k_{BZ}) = \hbar u_s k_{BZ} - \hbar\alpha k_{BZ}^2 \tag{3.114}$$

[7] The source code of that program reveals the path of the electron remains unperturbed.

In addition to AC phonon emission, there can also be absorption for T>0 which results in energy gain rather than energy loss. What we thus intend to determine is the *net average* of the energy loss and gain per event. We will follow the reasoning of Erik Kieft from FEI company: AC phonon absorption and emission should follow detailed balancing,[8]

$$\frac{\text{absorption}}{\text{emission}} \sim \frac{N_{BE}}{N_{BE}+1} = \exp\left(-\frac{\hbar\omega(k)}{k_B T}\right) \tag{3.115}$$

which states that the AC phonon absorption is proportional to $N_{BE}(k,T)$ and emission is proportional to $N_{BE}(k,T)+1$. He arrives at the following expression,

$$E_{AC} = \frac{4\pi \int_0^{k_{BZ}} [N_{BE}(k,T)+1-N_{BE}(k,T)]\hbar\omega_{AC}(k)k^2 \mathrm{d}k}{4\pi \int_0^{k_{BZ}} [2N_{BE}(k,T)+1]k^2 \mathrm{d}k} \tag{3.116}$$

where $N_{BE}(k,T)$ is the Bose-Einstein distribution, i.e.,

$$N_{BE}(k,T) = \frac{1}{\exp\frac{\hbar\omega(k)}{k_B T}-1} \tag{3.117}$$

Let us evaluate the integrals for silicon with a diamond cubic structure. We consider the quadratic phonon dispersion relations given in Ref. 69. There are three acoustic phonon modes in total: one longitudinal mode and two transverse modes, see Fig. 3.23. The coefficients for the longitudinal mode are $u_s = 9010\,\mathrm{m/s}$ and $\alpha = 2.00\times10^{-7}\,\mathrm{m^2/s}$. The coefficients for the transverse mode are $u_s = 5230\,\mathrm{m/s}$ and $\alpha = 2.26\times10^{-7}\,\mathrm{m^2/s}$. The boundary of the first Brillouin zone is effectively found at $k_{BZ} = 2\pi/a$, where the lattice constant a equals 5.43 Å. With a temperature T of 300 K, the average loss per event renders down to 25.2 meV for the longitudinal mode and 5.74 meV for the transverse mode. The *net* average (taking into account a factor of two for the transverse mode) evaluates to an energy loss of 12.3 meV per event. Consequently, the scattering of electrons by AC phonons is nearly (quasi) elastic. The low energy losses due to

[8] This reasoning was presented to us through email correspondence with Erik Kieft. It is not, however, included in the simulator from FEI company.

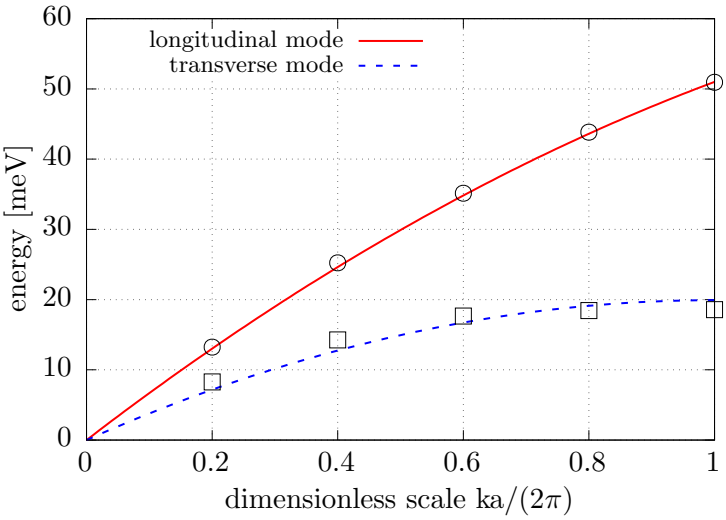


Figure 3.23: The phonon dispersion relations for silicon in the (100) direction are shown. The markers are taken from neutron scattering data.[70]

AC phonons are usually not included in the electron ELF, and when they do, they are marginally represented, see for example Fig. 3.13 and Fig. 3.22. Another reason to model the AC phonons separately is that the angular scattering distribution is not forward scattered.[68] We therefore would like to model the scattering of (hot) electrons with AC phonons in more detail.

We will pursue the model for AC phonons of Ref. 68, which is based on the expression of Sparks et al.[71] extended with a Coulomb screening parameter from Bradford and Woolf[72]. The idea is to use the model for (quasi) elastic AC phonons in the limit for low electron energies as a replacement for the elastic (Mott) scattering cross-sections.[50, 54, 67, 68, 44]. Let us work through the expressions for AC phonons of Schreiber and Fitting.[67, 68] The idea is to simplify the expressions such that they can be used more generically. For general applications, we assume a parabolic conduction band for electrons with an effective density of states (DOS) mass m_D. Furthermore, we assume that AC phonon scattering applies to electron energies where relativistic corrections can be neglected. The DOS as a function of energy

is given by the following relation,

$$D(E) = \frac{\sqrt{2m_D^3(E - E_{CB})}}{\pi^2\hbar^3} \tag{3.118}$$

where m_D is the DOS mass and E_{CB} is the energy at the bottom of the conduction band. The screened scattering rate P_{AC} in the low energy limit $E < \frac{1}{4}E_{BZ}$ is given by Fitting et al. and equals,

$$P_{AC}(E)|_{E<\frac{1}{4}E_{BZ}} = \frac{\pi\Xi^2 k_B T}{\hbar u_s^2 \rho_m} \frac{A}{A+E} D(E) \tag{3.119}$$

where Ξ is the AC deformation potential, ρ_m the mass density and A is the screening factor as introduced by Bradford and Woolf.[72] The electron energy at the Brillouin zone is estimated as follows,

$$E_{BZ} = \frac{(\hbar k_{BZ})^2}{2m_e} \tag{3.120}$$

The inverse MFP follows from dividing the scattering rate by the electron velocity,

$$\frac{1}{\lambda_{AC}(E)} = \frac{P_{AC}(E)}{u(E)} = \frac{m_e^\star}{\hbar k} \frac{P_{AC}(E)}{\rho_n} = \frac{m_e^\star P_{AC}(E)}{\sqrt{2m_e^\star(E - E_{CB})}} \tag{3.121}$$

where we have included the effective mass of electrons in the conduction band. The inverse MFP for AC phonon scattering in the low energy limit evaluates to,

$$\left.\frac{1}{\lambda_{AC}(E)}\right|_{E<\frac{1}{4}E_{BZ}} = \frac{\sqrt{m_e^\star m_D^3}\Xi^2 k_B T}{\pi\hbar^4 u_s^2 \rho_m} \frac{A}{A+E} \tag{3.122}$$

The screened scattering rate P_{AC} in the high energy limit $E > E_{BZ}$ is also given by Fitting et al. and equals,

$$P_{AC}(E)|_{E>E_{BZ}} = \frac{4\pi(2n_{BZ}+1)m_D\Xi^2}{\hbar\omega_{BZ}\hbar\rho_m} \frac{A^2}{E} \left(\ln\frac{A+E}{A} - \frac{E}{A+E}\right) D(E) \tag{3.123}$$

where n_{BZ} is the AC phonon population density. Note that, contrary to Ref. 67, a factor of $\hbar$ is missing in the denominator of Eq. 12b in Ref. 68. The inverse MFP in the high energy limit evaluates to,

$$\left.\frac{1}{\lambda_{AC}(E)}\right|_{E>E_{BZ}} = \frac{8\left(n_{BZ}+\frac{1}{2}\right)m_D\sqrt{m_e^\star m_D^3}\Xi^2}{\pi\hbar^4\rho_m\hbar\omega_{BZ}}\frac{A^2}{E}\left(\ln\frac{A+E}{A}-\frac{E}{A+E}\right) \tag{3.124}$$

The angular differential inverse MFP is obtained by normalizing the angular distributions from Ref. 68. It is convenient to define the following constant,

$$\lambda_0^{-1} = \frac{\sqrt{m_e^\star m_D^3}\Xi^2 k_B T}{\pi\hbar^4 u_s^2\rho_m} \tag{3.125}$$

which is the inverse MFP evaluated at $E = 0$. The angular differential inverse MFP for both regimes is then expressed as,

$$\frac{\mathrm{d}}{\mathrm{d}\Omega}\frac{1}{\lambda_{AC}(E,\theta)} = \begin{cases} \frac{\lambda_0^{-1}}{4\pi}\frac{1}{\left(1+\frac{1-\cos\theta}{2}\frac{E}{A}\right)^2} & \text{for } E < \frac{1}{4}E_{BZ} \\ \frac{\lambda_0^{-1}}{4\pi}\left(n_{BZ}+\frac{1}{2}\right)\frac{8m_D u_s^2 A}{\hbar\omega_{BZ}k_B T}\frac{\frac{1-\cos\theta}{2}\frac{E}{A}}{\left(1+\frac{1-\cos\theta}{2}\frac{E}{A}\right)^2} & \text{for } E > E_{BZ} \end{cases} \tag{3.126}$$

The inverse MFP is obtained from the differential inverse MFP by integration over the full solid angle,

$$\frac{1}{\lambda_{AC}(E)} = \begin{cases} \lambda_0^{-1}\frac{A}{A+E} & \text{for } E < \frac{1}{4}E_{BZ} \\ \lambda_0^{-1}\left(n_{BZ}+\frac{1}{2}\right)\frac{8m_D u_s^2 A}{\hbar\omega_{BZ}k_B T}\left(\frac{A}{E}\ln\frac{A+E}{A}-\frac{A}{A+E}\right) & \text{for } E > E_{BZ} \end{cases} \tag{3.127}$$

Now that we have generic expressions for AC phonons, let us try and reproduce the MFP of silica shown in Fig. 11 of Ref. 67. Note that in the work of Fitting et al., the effective mass and DOS mass for electrons in the conduction band is set equal to that of a free electron, i.e. $m_D = m_e$. Although it is not stated explicitly in that study, we assume that the parameters for the longitudinal and two transversal modes are identical. Note that in the work of Fitting et al., no details are given on the dispersion relation for AC phonons in silica. We therefore assume that the dispersion relation is linear, i.e. $\alpha = 0$. The MFP for silica is shown in Fig. 3.24, where

we have evaluated Eq. 3.127 using the parameters shown in Ref. 67. By visual inspection of the corresponding figures, we conclude that Eq. 3.127 reproduces the work of Fitting et al. What about Fig. 7 in Ref. 68? If we

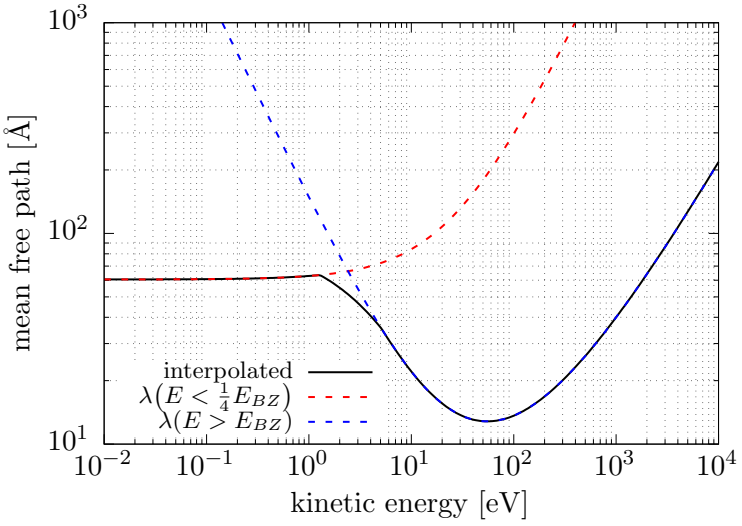


Figure 3.24: The (quasi) elastic MFP for AC phonons in silica is shown. The model of Fitting et al. involves a low and high energy limit with respect to the kinetic energy at the first Brillouin zone (dashed curves). The intermediate region between the two regimes is linearly interpolated (black curve). Details are given in the main text.

carefully consider the parameters used in that work (pay attention to ϵ_{AC}, the modified screening parameter and the rescaled scattering rate), then Fig. 7 is reproduced as well.

We would like to have a closer inspection on the characteristic shape of the MFP for AC phonons. At first we consider the energy at which the MFP is minimal. It is obtained by equating the first order derivative of the inverse of Eq. 3.127 to zero and solve for the energy. The smallest MFP is found at,

$$\lambda_{AC,\min} = \lambda_{AC}(E)|_{E=\alpha A} = \frac{\lambda_0}{n_{BZ} + \frac{1}{2}} \frac{\hbar \omega_{BZ} k_B T}{4 m_D u_s^2 A} \frac{\alpha(1+\alpha)}{(1+\alpha)\ln(1+\alpha) - \alpha} \tag{3.128}$$

where, $\alpha = 2.162581587\ldots$ is a transcendental number, which follows from solving the extremum of the first order derivative. Another property is

found by looking at the limit of the derivative for $E \to \infty$, which equals zero. This means that the MFP must have an inflection point for $E > \alpha A$. This point is found by equating the second order derivative and solve for the energy. The inflection point is found at,

$$\lambda_{AC,\mathrm{ip}} = \lambda_{AC}(E)|_{E=\beta A} \tag{3.129}$$

where, $\beta = 7.577356793\ldots$ is another transcendental number, which follows from solving the extremum of the second order derivative. After the inflection point the MFP is close to linear. The latter can be understood because $\lambda_{AC} \to \infty$ and $\mathrm{d}\lambda_{AC}/\mathrm{d}E \to 0$ as $E \to \infty$. The model of Fitting and Schreiber can thus also be explained as a phenomenological model with four parameters determined from the shape of the MFP curve, see Fig. 3.25. It is described as follows. At the lowest energy, the MFP starts flat out at λ_0, then reaches a cross-over point at $E_{\mathrm{co}} = E_{BZ}/4$ and decreases to a minimum at $E = \alpha A$. The MFP starts to increase again until $E = \beta A$ is reached, after which the MFP continues to grow linearly. The MFP can be expressed in terms of these parameters,

$$\lambda(E) = \begin{cases} \lambda_0\left(1 + \frac{\alpha E}{E_{\min}}\right) & \text{for } E < E_{\mathrm{co}} \\ \frac{\lambda_{\min}}{\alpha(1+\alpha)} \frac{\alpha E}{E_{\min}} \frac{(1+\alpha)\ln(1+\alpha)-\alpha}{\ln\left(1+\frac{\alpha E}{E_{\min}}\right) - \frac{\alpha E}{E_{\min}}\left(1+\frac{\alpha E}{E_{\min}}\right)^{-1}} & \text{for } E > 4E_{\mathrm{co}} \end{cases} \tag{3.130}$$

In any case, the most straightforward way is to use the AC deformation potential Ξ in Eq. 3.125, rather than the phenomenological model. Unfortunately, that parameter is usually not found in standard reference material property tables. There is, however, a work around for the special case of metals, which is also discussed in Ref. 44. The idea is that the mobility of electrons at the Fermi level relates to the scattering rate in the following way,

$$\mu = \frac{-e}{\rho_R n e} = \frac{-e}{m_e P_m} \tag{3.131}$$

where μ is the electron mobility, ρ_R the electric resistivity, n the density of conduction electrons and P_m is the scattering rate. The inverse MFP

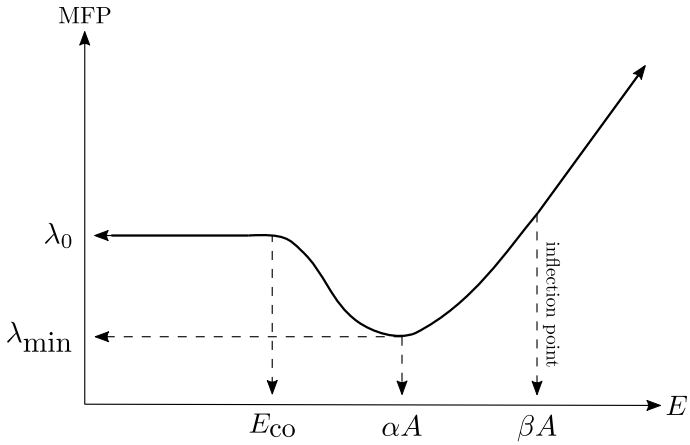


Figure 3.25: The schematic trend for the (quasi) elastic MFP for AC phonons on a log-log scale is shown. The MFP starts flat out at λ_0 at the far left, until it reaches the cross over at $E = E_{\rm co}$. The minimum MFP ($\lambda_{\rm min}$) is found at $E = \alpha A$. The inflection point is reached at $E = \alpha B$, after which it continues to grow linearly. The constants α and β are discussed in detail in the main text.

follows when the scattering rate is divided by the velocity of electrons at the Fermi level,

$$\left.\frac{1}{\lambda(E)}\right|_{E=E_F} = \frac{P_m}{u_F} = \frac{\rho_R n e^2}{\sqrt{2 m_e E_F}} = \frac{2}{3}\frac{m_e e^2}{\pi^2 \hbar^3}\rho_R E_F \tag{3.132}$$

In the latter we have used the non-relativistic relation for the electron velocity and,

$$n = \int_0^{E_F} D(E)\mathrm{d}E = \frac{2\sqrt{2}}{3\pi^2}\left(\frac{m_e E_F}{\hbar^2}\right)^{\frac{3}{2}} \tag{3.133}$$

We now assume that the inverse MFP for energies lower than E_F does not depend on energy, i.e.

$$\lambda_0^{-1} = \left.\frac{1}{\lambda(E)}\right|_{E=0} = \left.\frac{1}{\lambda(E)}\right|_{E=E_F} \tag{3.134}$$

The result of that gives the following relation for the effective AC deformation potential of metals,

$$\Xi^2 = \frac{2}{3\pi}\frac{e^2 u_s^2 \rho_m \rho_R \hbar}{m_e}\frac{E_F}{k_B T} \tag{3.135}$$

We conclude this section by giving a demonstration of the (quasi) elastic MFP (Fig. 3.26) associated with AC phonons for various materials: aluminum (metal), silicon (semiconductor), gold (metal) and silica (insulator). The curves are obtained by assuming a screening parameter equal to

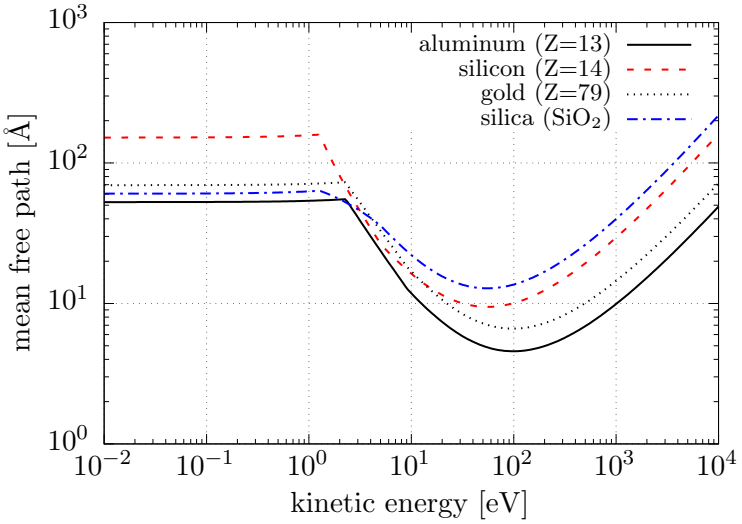


Figure 3.26: The (quasi) elastic MFP associated with AC phonons for various materials is shown. Specific details on the parameters are given in the main text.

five times the kinetic energy at the first Brillouin zone.[44] The corresponding parameters in the phenomenological model are given for completeness in Table 3.2. The deformation potentials for silicon and silica are obtained from Ref. 44 and Ref. 67 respectively. We have, on the other hand, no deformation potential at our disposal for aluminum and gold. Instead, we

material	λ_0	λ_{min}	$\mathbf{E_{ip}}$	$\mathbf{E_{co}}$
aluminum (Z=13)	52.7 Å	4.57 Å at 99.2 eV	347 eV	2.30 eV
silicon (Z=14)	152 Å	9.45 Å at 55.2 eV	193 eV	1.28 eV
gold (Z=79)	69.5 Å	6.62 Å at 97.7 eV	342 eV	2.26 eV
silica (SiO_2)	60.4 Å	12.8 Å at 55.1 eV	193 eV	1.28 eV

Table 3.2: The four parameters in the phenomenological AC phonon model for various materials are shown. The MFP in the limit of zero kinetic energy is given by λ_0, the minimum MFP is given by λ_{min} (including the energy at the minimum), the inflection point is given by E_{ip} and, finally, the cross-over point is given by E_{co}.

have used the Fermi level and electrical resistivity and evaluated Eq. 3.135 for an effective AC deformation potential. The DOS mass for aluminum, gold and silica[67, 68] is assumed to be m_e. For silicon we have used an effective mass and DOS mass of $m_e^\star = 0.26 m_e$ and $m_D = 1.09 m_e$ respectively for electrons in the conduction band.[73]. Moreover, we have included the experimental acoustic phonon dispersion relations for aluminum[74], silicon[69] and gold[75], all of which are obtained by neutron scattering experiments. The remaining parameters, such as Fermi level and mass density, can be found in standard reference material property tables.

The idea is now to use the AC phonon MFP as a replacement for the elastic MFP obtained with ELSEPA. We choose a regime where we only use the model for AC phonons and a regime where we exclusively use the Mott-scattering cross-sections from ELSEPA. That choice, however, is rather subjective and is determined at best for each material individually. At present, however, we employ a more generic approach, similar to Ref. 44, and define 100 eV as the limiting case for the Mott-scattering cross-sections. This means that below that 100 eV, the AC phonon MFP is used exclusively. The limiting case for AC phonons is set to 200 eV, after which the Mott-scattering cross-sections are used exclusively. In the intermediate range, we linearly interpolate the scattering cross-section of both models. The net result is an interpolated curve for the (quasi) elastic MFP (see Fig. 3.27), which includes the AC phonon model for the lowest energies and ELSEPA for the higher energies.

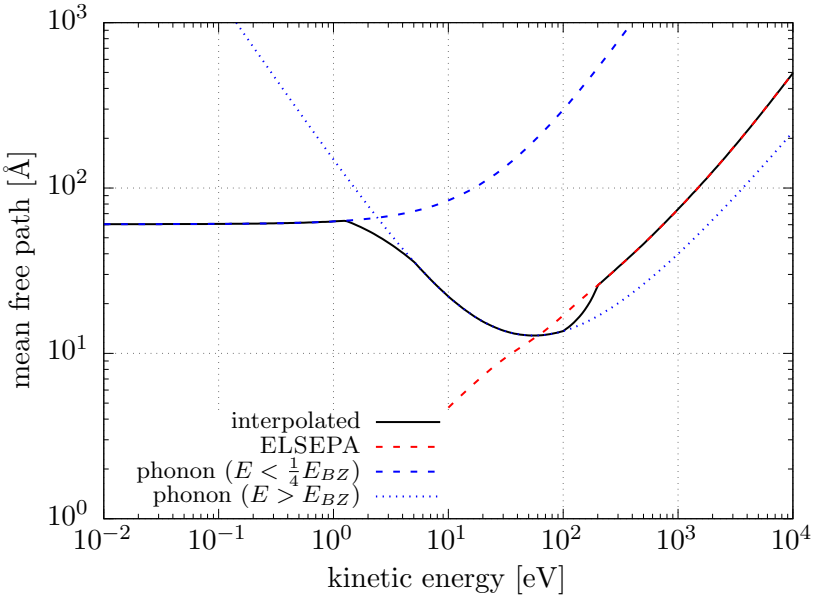


Figure 3.27: The interpolated (quasi) elastic MFP of silica from AC phonon and Mott-scattering cross-sections is shown. The MFP of the AC phonon model (green) is used exclusively for energies lower than 100 eV. The MFP for energies larger than 200 eV are determined exclusively by the Mott-scattering cross-sections of ELSEPA (red). The scattering cross-sections are linearly interpolated in the intermediate region. The net result is a single curve for the (quasi) elastic MFP (black).

3.5 Interface effects

In addition to the physics of elastic and inelastic scattering, special attention must be given to interface effects. An interface either marks the border between two different materials or it marks the vacuum-material boundary. There are, in either case, at least three different effects associated with interfaces: (1) refraction, (2) transmission/reflection and (3) interactions with surface plasmons. The transport of electrons across different materials requires us at first to properly define what we mean by the kinetic energy and momentum of the electron. To clarify matters, let us demonstrate the change in kinetic energy by virtue of a material-vacuum interface diagram, see Fig. 3.28. Suppose that an electron with kinetic energy E is injected from the vacuum side. Inside the material, the kinetic energy increases from E to $E + E_F + \Phi$, where E_F is the Fermi level and Φ is the work function.

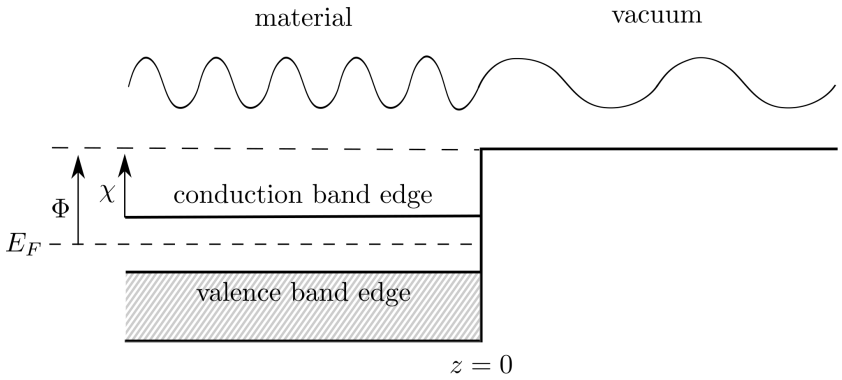


Figure 3.28: A Schematic diagram of a material-vacuum interface is shown. The dashed line between the conduction and valence band edge is the Fermi level E_F. The workfunction Φ is defined relative to the Fermi level. The electron affinity χ defines the energy difference between the vacuum level and the bottom of the conduction band.

When an electron crosses the boundary from one material to another, the resulting kinetic energy of the electron is calculated as follows,

$$E \to E + \left(E_F' + \Phi'\right) - \left(E_F + \Phi\right) = E + U' - U = E + \Delta U \tag{3.136}$$

where the primed terms correspond to the target material, U is defined as in the inner potential and ΔU denotes the net change in kinetic energy when moving from one material to another.

We now would like to discuss the effect of refraction. Refraction comprises an angular deflection due to a change in electron energy when crossing the interface. Suppose that we have an interface between two materials. Furthermore, suppose that an electron with kinetic energy E is propagating towards the interface at an angle θ with respect to the surface normal, see Fig. 3.29. The refraction angle θ' is found by applying momentum conservation parallel to the surface.

$$p \sin\theta = p' \sin\theta' \tag{3.137}$$

where p and θ relate to the momentum and angle of incidence. Similarly, p' and θ' relate to the momentum and angle of refraction. It is more convenient

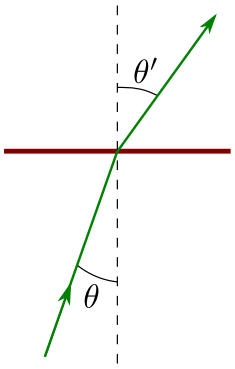


Figure 3.29: A Schematic diagram for refraction when crossing an interface is shown. The solid line at the center defines the interface and the dashed line is the surface normal. An electron is propagating at an angle θ towards the interface, which then deflects in the direction of θ' after the interface.

to express the momenta in terms of non-relativistic kinetic energy,

$$\sqrt{E}\sin\theta = \sqrt{E+\Delta U}\sin\theta' \tag{3.138}$$

Solving the latter for $\sin\theta'$ gives us the following relation for the angle of refraction,

$$\sin\theta' = \sqrt{\frac{1}{1+\frac{\Delta U}{E}}}\sin\theta \tag{3.139}$$

The angle of refraction is shown in Fig. 3.30 in the dimensionless variable $\Delta U/E$. Notice that for $\Delta U = 0$ we have that $\sin\theta' = \sin\theta$ and thus the path of the electron remains unperturbed. This happens when $E_F + \Phi = E'_F + \Phi'$. For $\Delta U > 0$, we can see that $\sin\theta' < \sin\theta$ which implies that we have deflection towards the surface normal. Analogously, we have deflection away from the surface normal for $\Delta U < 0$. Also notice that $\sin\theta' \to \infty$ as $\Delta U \to -E$. This corresponds to the critical angle, where the angle of refraction becomes parallel to the plane of the interface.

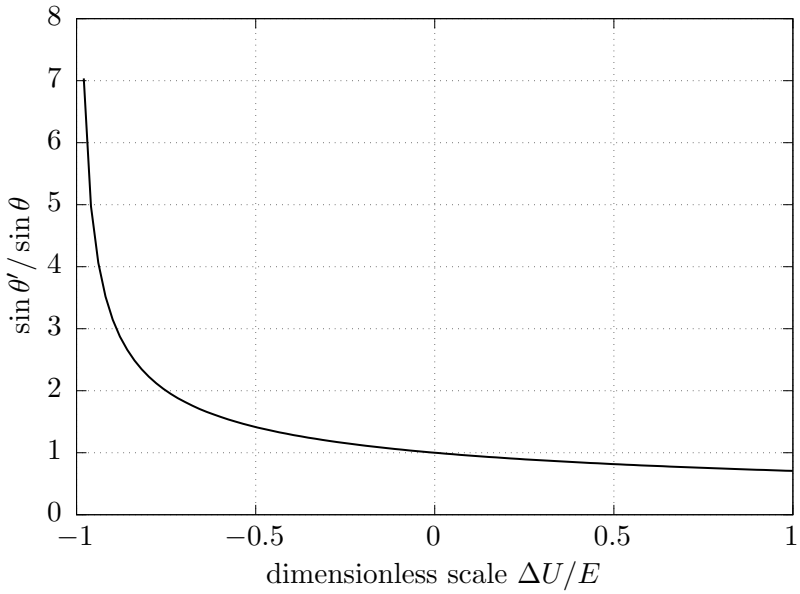


Figure 3.30: The angle of refraction for the special case of a free electron in the dimensionless variable $\Delta U/E$ is shown.

The probability for transmission and reflection of electrons at the interface is determined in a quantum mechanical way. In a one-dimensional potential, we define the interface between two materials and consider the solution to the Schrödinger equation on both sides, see Fig. 3.28. The electron wave function before the interface is defined as,

$$\phi(z) = A\exp(ikz) + B\exp(-ikz) \tag{3.140}$$

where A is the initial amplitude of the electron wave, B is the amplitude of the reflected electron wave and k is the momentum of the electron before the interface,

$$\hbar k = \sqrt{2m_e E}\cos\theta \tag{3.141}$$

The solution on the other side is given by,

$$\phi'(z) = A'\exp\left(ik'z\right) \tag{3.142}$$

where A' is the amplitude of the transmitted electron wave and k' is the

corresponding momentum,

$$\hbar k' = \sqrt{2m_e(E + \Delta U)}\cos\theta' \tag{3.143}$$

The transmission coefficient T of the electron wave is obtained in terms of the reflection coefficient as follows,

$$T = 1 - R = 1 - \left|\frac{B}{A}\right|^2 = 1 - \left(\frac{k' - k}{k' + k}\right)^2 = \frac{4k'k}{(k' + k)^2} \tag{3.144}$$

where we have applied continuity of the electron wave and its first order derivative at the interface $z = 0$. We now use the definitions of the non-relativistic electron momenta to evaluate the transmission coefficient in terms of the angle of incidence and angle of refraction,

$$T = \frac{4\tan\theta\tan\theta'}{(\tan\theta + \tan\theta')^2} = \frac{4\sqrt{1 + \frac{\Delta U}{E\cos^2\theta}}}{\left(1 + \sqrt{1 + \frac{\Delta U}{E\cos^2\theta}}\right)^2} \tag{3.145}$$

The transmission coefficient versus the ratio of the tangent of the angle of refraction to the tangent of the angle of incidence is shown in Fig. 3.31. For $\theta'/\theta < 1$ we have deflection towards the normal. Notice that the transmission coefficient is unity when $\theta' = \theta$.

At last we would like to discuss the coupling of electrons to surface plasmons. The physical picture is that the electron interacts with induced surface charges.[42] Consequently, there is, analogously to the plasmon peak for bulk losses, a plasmon peak associated to surface losses. Far away from the surface, the coupling to a surface plasmon is negligible and we can rely on bulk losses only. When the electron approaches the surface, the coupling to the surface plasmon increases and the coupling to bulk plasmons diminishes. We therefore employ, in the vicinity of the surface, a modified electron ELF associated to surface losses,[54, 76, 42]

$$\mathrm{Im}\left[-\frac{1}{\epsilon(\omega) + 1}\right] = \frac{\epsilon_2(\omega)}{(\epsilon_1(\omega) + 1)^2 + \epsilon_2(\omega)^2} \tag{3.146}$$ (A.U.)

where $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$ relate respectively to the real and imaginary part of

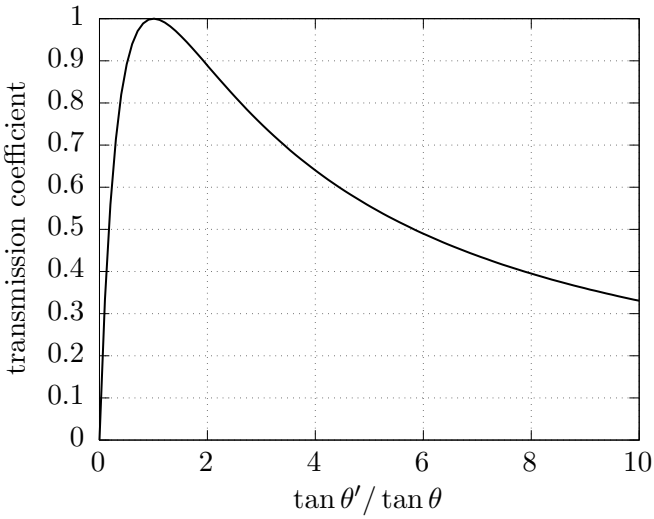


Figure 3.31: The transmission coefficient in terms of the angle of incidence and angle of refraction is shown.

the complex index of refraction for the bulk of a dielectric material. The real and imaginary part can be expressed in terms of the electron ELF as follows,

$$\epsilon_1(\omega) = \frac{-\mathrm{Re}\left[-\frac{1}{\epsilon(\omega)}\right]}{\left(\mathrm{Re}\left[-\frac{1}{\epsilon(\omega)}\right]\right)^2 + \left(\mathrm{Im}\left[-\frac{1}{\epsilon(\omega)}\right]\right)^2} \tag{3.147}$$ (A.U.)

$$\epsilon_2(\omega) = \frac{\mathrm{Im}\left[-\frac{1}{\epsilon(\omega)}\right]}{\left(\mathrm{Re}\left[-\frac{1}{\epsilon(\omega)}\right]\right)^2 + \left(\mathrm{Im}\left[-\frac{1}{\epsilon(\omega)}\right]\right)^2} \tag{3.148}$$ (A.U.)

where the real part is obtained from the imaginary part through the Kramers-Kronig relation, viz.,

$$\mathrm{Re}\left[-\frac{1}{\epsilon(\omega)}\right] = \frac{2}{\pi}\mathcal{P}\int_0^\infty \mathrm{Im}\left[-\frac{1}{\epsilon(\omega)}\right]\frac{\omega'}{\omega'^2 - \omega^2}\mathrm{d}\omega' - 1 \tag{3.149}$$ (A.U.)

We have calculated the electron ELF for surface losses in silicon. The real part of the complex index of refraction is obtained by numerical integration of the Kramers-Kronig relation using Maclaurin's formula.[77] The electron ELF for surface losses is then obtained through substitution in

Eq. 3.146. The result of that is shown in Fig. 3.32. Notice that, in comparison to the bulk plasmon peak, the amplitude of the surface plasmon is smaller and the characteristic energy loss is lower. The excitation of a surface plasmon depends on the amount of parallel momentum transfer and the distance to the surface. The resulting electric field of a surface plasmon – in the direction perpendicular to the plane – is proportional to,[54]

$$\mathcal{E}(z) \sim \exp\left(-|\boldsymbol{q} - \boldsymbol{q} \cdot \hat{\boldsymbol{n}}|z\right) \tag{3.150}$$

where z is the distance to the surface, $\hat{\boldsymbol{n}}$ the unit normal of the surface and the vectorial quantity $\boldsymbol{q} - \boldsymbol{q} \cdot \hat{\boldsymbol{n}}$ is the momentum transfer parallel to the plane of the surface. As a rule of thumb, the coupling depth to surface plasmons for an electron with kinetic energy of 1 keV is on the order of a few Ångstroms.

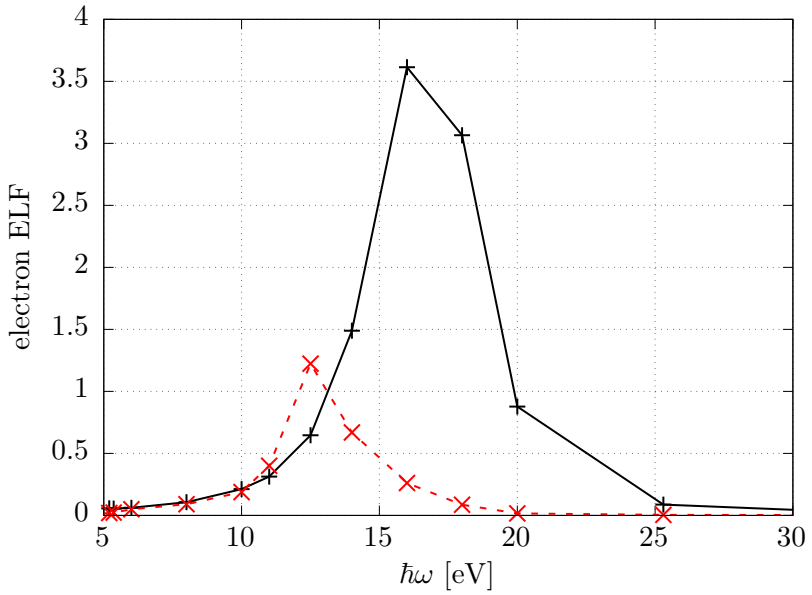


Figure 3.32: The bulk and surface plasmon peak in the electron ELF of silicon is shown. The markers are actual measurements, which are connected in the log-log figure by straight lines. The bulk loss function (black) is taken from Ref. 44. The surface loss function (red) is obtained through Kramers-Kronig analysis. Details are given in the main text.

3.6 Conclusion

We have discussed the determination of the scattering cross-sections for elastic and inelastic processes throughout this chapter in detail. In this concluding section, we would like to put the pieces of the puzzle together. Let us consider the ratio of the inelastic to elastic scattering cross-sections for various materials as a function of kinetic energy. This ratio will provide us with insight in the most probably scattering process at a given energy for a particular material. The ratio $\sigma_{\text{inelastic}}/\sigma_{\text{elastic}} = \lambda_{\text{elastic}}/\lambda_{\text{inelastic}}$ for aluminum (metal), silicon (semiconductor), gold (metal) and silica (insulator) is shown in Fig. 3.33. We can see some interesting effects in Fig. 3.33.

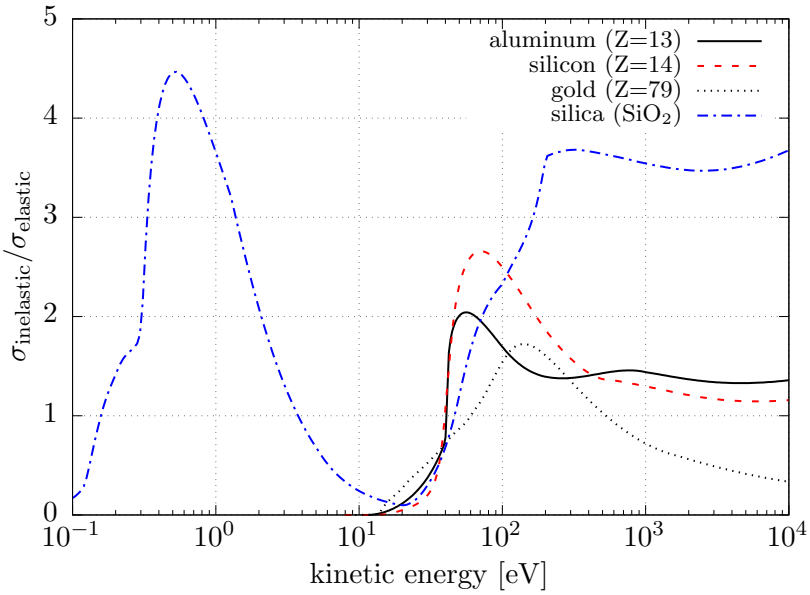


Figure 3.33: The ratio of the inelastic versus elastic scattering cross-sections is shown for various materials.

First of all, observe that $\sigma_{\text{elastic}} > \sigma_{\text{inelastic}}$ for kinetic energies lower than approx. 40 eV. What happens is that we run out of inelastic channels as we approach the Fermi level. Instead, the electrons primarily interact with (quasi) elastic AC phonons with a (nearly) isotropic angular scattering distribution. The result of that is electron diffusion, which has a profound effect on the spatial distribution of the electrons. Especially since, given

the fact that the loss per event is on the order of meV, the number of isotropic scattering events is abundant. The insulator silica, however, is an exception. At the lowest energies we can see that AC phonon scattering is overshadowed by the strong coupling of the electron to LO phonons. We would like to remind the reader that, although the effect of LO phonons is that $\sigma_{\text{inelastic}}$ becomes larger than σ_{elastic}, no secondary electrons result from that process. Moreover, in the case of electron emission from solids, only electrons with sufficient energy to escape the material are considered. Notice that for the metals aluminum and gold, most secondary electrons are created in the vicinity of the inelastic peak at approx. 100 eV. At increasing kinetic energies, the inelastic scattering process remains dominant for aluminum, but not for gold. A plausible explanation is that the high Z of gold is responsible for the dominance of elastic scattering. On the other hand, the insulator silica and the semiconductor silicon seem to maintain the preference for inelastic scattering in the higher kinetic energies. Notice that on average, most secondary electrons are created in silica. We can already anticipate that the number of emitted secondary electrons from the insulator is going to be significantly larger. Without performing any detailed calculation at present, the scattering cross-sections alone already provide deeper insight in the effects of electron scattering. As a final remark, we would like to mention that, despite the details and variety of the physics described in this chapter, there are still effects which have not been included:

- Detailed dispersion relations (instead of free electron model) for electrons in solids.
- Crystallographic anisotropy: the effective mass of the electron is dependent on the wavevector direction.
- Delocalization of secondary electron generation.
- Additional surface effects: roughness and super surface scattering.
- Electric fields: charging effects, trapping of electrons and charge dissipation.

The only additional effect that will be discussed in more detail is trapping (chapter 4) when we model the insulator PMMA. Perhaps the most critical effect neglected at present is the wavelike nature of the electron in the Monte-Carlo framework. What is assumed is that electrons behave like infinitesimally small scattering point particles. In reality, the extent of the electron wavefunction cannot be ignored and especially at lower electron energies. This brings us to the following question: have we included and modeled the physics sufficiently? What is and what is not important strongly depends on what is being measured/calculated. There are two approaches to justify the physics, which both are discussed in the upcoming chapters. The first, and most direct way, is to make a comparison to experimental observables (chapter 4). Another way is to consider a sensitivity analysis (chapter 6): exactly how sensitive is the outcome of the experiment with respect to the physics of the simulation?

Chapter 4

Electron-matter interaction simulator

Full Monte-Carlo simulation programs for SEM image acquisition are known to be notoriously slow. Our quest in reducing the computation time of SEM image simulations has led us to investigate the use of graphics processing units (GPUs). We have succeeded in creating a rigorous Monte-Carlo simulation program for SEM images, which runs entirely on a GPU.

In this chapter we present the details of the implementation of our GPU accelerated Monte-Carlo simulator for electron-matter interaction. As a case study for the performance, we consider the simulated exposure of a complex feature: an isolated silicon line with rough sidewalls located on a flat silicon substrate. We have also included a preliminary comparison to experimental electron yields and EELS spectra.

The content of this chapter is a modified version of an article we have published in the Proceedings of SPIE: *Verduin, T., Lokhorst, S.R., Kruit, P., and Hagen, C.W., "GPU accelerated Monte-Carlo simulation of SEM images for metrology," Proceedings of SPIE 9778, Metrology, Inspection, and Process Control for Microlithography, 97780D (2016).* This work was presented at the SPIE 2016 conference in San Jose.

4.1 Introduction

Scanning electron microscopy (SEM) image simulators can be of great benefit in the study of dimensional metrology. An example is the interpretation of the true size, shape and roughness characteristics of three dimensional resist features in top-down SEM images.[24, 25, 78] The computation time of a rigorous Monte-Carlo electron-matter interaction simulation is known to be notoriously slow. In Table 4.1, we present the simulation times for a yield experiment, which consists of an infinitely wide and thick slab of silicon. The simulation times are obtained by using the simulator from Ref. 78. We have calculated the average time it takes for a primary electron, including the cascading process, to come to rest or to be detected for various primary electron energies with an Intel Xeon X5650 processor using a single thread. In the case of a simulation with a rough feature, which is typically decom-

kinetic energy	time per primary	time per 10^6 primaries
5 keV	128.937 ms	35h49m
3 keV	78.048 ms	21h41m
1 keV	22.851 ms	6h21m
800 eV	17.882 ms	4h58m
500 eV	9.571 ms	2h40m
300 eV	4.193 ms	1h10m

Table 4.1: The simulation times for a simple yield experiment of an infinitely wide and thick slab of silicon are shown. We have calculated the average time it takes for a primary electron, including the cascading process, to come to rest or to be detected for various primary electron energies. Simulation times are given in terms of milliseconds (ms), minutes (m) and hours (h). Results are obtained with an Intel Xeon X5650 processor using a single thread.

posed into many geometrical elements, the computation times of Table 4.1 are increased dramatically. This problem is tackled, for example, with voxel based geometries[25], regions and height maps[79], and tetrahedra with shared faces[78]. Although these solutions seem to work well, computation time is still a problem, especially when statistics in the metrology play a role. A typical example is the determination of line edge roughness (LER) using the power spectral density (PSD).[27, 80] Another example is when one or more input parameters are varied over a range of values.[78]

We now ask ourselves the question: how can we reduce the simulation time per primary? One solution is to introduce range or depth cuts for electrons. The idea is to estimate the probability for an electron to escape the material, which is based on the kinetic energy of the electron and the shortest distance to the material boundary. If that probability is below some threshold, then we stop tracking that electron. This, however, typically affects the higher energetic primary electrons because they penetrate deeper into the material. We note that range or depth cuts have not been used in the calculation of Table 4.1. Another solution is to use more cores, which could be distributed over a cluster of workers. Our quest in reducing the computation time of SEM image simulation has led us to investigate the use of graphics processing units (GPUs).[81] The aim of this work is to investigate the advantage in computation time of SEM image simulation for metrology by using a GPU.

4.2 Design of the simulator

Let us begin by describing the GPU, which can be seen as a special type of parallel processor. At our disposal, we have the GTX480 from NVIDIA. The architecture of this GPU consists of 15 streaming multi-processors (SMs) with multiple processing cores running at 1.4GHz. Each SM has 32 processing cores and executes in a single instruction, multiple thread (SIMT) fashion. Instructions are issued per warp of 32 threads, where instruction divergence within a warp (due to if-then statements for example) is serialized. The maximum number of threads that can be *resident* on-chip is calculated as follows: 15 SMs times 48 warps per SM times 32 threads per warp equals 23 040 threads. We emphasize that this is *not* a fair number to compare against regular multicore CPUs. The *throughput* or peak performance of the GPU depends, for example, on register pressure, latencies due to arithmetic pipelines and memory transactions. More technical details on the GPU architecture and the graphics pipeline can be found in the documentation of CUDA from NVIDIA.[1]

[1] http://docs.nvidia.com/cuda

The implementation of the physics discussed in the previous chapter in a fully functional program running on the GPU is not straightforward. In order to match the parallel architecture of the GPU, the simulator needs to have a special design. The most basic assumptions of our simulation tool with respect to the tracking of electrons are the following,

- There are four possible discrete events: elastic scatter event, inelastic scatter event, detection event and boundary crossing.
- One electron can produce at most one secondary per inelastic event.
- Electrons move in a straight line from event to event.
- All electrons can be tracked independently.
- There are no time-dependent effects.

The elastic and inelastic scatter events require the random sampling of scatter angles and energy losses by means of cumulative distribution functions. The cumulative distribution functions are pre-computed and stored as two-dimensional tables in memory. We have employed the following strategy for the tables. One dimension accounts for the cumulative probability,

$$0 \leq P_i = \frac{i}{n_i - 1} \leq 1 \tag{4.1}$$

where the index i runs from zero to $n_i - 1$. The second dimension accounts for the kinetic energy of the electron,

$$E_1 \leq E_j = E_1 \exp\left(\frac{j}{n_j - 1} \ln \frac{E_2}{E_1}\right) \leq E_2 \tag{4.2}$$

where the index j runs from zero to $n_j - 1$. Notice that the kinetic energy is logarithmically spaced between the end points E_1 and E_2. Random

sampling of the tables is achieved through bilinear interpolation of the table,

$$t = (1-u)(1-v)T_{i,j} + u(1-v)T_{i+1,j} + (1-u)vT_{i,j+1} + uvT_{i+1,j+1} \quad (4.3)$$

where t is the interpolated value. The indexes i and j are related to P and E as follows,

$$i = \lfloor (n_i - 1)P \rfloor \quad (4.4)$$

$$j = \left\lfloor (n_j - 1)\frac{\ln E - \ln E_1}{\ln E_2 - \ln E_1} \right\rfloor \quad (4.5)$$

The corresponding interpolation constants u and v are determined by,

$$u = P - P_i \quad (4.6)$$

$$v = E - E_j \quad (4.7)$$

The interested reader might point out that dedicated hardware exists on the GPU for interpolating textures. Naively we would say it is beneficial to exploit this hardware. First of all, two-dimensional texture memory on the GPU is spatially cached by means of space filling curves. Second, the actual interpolation is performed on-chip and only a single instruction for a texture fetch is issued. We have tried to implement hardware-driven bilinear interpolation for the GTX480 and it turns out that it is actually slower than a global memory implementation. The reason is that the bandwidth of the texture cache of the GTX480 is significantly slower than the L1 cache of global memory. We therefore conclude that, performance wise, no advantage is found in exploiting the texture hardware of the GPU for bilinear interpolation.

The geometry, which is implemented in a special way, includes the vacuum, sample, electrons and shape of the detectors and is subdivided hierarchically into a (three dimensional) octree of cuboid cells. The shape of the detectors and the material boundary, which defines the interface between two different materials, is triangulated. Every leaf of the octree has a list of precisely those triangles which overlap with this particular cell. Let us

briefly discuss the construction of the octree. The idea is similar to Ref. 82, but then applied to triangles instead of particles. At first we define the root cell which has dimensions that determine the bounding box of the geometry. We then insert, one by one, the triangles that overlap with the root cell. The overlap of a triangle with a particular cell is calculated using the method of Ref. 83. There comes a moment when the number of triangles in the root cell surpasses a predefined threshold n_T. It is at this point that we split the root cell into eight octants by subdividing each dimension in half (the volume of each octant is 1/8th of the root cell). The triangles of the root cell are then re-inserted recursively into each overlapping octant. Note that each triangle in the root cell overlaps with at least one octant and possibly overlaps with multiple octants of the root cell. Another triangle from the ensemble is inserted by recursively traversing through the tree starting from the root cell until the overlapping leafs are found. The triangle at hand is then inserted in every overlapping leaf of the tree. When the number of triangles in a leaf surpasses the threshold, then that particular leaf is split into eight octants and the triangles of that leaf are re-inserted recursively. Note that every time we traverse one level down the tree, the dimensions of the cells are halved. This procedure is repeated until all triangles have been inserted. The net result is a nested tree in which no more than n_T triangles are found per leaf. For each leaf, only a small subset of the ensemble of triangles is found within that particular subspace. The determination of the intersection of a line segment, i.e. the straight path of an electron following Eq. 3.8, with the triangles only requires us to investigate the leafs of the tree that actually intersect with the line segment. For the latter we use the the fast, efficient and robust ray-box intersection algorithm of Williams et al.[84]. The idea of this intersection method is demonstrated for the two-dimensional case (quadtree) in Fig. 4.1. This limits the number of electron-triangle intersections per electron and hence reduces the computational stress. Please note that without any special implementation, we would have to determine the intersection of *every* electron with *all* triangles.[2]

[2]We assume a large number of electrons ($\sim 10^6$) and a large number of triangles ($\sim$

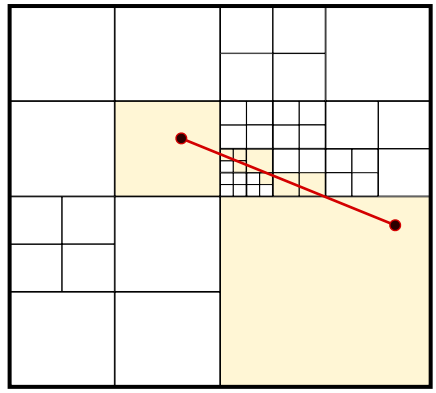


Figure 4.1: Schematic drawing of the intersection of a line segment with the triangles distributed in an hierarchical tree structure is shown. Every cell corresponds to a leaf of the quadtree and contains no more than n_T triangles. The determination of the intersection of a line segment (red) with the triangles only requires us to investigate the triangles in the (yellow) leafs that intersect with the line segment. The triangles in the remaining leafs (white) are not considered. The result is a significant reduction in electron-triangle intersections. The construction of a three-dimensional octree is explained in the main text.

The tracking of electrons within a cell can be summarized by the following steps. At first, for any electron, we determine the next event. To that order, we calculate (1) the distance to the material boundary (if any) (2) the elastic and inelastic attenuation length and (3) distance to the next detector in the direction of the electron (if any). The next event is determined by the *smallest* distance associated with a particular event. Finally, the electron is displaced and the event is executed. This is one iteration in our simulation tool, which is repeated until the electron is terminated. An electron is terminated in our simulation tool in the following four cases: (1) the electron is detected, (2) the electron is in vacuum and there is no material interface and no detector in the direction of the electron, (3) the electron is inside material and the kinetic energy is lower than the energy barrier and (4) the electron moves out of the border, i.e. the enclosed space of the full geometry.

10^6) which gives a problem size of order $\sim 10^{12}$.

To clarify matters, we now give an example of a few iterations for a single primary electron. In Fig. 7.3, we have a primary electron (1) moving in the direction of event 2, where the electron is transmitted and refracted from vacuum into silicon. The arrow at event 2 (also visible at event 8) indicates the normal of the surface at the intersection of the electron with the material boundary. The electron then moves to event 3 where the electron intersects with the cell boundary and traverses into the cell on the bottom left. We remark that no physical property of the electron is changed when the electron moves from one cell to another. The electron then scatters elastically at event 4 and moves in a new direction towards event 5 where the electron traverses into the cell on the bottom right. From event 5, the electron moves towards event 6, scatters inelastically and moves in the direction of event 7 where it moves in the adjacent cell (not shown). The secondary electron from inelastic event 6 moves towards the silicon-vacuum boundary and refracts at event 8 into the vacuum. This secondary electron then moves towards the boundary of the cell at event 9 and proceeds to event 10 where it traverses into the adjacent cell (not shown).

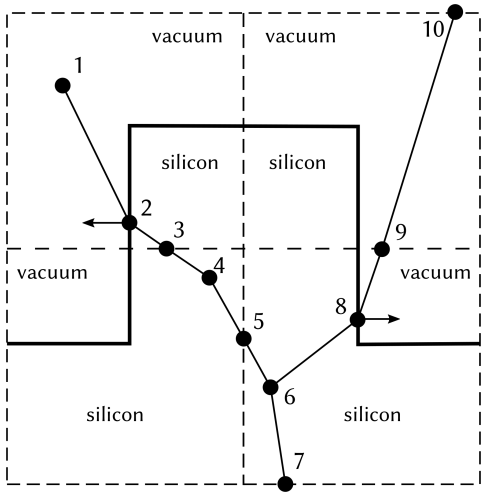


Figure 4.2: Schematic of the scattering trajectories in the cells. The cell boundary is given by the dashed lines. In reality, the simulation is three dimensional and the boundary consists of triangular shaped elements. The propagation of the electron via the numbered events is explained in detail in the main text.

In view of of parallel processing with the GPU, we face three main problems which we would like to discuss. The first problem is that the total number of electrons to be traced during a simulation increases because of inelastic events. In practice, memory on the GPU is contiguous and pre-allocated and so we must allocate contiguous arrays which are large enough to fit the cascading process. The size of the arrays, related to the properties of an electron in our simulation tool, is called the electron capacity. What is an appropriate value for the electron capacity? We will answer this question later when we discuss the saturation of the GPU hardware. The second problem can be stated as follows: if the properties of the electrons are indexed in contiguous arrays, then at which indices do we create secondary electrons? The third problem is related to the different types of events per electron. The next event for any electron is, due to the Monte-Carlo procedure, of random nature. This means that we expect to find different types of events per warp of 32 threads. This is a performance penalty because instructions on the GPU are, as mentioned before, issued per warp, hence different events within a warp are serialized.

Fortunately, the last two problems are both solved by grouping similar electrons, i.e. sharing the same type of event, by using a parallel GPU radix sorting method.[85] Performance wise, the use of sorting is justified by the fact that the method of Ref. 85 is capable of sorting a mind boggling $\sim 10^{10}$ 4-bit keys[3] per second on the GTX480. The idea of sorting the electrons is illustrated in Fig. 4.3. At the left side we have the contiguous array of n electrons, where each index corresponds to an electron in the simulation. An iteration begins with the determination of the next event for all electrons. In the illustration, we have designated the possible events with labels: inelastic events (I), elastic events (E), boundary events (B), detected electrons (D) and terminated electrons (X). Note that the events at this stage in sequential order are randomly distributed. Now comes the crucial step: the indices of the electrons are sorted by event to another array.[4] The result of that is a contiguous array where similar events are

[3]We only need four bits to discriminate between the different possible events in our simulation tool.

[4]We do not explicitly sort all properties of the electrons, instead we only sort the

grouped together and groups of different type of events are sequentially ordered. By using the sorted array on the right, we minimize the risk of having serialized execution within a warp.

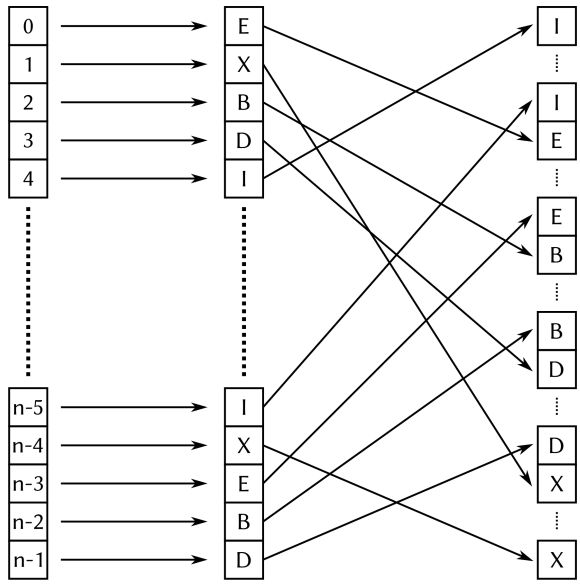


Figure 4.3: The grouping of similar electrons in an iteration is shown. There are inelastic events (I), elastic events (E), boundary events (B), detected electrons (D) and terminated electrons (X). At the left side we have the contiguous array of n electrons, where each index corresponds to a particular electron. In the simulation, the upcoming event (pointed by the arrow on the left) is determined. The indices of the electrons are then sorted by event to another array. The events in the sorted array (pointed by the arrows on the right) are grouped and the groups of different types are sequentially ordered.

The sorted array has an additional advantage with respect to the creation of secondary electrons. Suppose that we sort the events in such a way that the inelastic events are upfront (having the lowest indices) and the terminated electrons are at the back (having the highest indices). This actually corresponds to the ordering shown in Fig. 4.3 on the right. Note that terminated electrons are no longer used, and therefore can be replaced by new electrons. A secondary electron from an inelastic event with index i can now be created at index $n-1-i$, which due to the order in sorting, is a terminated electron.

indices pointing to electrons.

It is possible, however, that a terminated electron is not available at the given index and so a secondary electron cannot be created. The latter typically happens when the electron capacity is not sufficiently large. We have solved this issue by setting the inelastic event in that particular case on hold. In the next iteration, another attempt is made to create a secondary electron. This procedure is repeated until a free slot, i.e. terminated electron, becomes available. We emphasize that, in view of performance, this situation should be avoided by increasing the electron capacity. We also mention that an over-sized electron capacity, i.e. an array with too many free slots, is not beneficial either, because the sorting procedure also includes any excess of free slots in the array. In other words, an over-sized electron capacity would waste computational resources to redundant sorting.

This brings us to our first problem, i.e. what is an appropriate value for the electron capacity? To answer this question, let us suppose that we simulate a particular geometry, in which vacuum, material boundaries and detector shapes are defined. Also suppose that we have a particular field of view and that we have $n > 10^6$ primary electrons ready for the exposure. The electron capacity should be such, that enough room is available in the array to hold both the primary and future secondary electrons, which will be created during the exposure of the sample. But how many secondary electrons do we create? To answer this question, we initially oversize the electron capacity and pre-expose the geometry with a limited number, say $\sim 10^3$, of primary electrons at random within the field of view. During the pre-exposure in our GPU simulation tool, we count the total number of electrons (which includes both primary and secondary electrons) per iteration. An example of this counting for an exposure of an infinitely wide and thick slab of silicon is shown in Fig. 4.4. The result is a (set of) curve(s), which tells us how many electrons we are tracking on average during the simulation per incident primary.

Note that a curve from Fig. 4.4 starts with one electron, which is the incident primary, and ends with no electrons at all. The shape in-between is explained as follows. The number of electrons increases as soon as the

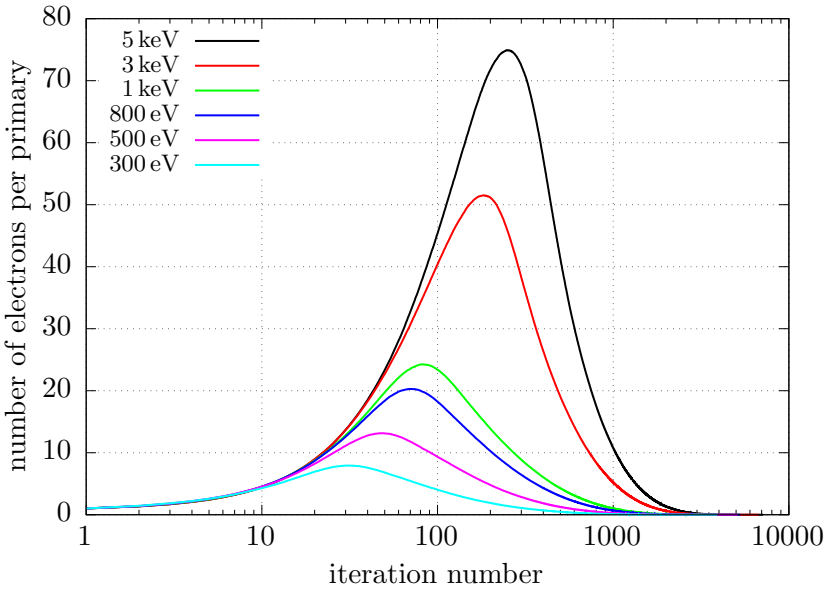


Figure 4.4: The total number of electrons being traced per iteration for various primary electron energies is shown. The curves are obtained by counting the total number of electrons during a pre-exposure of an infinitely wide and thick slab of silicon. Each curve is normalized by dividing the obtained counts with the number of incident primary electrons.

primary electrons enter material. The increase means that more secondary electrons are created in the inelastic events than terminated per iteration. This increase of number of electrons continues up to the top of the curve where the number of created secondary electrons equals the number of terminated electrons per iteration. This is the point where the electrons have lost most of their initial kinetic energy and the (quasi) elastic scattering starts to dominate over the inelastic scattering process. We now have, per iteration, more electrons terminating, which results in a decrease of the total number of electrons. This continues until all electrons are terminated.

There are, however, two more aspects to the number of electrons which we have not yet discussed. How many electrons do we need at least in parallel to saturate the GPU hardware? The answer can only be found by profiling the GPU using NVIDIA's visual profiler, which is included in the CUDA toolkit from NVIDIA. We have found that the GTX480 is saturated at a total of $n > 1.5 \cdot 10^5$ electrons per iteration, which means

that the *throughput*, in terms of electrons per second, is maximized. In view of the latter, we would like to make a remark with respect to the tails visible in Fig. 4.4. No matter how many electrons we begin with, in the end, all electrons will terminate and so the tails in Fig. 4.4 cannot be avoided. We must accept that in those cases the GPU is not saturated, hence the performance is severely reduced.

The last aspect relates to the limited amount of available memory on a GPU. In practice, we typically have the problem that the electron capacity can be sufficient to hold all primary electrons, but cannot be sufficient to include all future secondary electrons. The latter is solved, for example, by using a batch process with regular intervals. At the beginning of each interval, we push new primary electrons to the GPU. We do this in such a way that the total number of electrons is sufficient to saturate the GPU, but never exceeds the amount of available memory. This introduces two new questions: what size of the interval do we choose and how many primary electrons do we push to the GPU at the beginning of each new interval? Let us first discuss the choice for the size of the interval. The consequence of a small interval size is that the frequency of the batch process is increased. We should avoid this because memory copies between CPU and GPU are costly and reduce the overall performance. On the other hand, a large interval size should be avoided as well, which is best understood by looking at the tails in Fig. 4.4. With a large interval size, we run the risk that the number of electrons has dropped significantly at the moment we push new primary electrons to the GPU. The repeated result of that is a large fluctuation in the number of electrons throughout the simulation. In general, we choose the interval size equal to where the total number of electrons, which is determined from a pre-exposure with randomly distributed primary electrons, reaches a maximum. Let us now discuss the choice for the number of electrons per interval. The electron capacity can be decomposed as follows,

$$\text{capacity} \sim \text{primary count} + \text{secondary count} + \text{terminated count}$$

We assume that the electron capacity is a fixed number, which is larger than the saturation level of the GPU. At the very beginning of the simulation, we

only have primary electrons and no secondary electrons. Remember that, in our simulation tool, terminated electrons are replaced by new electrons. This means that the number of terminated electrons must be larger than the (expected) number of secondary electrons. In order to estimate the size of the batch, we need to know the expected number of primary electrons and secondary electrons throughout the simulation in advance. The expected number of primary and secondary electrons can be estimated by multiplying the curve from Fig. 4.4 with the size of the batch, which is then superimposed at regular intervals. This way we can optimize the size of the batch such that at each interval during the bulk of the simulation time: (1) the total number of electrons does not exceed the capacity, (2) the total number of electrons is higher than the saturation level and (3) the number of terminated electrons is larger than the number of secondary electrons. An example of an optimized batch process is shown in Fig. 4.5 for the exposure of an infinitely wide and thick slab of silicon with 5 keV electrons using an electron capacity of $2 \cdot 10^5$.

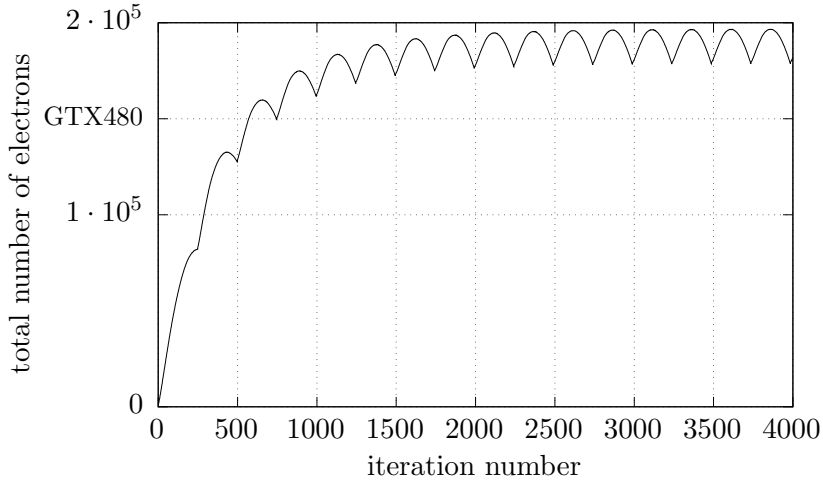


Figure 4.5: The expected total number of electrons per iteration for an optimized batch process is shown. The curve corresponds to an exposure of 5 keV electrons to an infinitely wide and thick slab of silicon. The number of electrons is determined by multiplying the curve from the pre-exposure with the optimized size of the batch, superimposed at regular intervals. The saturation level of the GTX480 is also given. Note that a limited number of iterations is shown because, ultimately, the total number of electrons vanishes to zero.

In this particular case we have on average, at each regular interval of 245 iterations during the bulk of the simulation time, 1024 primary electrons, 194836 secondary electrons accumulated from previous intervals and 2033 new secondary electrons. At the beginning of an interval we have approximately $2 \cdot 10^5 - 194836 = 5164$ free slots, i.e. terminated electrons. After pushing the primary electrons, we have a total of $5164 - 1024 = 4140$ terminated electrons available for replacement with new secondary electrons, which is approximately twice the expected number during that interval.

4.3 Performance results

For the demonstration of the performance, we now give an example of an exposure in our GPU simulation tool. At first we define, similar to Ref. 78, a geometry of a silicon line with rough sidewalls with dimensions 1.5 µm×32 nm (length, width) and 32 nm in depth.[5] This silicon line is centered on top of a flat silicon substrate with dimensions 1.5 µm×1.5 µm (length, width) and 2 µm in depth. The field of view for our exposure is 256 nm×64 nm (length and width) and so we only image a small part of the line. We have verified that the actual length of the line and the dimensions of the substrate are sufficiently large enough for electrons to scatter freely. We choose a spot size of 1.5 nm and a beam step size of approximately 0.5 nm, such that the resulting image has 512×128 pixels (length, width). For each exposure we use a dose of 6 mC/cm^2, which corresponds to a total of 6 553 600 primary electrons on average.

There are a few differences with respect to Ref. 78. In this work we use a resolution of 2 nm to define the roughness, which is increased because of memory limitations (1536 MB) of the GTX480. The model for generating the roughness is actually, in view of performance not important. This is why we have used a purely Gaussian roughness with a 3σ of 1.5 nm. For completeness, we mention that the total number of triangles for the rough isolated line equals 408 012.

[5] The interested reader is referred to Chapter 5 for a detailed description.

At first we run a pre-exposure of 10^3 randomly distributed electrons within the field of view. From this pre-exposure, we determine the size of the interval and the number of primary electrons per interval. The result of that, for various primary electron energies, is shown in Table 4.2. We

kinetic energy	size of the interval	primaries per interval
5 keV	249	1095
3 keV	184	1424
1 keV	83	2137
800 eV	70	2370
500 eV	49	3226
300 eV	31	4464

Table 4.2: The optimized batch process parameters for various primary electron energies are shown. The values are determined from a pre-exposure with 10^3 randomly distributed electrons within the field of view. The size of the interval is given as the number of iterations.

are now ready to fully simulate the SEM images by exposing the pattern with all of the primary electrons. The primary electrons are pushed to the GPU in batches at regular intervals of which the details are shown in Table 4.2. The resulting simulated SEM images are shown in Fig. 4.7. We have duplicated, for the sake of comparison, the simulation on the CPU by using the program of Ref. 78. We conclude that, apart from statistics in the simulation, no difference is found between the CPU and GPU simulated results. This, however, cannot be said about the total simulation time, which is shown in Table 4.3 for the GPU (GTX480) and the CPU (single threaded Intel Xeon X5650).

For completeness, we have also determined the maximum throughput of the GTX480, i.e. total number of electrons being tracked per second. The maximum throughput is given in Fig. 4.6 as a function of the kinetic energy of the primary electrons. Note that this throughput corresponds to the geometry of an isolated silicon line with rough sidewalls, located on a flat silicon substrate. We emphasize that the saturation level of the GPU relates to the number of electrons *per iteration*. What is shown in Fig. 4.6 is the number of electrons *per second*.

kinetic energy	GTX480	Intel X5650 (single thread)	speedup
5 keV	32m	2w5d12h	894×
3 keV	22m	1w4d20h	796×
1 keV	10m	3d11h	538×
800 eV	8m	2d17h	530×
500 eV	5m	1d10h	472×
300 eV	3m	15h16m	387×

Table 4.3: Comparison of simulation times for an exposure of an isolated silicon line with rough sidewalls, located on a flat silicon substrate. Each simulation involves 6 553 600 primary electrons on average. The geometry of the rough feature is decomposed into a total of 408 012 triangles. Simulation times are given in terms of weeks (w), days (d), hours (h) and minutes (m).

Although the speedup factors in Table 4.3 are tremendous, we would like to make two remarks about further improvements. First of all, we have not used a range or depth cut in the simulation. This means that electrons are being tracked, even if they have practically zero probability of escaping the material. Second, we would like to mention that newer GPUs are more powerful than the GTX480. The more recent GTX970 from NVIDIA, for example, has more multiprocessors, a faster clock-rate and faster memory in comparison to the GTX480.

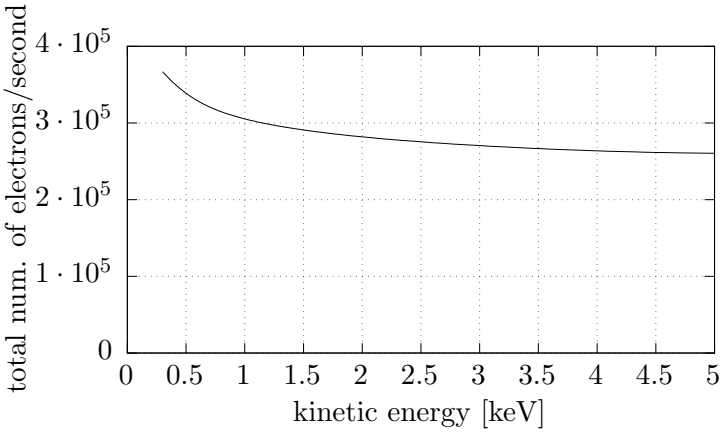
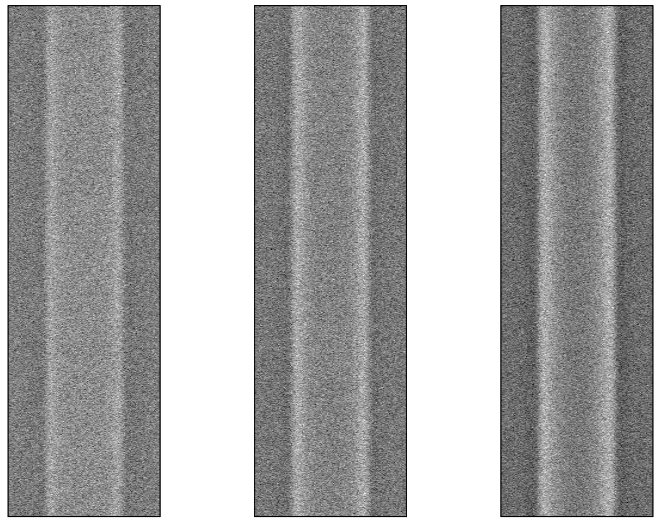
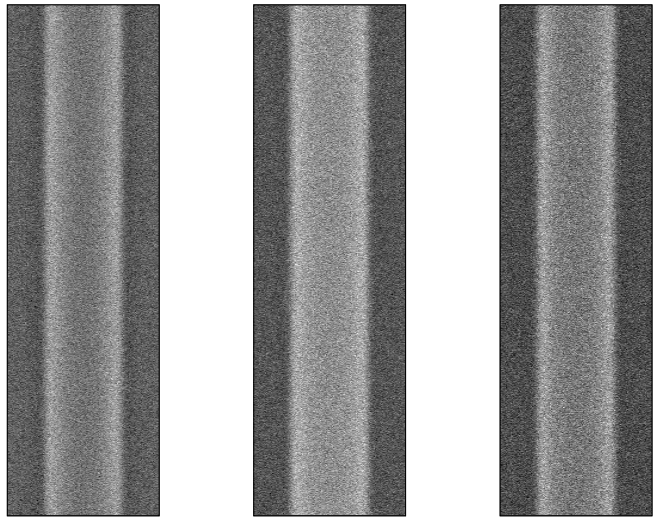


Figure 4.6: The throughput of the GPU measured in terms of total number of electrons per second. This graph corresponds to the geometry of an isolated silicon line with rough sidewalls, located on a flat silicon substrate, of which the details are given in the main text. For various primary electron energies, the total number of electrons, which includes primary and secondary electrons, is determined per second.



(a) 300 eV **(b)** 500 eV **(c)** 800 eV

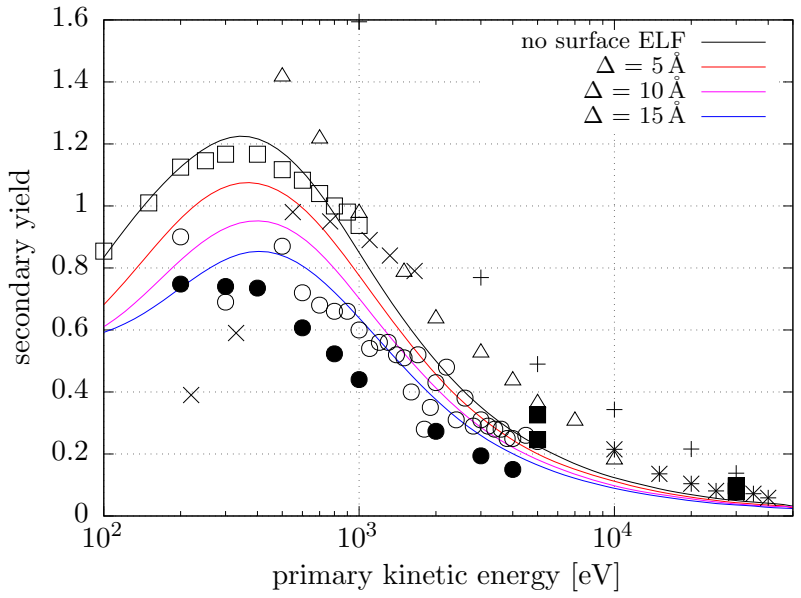


(d) 1 keV **(e)** 3 keV **(f)** 5 keV

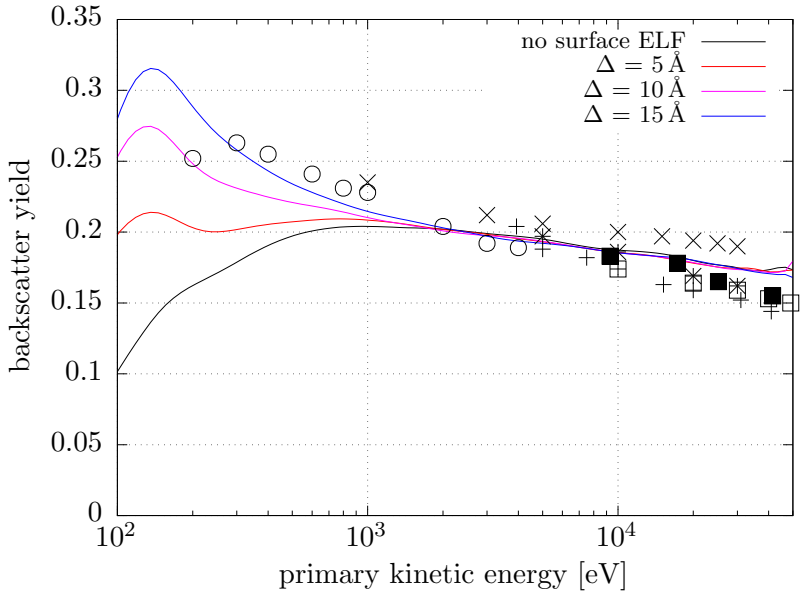
Figure 4.7: The GPU simulated SEM images of an isolated silicon line with rough sidewalls, located on a silicon substrate are shown. The field of view equals 256 nm×64 nm (length and width) and so only a small part of the rough line is exposed. The spot size equals 1.5 nm and the beam step size is approximately 0.5 nm. Each image is made out of 512×128 pixels (length and width). The exposure dose equals 6 mC/cm^2.

4.4 Comparison to experiment

In this section we would like to make a connection to experiment. At first we consider the emission of secondary electrons and backscattered electrons from a half-infinite sample. In the simulation, we expose the sample to a (infinitely small) beam of electrons. A detector is located just above the surface to count the number of electrons that have escaped the material relative to the number of primary electrons, which is known as the electron yield. The idea is to vary the kinetic energy and measure the secondary yield (SEY) and backscatter yield (BSY) as function of primary kinetic energy. We have verified that the dimensions of our bulk sample are sufficiently large enough for electrons to scatter freely. In addition, we have added a small layer with varying thickness Δ to account for surface losses. Within the layer, we exclusively use the surface ELF, rather than the bulk ELF. This, however, is a crude simplification as the coupling of the electron to surface plasmons is more complicated.[54] Nevertheless, the strength of the coupling can be controlled by varying the thickness Δ. The latter gives way to investigate the sensitivity of the electron yield with respect to surface losses. The result of this experiment for silicon is shown in Fig. 4.8. Notice that the inclusion of a small layer of a few Ångstrom already has a seizable effect on the simulated yield. In particular, notice that for increasing thickness, the SEY decreases and that the low energetic electrons in the BSY increase. There are two reasons for this effect. The inclusion of a surface layer affects the inelastic versus elastic scattering ratio. The calculated surface ELF is typically smaller in magnitude than the bulk ELF. This means that the scattering cross-sections for the surface ELF are smaller which results in a larger MFP for inelastic scattering. The latter implies that the balance in the inelastic versus elastic scattering ratio shifts in the direction of elastic scattering, which in turn results in lower probabilities for creating secondary electrons. Another effect, although of lesser importance, is that the surface plasmon peaks at a loss somewhat lower than the bulk plasmon. This means that the energy loss per inelastic event in the vicinity of the surface is most likely lower: electrons near the surface with an energy in the backscatter energy range dissipate slower to energies in the secondary energy range.



(a)



(b)

Figure 4.8: The secondary and backscatter yield of silicon is shown. All markers correspond to the experimental database of David Joy.[86]

In the next chapter we will consider the simulation of rough PMMA lines on a silicon substrate. We would therefore like to repeat the yield experiment for the case of PMMA. The simulation of an insulator, such as PMMA, however, is much more complicated since effects such as charging and trapping of electrons must be included.[87] We will strictly follow the approach of Dapor et al.[88] and include the trapping of low-energetic electrons.[6] The idea is that a low energetic electron propagating through an insulator induces a polarization field, which in turn induces a stabilizing effect. The resulting electron (or hole) with a polarization cloud is called a polaron. We have assumed, similarly to Ganachaud et al.[89], that the effective mass of polaron is large and therefore effectively remains localized at the trapping site. The model of Ganachaud et al. predicts the following energy-dependent probability for a polaron,

$$\lambda_{\text{trapping}} = C \exp\left(-\gamma E\right) \tag{4.8}$$

In the work of Dapor et al., the parameters $C = 1\,\text{nm}$ and $\gamma = 0.14/\text{eV}$ for PMMA are obtained from best fits of experimental data to a phenomenological model for the SEY. The interested reader is referred to Ref. 88 for more details on the determination of the parameters. We would like to make one final remark about the polaron. Note that the MFP, associated to the emergence of a polaron, is more than 1 µm for electrons with kinetic energy of 50 eV. This means that the polaron has no influence on the backscatter yield.

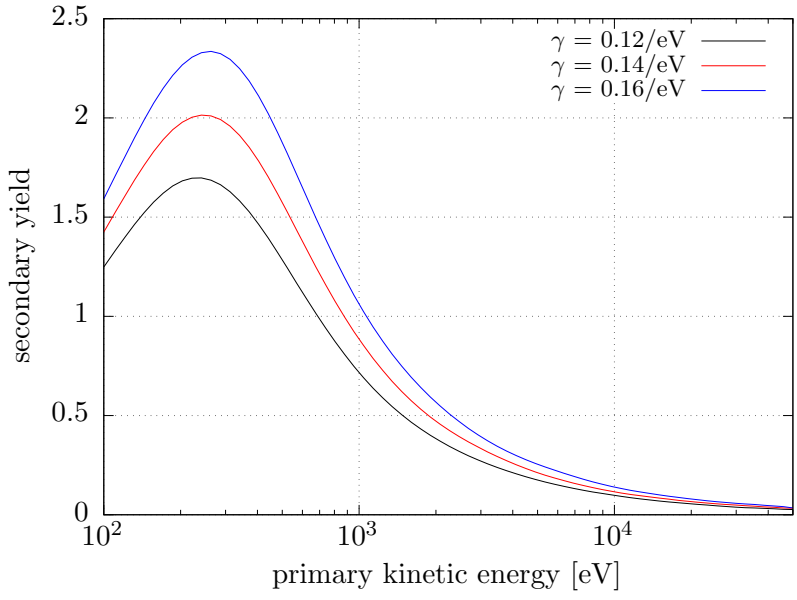
In addition, Dapor et al. have included optical phonons through Fröhlich theory, but neglected the acoustic phonons. We do not include Fröhlich theory explicitly, instead, we have used the ELF from Ref. 44 for PMMA, which includes the losses due to optical phonons below the bandgap.[7] For the (quasi) elastic scattering, we have exclusively used the ELSEPA cross-sections all the way down to a few electron volts. We assume that no surface effects have been included in the work of Dapor et al. The results

[6] The reason that we have not included this in chapter 3 is that the theory presented here is too specific for the case of PMMA.

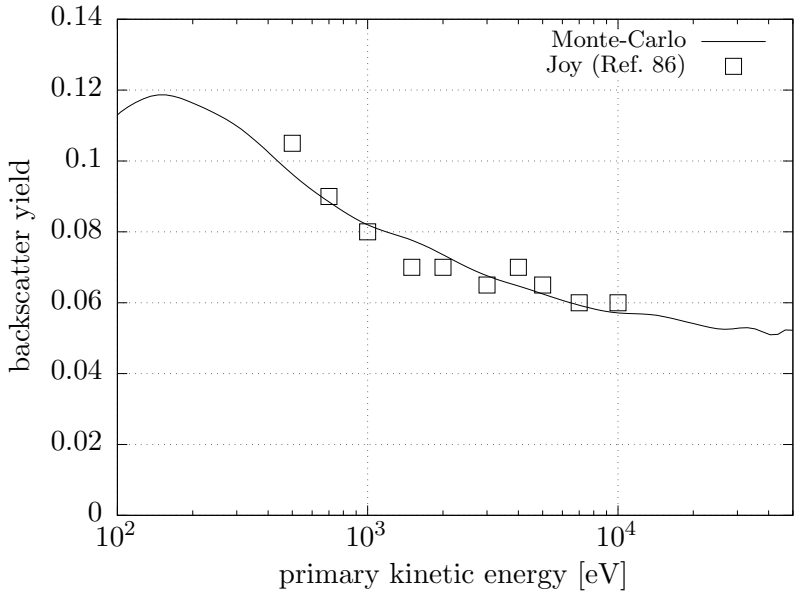
[7] Notice in the work of Dapor et al. that optical phonons are indeed missing in the ELF.

for PMMA are shown in Fig. 4.9, where we have used the same constant $C = 1\,\mathrm{nm}$ as Dapor et al. The interested reader is invited to compare the calculated yield to the experimental results shown in Ref. 88.

We now consider another experiment in which we expose a sample of finite thickness (50 nm) with a beam of electrons of fixed energy (E=50 keV). The idea is to measure the spectrum of energy losses as electrons escape the material. Note that electrons can emit from the top (reflection) as well as from the bottom (transmission) of the sample. That is why we have included another detector at the bottom of the sample. The acquisition of such an energy-loss spectrum is known as electron energy-loss spectroscopy (EELS). We emphasize that multiple inelastic events cannot be ignored and hence form an intrinsic difficulty in the acquisition of an electron energy-loss spectrum. In experiment, the electron energy-loss spectrum is retrieved nevertheless by means of one or more deconvolution procedures. In simulation, we simply neglect (1) the creation of secondary electrons and (2) neglect the electrons with more than one inelastic event. The result of this EELS experiment for silicon is shown in Fig. 4.10.

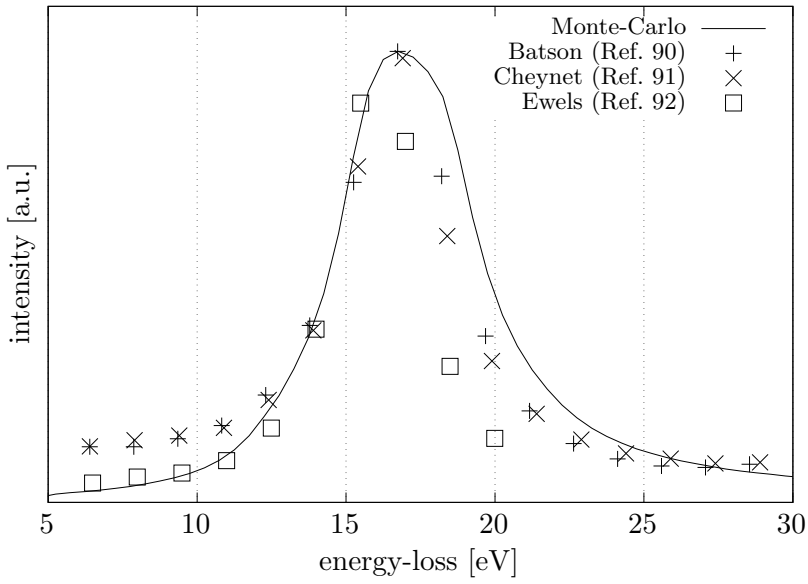


(a)

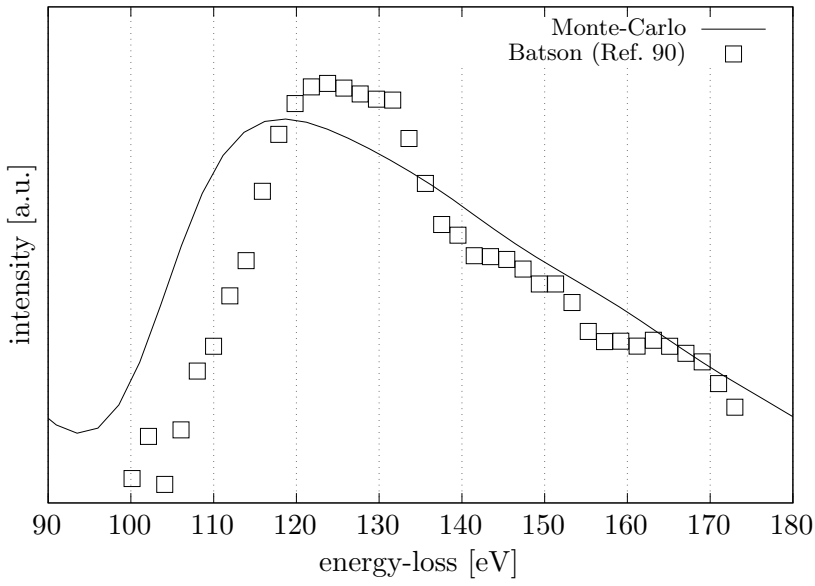


(b)

Figure 4.9: The secondary and backscatter yield of pmma is shown. The constant γ refers to a parameter in the model for electron trapping. The markers correspond to the experimental database of David Joy.[86]. More details are given in the main text.



(a)



(b)

Figure 4.10: The EELS spectrum for silicon is shown. The spectrum near the bulk plasmon is shown in (a). The spectrum near the L-shell excitation is shown in (b). All markers correspond to experimental measurements.

4.5 Conclusion

We have succeeded in creating a full Monte-Carlo simulation program for SEM images, which runs entirely on a GPU. We have used as a case study the simulation of a complex feature: an isolated silicon line with rough sidewalls located on a silicon substrate. At first we run the simulation on a GeForce GTX480 from NVIDIA. The very same simulation is duplicated on a CPU-based program, for which we have used an Intel Xeon X5650. We conclude that, apart from statistics in the simulation, no difference is found between the CPU and GPU simulated results. This, however, cannot be said about the total simulation time. We have determined that the GTX480 generates the images (depending on the primary electron energy) 387 to 894 times faster than a single threaded Intel X5650 CPU.

The performance increase is achieved as follows. First of all, we have used a special implementation for the geometry. The geometry includes the vacuum, sample, electrons and shape of the detectors and is subdivided into a special three dimensional octree of cuboid cells. The shape of the detectors and the material boundary, which defines the interface between two different materials, is triangulated. Every cell in the grid has a list of precisely those triangles which overlap with this particular grid cell. This gives a significant reduction on the number of electron-triangle intersections during the simulation of a complex geometry. Second, electrons with similar events are grouped by using a parallel radix sorting method, which also runs on the GPU. By sorting the electrons by event, we reduce the risk of instruction divergence within a warp of 32 threads. Finally, we saturate the GPU by using sufficient electrons per iteration to track in parallel. We have found, by using the visual profiler from NVIDIA, that the GTX480 is saturated at a total of $n > 1.5 \cdot 10^5$ electrons per iteration. In practice, the GTX480 does not have sufficient memory to process all electrons of the SEM image at once. Instead, a batch process is used to push primary electrons at regular intervals.

The speedup enables the fast acquisition of simulated SEM images for metrology. This means, for example, that many parameters of a rough feature (critical dimension, roughness, height, and so on) can be simulated in a reasonable amount of time. Moreover, because of the speedup, statistics can be generated as well by simulating a multitude of SEM images.

We conclude that with a GPU accelerated SEM image simulator, more results can be generated in less time. We now have the potential to investigate case studies in CD-SEM metrology, which otherwise would take unreasonable amounts of computation time.

Chapter 5

Simulation of side-wall roughness imaging

The true size, shape and roughness characteristics of resist features are not fully investigated in the analysis of two-dimensional SEM images. In reality, rough resist features are complex three-dimensional structures. The characterization of roughness of resist features naturally extends to the analysis of sidewall roughness (SWR) which can be measured, for instance, by using an atomic force microscope (AFM). However, in view of the large volume of wafers being produced in a typical production line, the AFM is not considered as a suitable metrology tool. Another problem with the AFM involves the complexity of the measurement, in which the exact shape of the tip plays a crucial role. Another possibility is to create virtual rough samples of patterns of lines and spaces and simulate the image acquisition using a SEM image simulator. Numerical studies on SWR are difficult to perform as it requires (1) the acquisition of many images for statistics, (2) variation of many parameters (critical dimension, roughness parameters, feature height and beam energy) and (3) spatially resolved details of micrometer-sized features with roughness defined at the nanometer scale. Studies involving the discrete modeling of the roughness on the sidewalls are therefore subject to time consuming simulations, and especially in the case of Monte-Carlo simulations. Because of the latter, detailed Monte-Carlo simulations are avoided and simulators with simplified physical models are used instead.

In this chapter we present a study on the effect of the SWR on the LER from two-dimensional SEM images. The core idea of this chapter is to randomly generate patterns of three-dimensional patterns of rough lines and spaces, where the SWR of the lines is modeled by means of a power spectral density (PSD) function. The three-dimensional patterns are then used in our GPU accelerated Monte-Carlo electron-matter simulator to simulate the acquisition of two-dimensional SEM images. The influence of the (three-dimensional) SWR on the resulting (two-dimensional) images is quantified using LER analysis. The analysis of this chapter provides new insight in the interpretation of the true size, shape and roughness characteristics of resist features from SEM images.

The content of this chapter is an updated version of an article we have published in the Proceedings of SPIE: *Verduin, T., Lokhorst, S.R., Kruit, P., and Hagen, C.W., "The effect of sidewall roughness on line edge roughness in top-down scanning electron microscopy images," Proceedings of SPIE 9424, Metrology, Inspection, and Process Control for Microlithography, 942405 (2015).* This work was presented at the SPIE 2015 conference in San Jose.

5.1 Introduction

In a previous chapter, we proposed a method for the determination of line edge roughness (LER) in low dose top-down SEM images.[28] However, the true size, shape and roughness characteristics of resist features are not fully investigated in the analysis of top-down SEM images. The studies that we found in literature indicate that the true SWR is larger than the measured LER in a top-down SEM image.[24, 25] We have a few remarks with respect to these studies. In the study of Li et al., the focus is on pure polycrystalline silicon lines with a Gaussian roughness model for the sidewalls.[24] The roughness of a resist feature, however, is typically characterized by more than just the standard deviation of a Gaussian-like distribution and involves additional parameters, such as correlation length and a roughness exponent.[26, 16, 27, 28] The study of Lawson et al. is different for two

reasons.[25] In the first place, Lawson et al. have used a more sophisticated model for the roughness of the sidewalls.[29] Unfortunately, the relation of the roughness parameters to correlation length and roughness exponent are not explicitly mentioned nor are the roughness parameters varied to study its influence on the LER. Second, the lines used in the study of Lawson et al. are made of pure poly-methyl methacrylate (PMMA) coated on a pure silicon substrate. Not only is PMMA a different material, it also comes, contrary to pure silicon lines, with a risk of resist shrinkage caused by the electron beam.[16] Although it is not addressed in the work of Lawson et al., this risk can be reduced, for example, by lowering the beam voltage and total electron dose, i.e. reduce the number of integration frames.[16, 28]

We would like to extend the study of Refs. 24, 25 and examine the relation between the SWR and the LER measured from simulated top-down SEM images by varying the parameters of a non-Gaussian roughness model. What happens, for example, to the measured LER when the correlation length of the SWR changes? In particular, we consider a case similar to that of Lawson et al. with rough PMMA lines on a silicon substrate. However, in order to reduce the risk of shrinkage, we will consider a beam energy of 300 eV and an approximate dose of 20 incident electrons per pixel. We have already shown in a previous study, albeit theoretically, that the LER can be determined under such noisy circumstances.[28] For this study, we use the power spectral density (PSD) function to model the roughness as described by Palasantzas.[26]

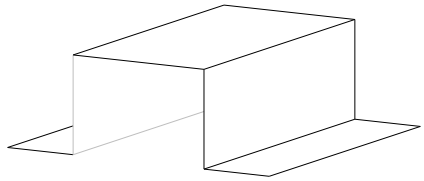
$$\mathrm{PSD}(k) = \frac{\sqrt{\pi}}{2\pi} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma\left(\alpha\right)} \frac{2\sigma^2\xi}{\left(1 + k^2\xi^2\right)^{\alpha+\frac{1}{2}}} \tag{5.1}$$

which we have also used in our previous study and seems a logical extension to the case of rough surfaces.[28] In the model of Eq. 5.1, roughness is defined with σ as the standard deviation, ξ the correlation length and α the roughness exponent. Random rough surfaces with a PSD equal to that of Eq. 5.1 are generated by using the method of Thorsos, which is explained for example in Ref. 27. The idea is to compute the inverse two-dimensional Fourier transform of the amplitude of the PSD with a random phase. We

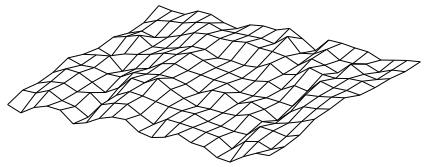
remark that the roughness generated via the method of Thorsos is biased. This can be corrected for by multiplying the resulting displacements of the roughness with a scalar, which is also explained in Ref. 27. We emphasize that the statistics of the rough surfaces generated in this way are isotropic.

5.2 Simulation of rough lines

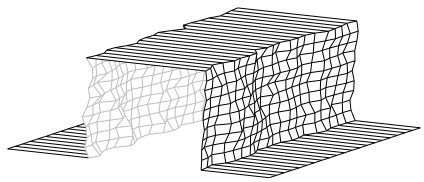
We would like to study a pattern of rough lines and spaces, where each line is made of PMMA located on a silicon substrate. The dimensions of the lines are 2 µm in length, 32 nm wide and 32 nm in height. We randomly generate rough lines as follows. At first we create a template line with flat surfaces of which, for example, a small section is shown in Fig. 5.1a. A randomly generated rough surface is made by using the PSD of Palasantzas, given by Eq. 5.1, and the method of Thorsos. The area of the rough surface is equal to that of the flat sidewall (2 µm × 32 nm). We define the rough surface with a resolution of 1 nm in both dimensions. The justification for this number will be given later, when we define the image pixel size and beam spot size. An example of a randomly generated rough surface is shown in Fig. 5.1b. Rough lines are produced by replacing the flat sidewalls of the line shown in Fig. 5.1a with randomly generated rough surfaces. The result of that is illustrated in Fig. 5.1c. We emphasize that the top of the PMMA line remains flat, and the silicon substrate remains flat too. This is analogous to the work of Li et al., and Lawson et al. In reality, the top of the line and the substrate are expected to be rough as well. For more realistic cases, we should include the effects of post lithographic processing as well. The effect of that could be, for instance, that the base of the sidewall is inclined or even curved. None of these effects are taken into account in this study, which means that we investigate a rather ideal case of SWR. We do not use a single isolated line, but instead, we symmetrize the sample in the direction perpendicular to the line by introducing ideal electron mirrors. The mirrors are located to the left and right of the isolated line such that the spacing between the lines equals exactly 32 nm. The resulting symmetrized sample is effectively an infinitely repeated pattern of lines and spaces. The reason



(a) A small section of an isolated line with flat surfaces is shown. This flat line is used as a template for the construction of a rough line.



(b) A small section of a randomly generated rough surface is shown. The roughness is generated by effectively sampling height displacements over a discretized flat surface by using the PSD function of Palasantzas and the method of Thorsos. The width and height of the rough surface is equal to the full size of the sidewall of the flat line of which a small section is illustrated in (a).



(c) A small section of a randomly generated isolated rough line is shown. The flat sidewalls shown in (a) are replaced with rough surfaces, of which one example is shown in (b).

Figure 5.1: The construction of a randomly generated isolated rough line is demonstrated. At first, a flat line (a) with a length of 2 µm, width of 32 nm and a height of 32 nm is constructed. The sidewalls of the flat line are replaced with randomly generated rough surfaces (b). The result of that is a randomly generated rough line of which a small section is shown in (c). Note that the top of the line and the substrate, on which the line is located, remains flat.

for this more complicated geometry, is to accommodate for the scattering of electrons into neighboring lines. Now that we have the sample defined, we can instruct the program to scan the line and count the secondaries as follows. An electron is instantaneously detected whenever the following two criteria are satisfied at the same time: (1) The electron is transmitted from material to vacuum and (2) the kinetic energy of the electron is less than 50 eV. We have assembled illustrations of induced scattering events as calculated by our home-built simulator in Fig. 5.2.

Let us now discuss the actual simulation of an image of a randomly generated rough line. The pixel size is fixed at the size of 0.43 nm × 2.7 nm (width times length), which is in accordance with our previous study of rough lines.[28] We choose a beam with an energy of 300 eV and a spot size of 3 nm. The decision for this spot size (instead of a smaller one) is due to the increased aberrations related to the low energy of the beam. The decision for the low beam energy is to reduce the risk of shrinkage that is involved in samples made of the organic resist PMMA. Note that the spot size and the pixel size in the direction of the edges are larger than the resolution of 1 nm at which the rough surfaces of the sidewalls are defined. Each pixel is exposed with 20 electrons on average following the Poisson distribution, by which we simulate the effect of illumination shot noise. The resulting dose is approximately 276 $\mu C/cm^2$ on average. An example of a simulated SEM image of a randomly generated rough line is shown in Fig. 5.3. There are a few remarks to be made with respect to this simulated image. In the first place, we have not simulated the effect of detector noise, i.e. the detection of electrons is assumed to be perfect. For a more realistic image we should include, for example, a detection threshold, detection efficiency, Gaussian-like background noise and additional Poisson noise. The precise position of a detector also plays a role in the image formation as the position of the detector could introduce a shadow effect. This means that we expect to find more noise in a real SEM image with 20 electrons per pixel on average than shown in Fig. 5.3. Moreover, in a real (CD-)SEM, there is also a control mechanism for image contrast. We, however, simply scaled the intensity linearly over the full range of secondary

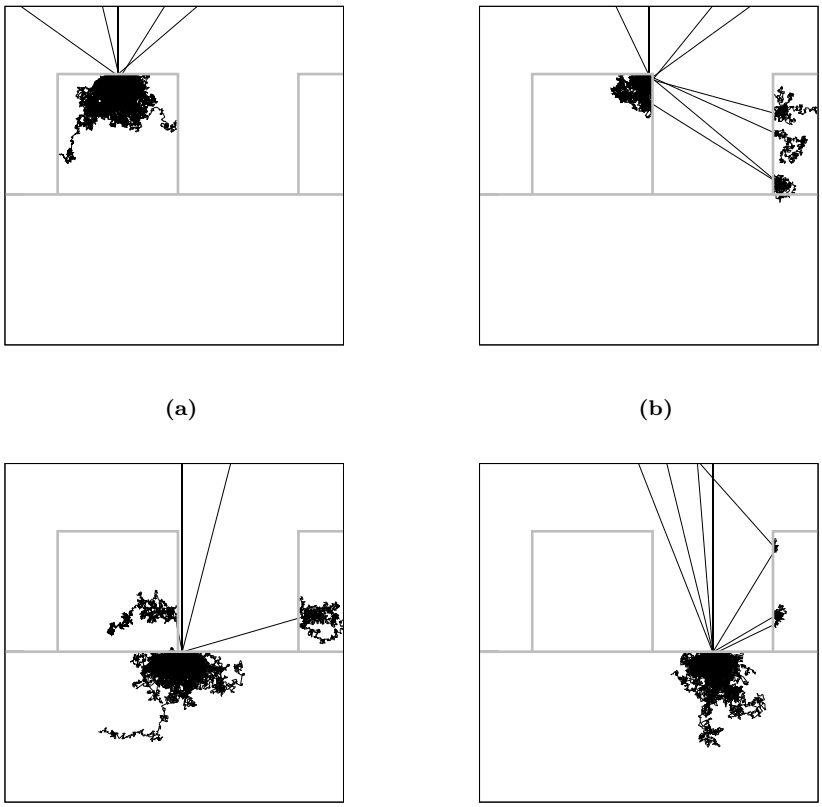


Figure 5.2: Demonstration of the scattering events induced by an incident stream of primary electron onto a sample of lines and spaces at different positions. The lines are made of PMMA located on a silicon substrate. The detection of secondary electrons with an energy less than 50 eV at the surface of the lines and substrate is not shown. In reality, the lines have rough sidewalls and the electron beam has a finite spot size of 3 nm. Notice that backscattered electrons may travel through vacuum into the neighboring lines where more scattering events are induced.

counts. Although none of these additional effects in the image formation are taken into account in our simulations, the resulting image, i.e. Fig. 5.3 still appears realistic.

In order to generate statistics, we repeat the simulation of a randomly generated rough line 10 times. This means that, for each roughness, we calculate a full image of rough lines and spaces such as shown in Fig. 5.4.

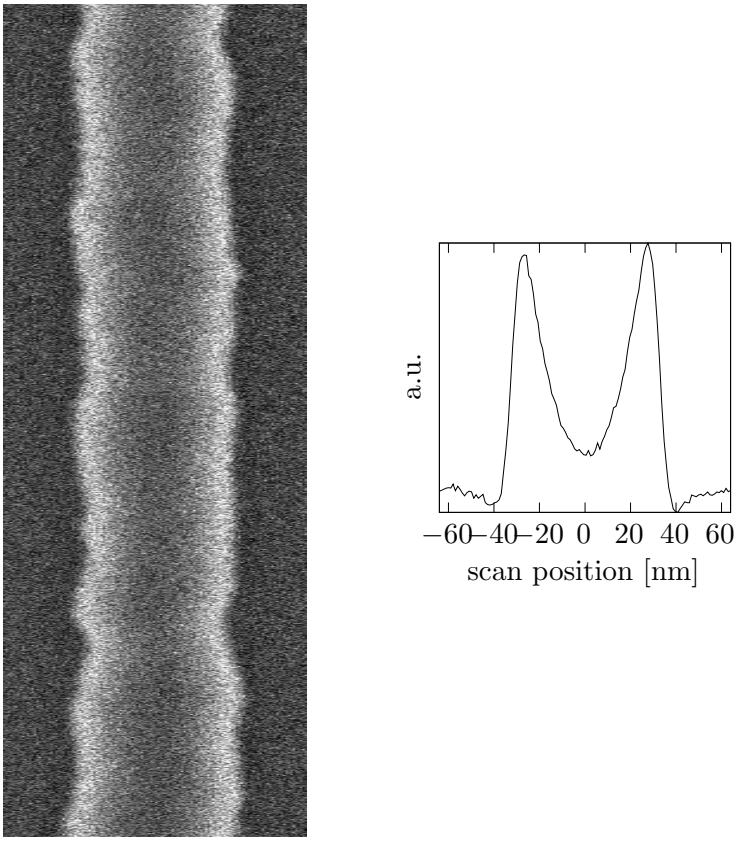


Figure 5.3: Simulated top-down image of a randomly generated rough line (left), including its integrated profile (right) is shown. The line is made of PMMA, located on a silicon substrate and is 1 µm long, 32 nm wide and 32 nm in height. The area of the image is 64 nm × 1 µm (width times length). The SWR that was generated has a standard deviation σ_{3D} of 1 nm, correlation length ξ_{3D} of 20 nm and a roughness exponent α_{3D} of 0.75. The pixel size is 0.43 nm × 2.7 nm (width times length). Each pixel is exposed with 20 primary electrons on average by sampling the Poisson distribution. The primary beam has a spot size of 3 nm and the kinetic energy of the primary incident electrons equals 300 eV. The detector for secondary electrons with an energy less than 50 eV is assumed to be perfect and does therefore not introduce additional noise.

The resulting image is used for analysis by applying the method as presented in our previous study.[28] In that method we (1) measure the edge displacements using a model for the integrated profile function, (2) determine the PSD by Fourier transform and finally (3) fit the measured PSD against Eq. 5.1 extended with a white noise term in order to capture the pixel noise. Let us discuss the two examples of the PSD analysis shown in Fig. 5.5. The SWR that was generated for both cases has an equal standard deviation σ_{3D} of 1 nm and equal roughness exponent α_{3D} of 0.75. The correlation length, however, was taken differently. In Fig. 5.5a we have used a correlation length ξ_{3D} of 6 nm, and for Fig. 5.5b we have used a correlation length ξ_{3D} of 25 nm. Note first of all that the bending point at the far left shifts in the direction of lower frequencies as the correlation length increases. This means that, in the case of the increased correlation length, the high frequencies, relative to the low frequencies, are more suppressed. This can also be seen from Eq. 5.1, where the correlation length ξ is coupled to the wave number k. Suppose, for the sake of argumentation, that the PSDs in Fig. 5.5 are unbiased measurements of the actual SWR. Note that the total variance σ_{3D}^2 is obtained by integrating Eq. 5.1. Since the standard deviation σ_{3D} is kept constant, the area under the PSDs must be equal as well. As we increase the correlation length, the PSD essentially shifts to the left, due to the coupling with wave number k, and the total area remains invariant because of an increase in power. This can also be seen from Eq. 5.1, where the power (and thus the integral, hence the variance) scales as a function of correlation length ξ.

Although the (pixel) noise is found to be close to equal (0.5 nm), the measured LER for Fig. 5.5a (0.63 nm) is smaller than for Fig. 5.5b (0.90 nm). This means that the determination of the SWR from top-down SEM images is biased. Since the actual three-dimensional roughness has a standard deviation σ_{3D} of 1 nm, we conclude that the bias in the LER has increased for decreasing correlation length ξ_{3D}.

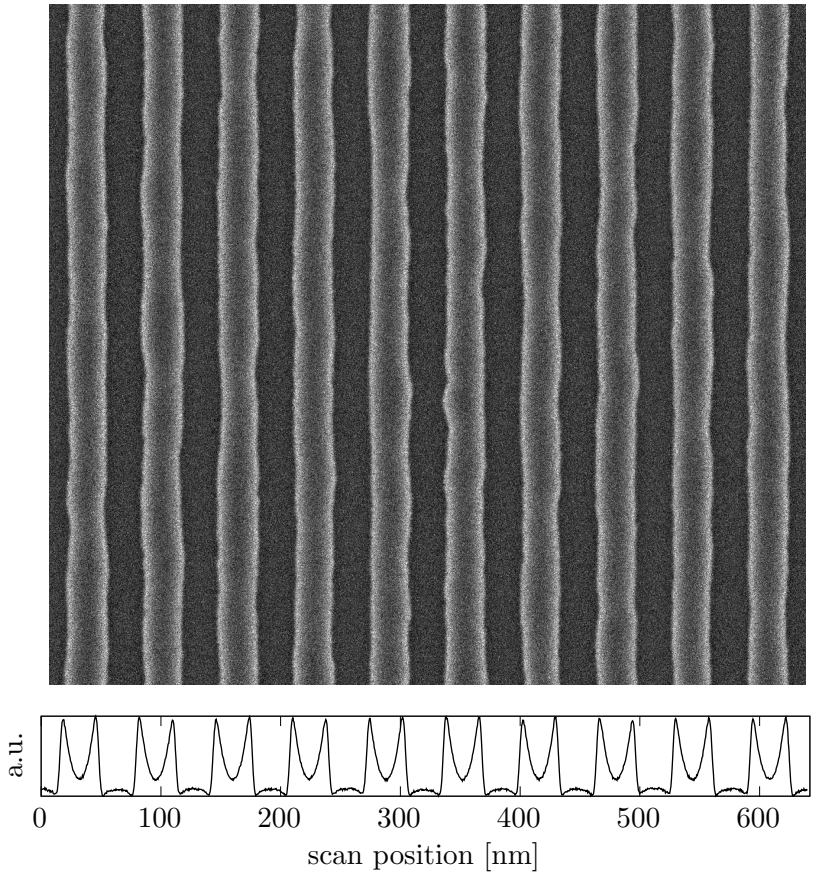
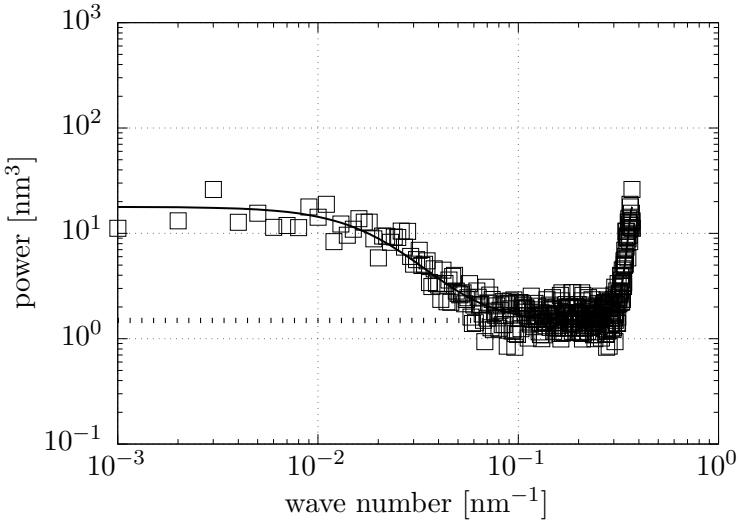
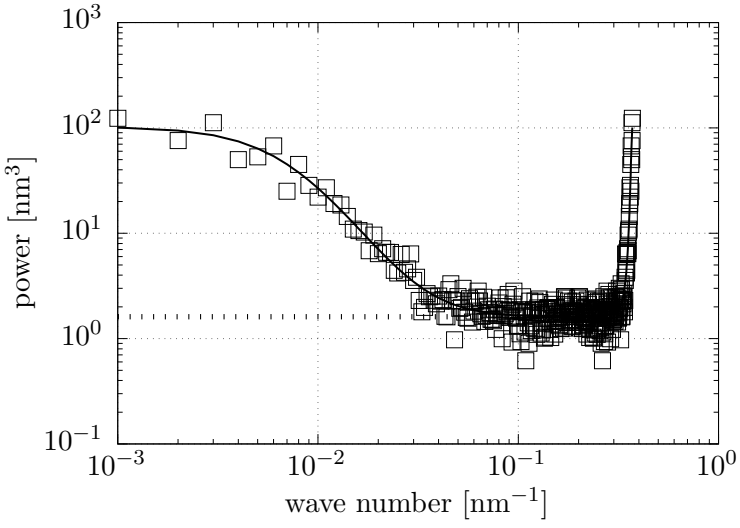


Figure 5.4: Simulated top-down image of randomly generated rough lines and spaces (top), including its integrated profile (bottom) is shown. This image is constructed by merging the individual images of ten randomly generated lines of which one is shown in Fig. 5.3. Each line is made of PMMA, located on a silicon substrate and is 1 µm long, 32 nm wide and 32 nm in height. The SWR that was generated has a standard deviation σ_{3D} of 1 nm, correlation length ξ_{3D} of 20 nm and a roughness exponent α_{3D} of 0.75. The pixel size of this image is 0.43 nm × 2.7 nm (width times length). Each pixel is exposed with 20 primary electrons on average by sampling the Poisson distribution. The primary beam has a spot size of 3 nm and the kinetic energy is set to 300 eV. The detector for secondary electrons with an energy less than 50 eV is assumed to be perfect and does therefore not introduce additional noise.



(a)



(b)

Figure 5.5: The measured PSD, including a model fit, derived from a simulated top-down image of randomly generated rough lines and spaces is shown. The open squares are the measurements obtained from analyzing ten lines, the solid line is the model fit (Palasantzas) and the dashed line is the corresponding (pixel) noise level. The SWR that was generated has a fixed standard deviation σ_{3D} of 1 nm and roughness exponent α_{3D} of 0.75. The only difference between (a) and (b) is the roughness exponent ξ_{3D}, which equals 6 nm and 25 nm respectively. The measured LER (one-sigma) for (a) is found by fitting and equals 0.90 nm with a pixel noise of 0.56 nm. The measured LER (one-sigma) for (b) is found by fitting and equals 0.63 nm with a pixel noise of 0.54 nm.

Now that we are able to create, simulate, measure and analyze randomly generated rough lines, we demonstrate and discuss our simulations in which we vary the parameters of the Palasantzas roughness model as given by Eq. 5.1. At first we vary the correlation length ξ_{3D} of the SWR while keeping the standard deviation σ_{3D} at the fixed value of 1 nm. The result of that simulation is shown in Fig. 5.6. We already concluded that the determination of the SWR in top-down SEM images is biased when changing the correlation length of the SWR. Here, in Fig. 5.6, we see that the bias in SWR determination is actually a non-linear function of the correlation length ξ_{3D} of the SWR. The conclusion from the work of Lawson et al. was that the LER is typically 50% smaller than the true SWR. In our simulation, this corresponds to a case where $\xi_{3D} < 5$ nm. In order to compare, we must know the effective correlation length ξ_{3D} for the SWR of PMMA in the mesoscopic roughness model of Lawson et al. Unfortunately, the parameters that they have used in the mesoscopic model for PMMA are not mentioned in Ref. 25 and a direct comparison can therefore not be made. In any case, the result of Fig. 5.6 indicates that the effective correlation length ξ_{3D} for the SWR of PMMA in the mesoscopic roughness model of Lawson et al. is probably less than 5 nm.

We have found, by trial and error, that a particular function fits our simulations very well,

$$\frac{\mathrm{LER}}{\sigma_{3D}} = \frac{1}{a^{-1} + (b^{-1} - a^{-1}) \exp\left(-c^{-1} \cdot \xi_{3D}\right)} \tag{5.2}$$

where the fit parameter $0 < a < 1$ relates to the lowest possible bias on the far right ($\lim_{\xi_{3D}\to\infty}$), $0 < b < 1$ relates to the highest possible bias on the left ($\lim_{\xi_{3D}\to 0}$) and c relates to the bending point. Notice that the LER/σ_{3D} is *always* less than one. We now question if and how the fit parameters a, b and c depend on (1) the kinetic energy of the primary beam and (2) the height of the feature. We therefore extended the study by repeating the simulation for the following kinetic energies: 100 eV, 300 eV and 500 eV. In addition, we have adjusted the height of the feature to 16 nm, 32 nm and 48 nm while maintaining a fixed width and spacing of

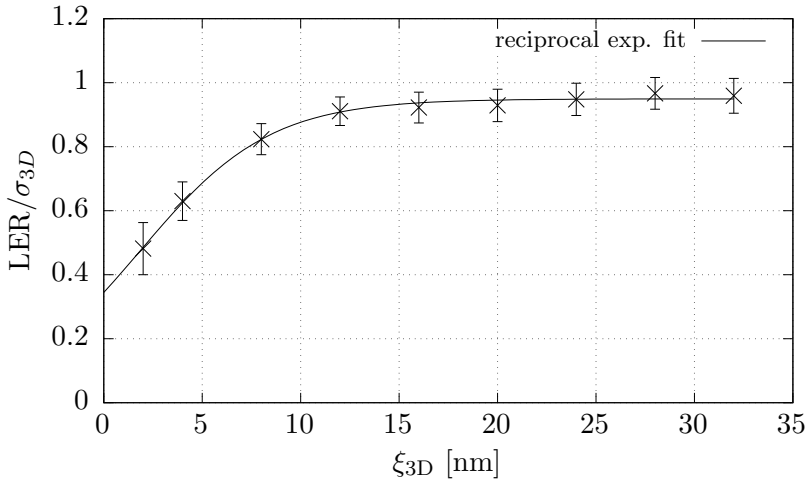
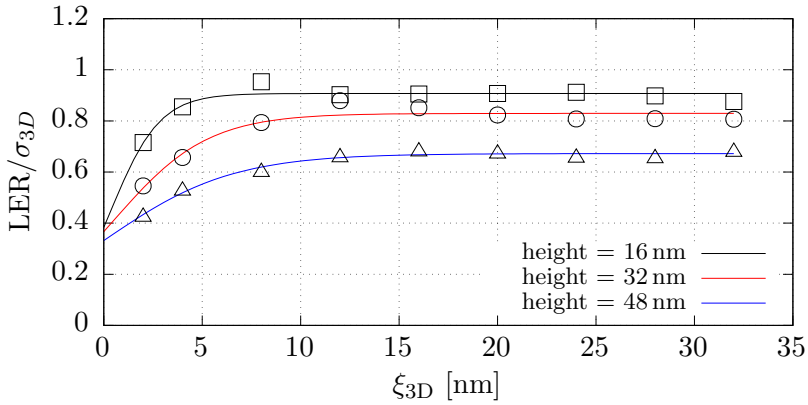
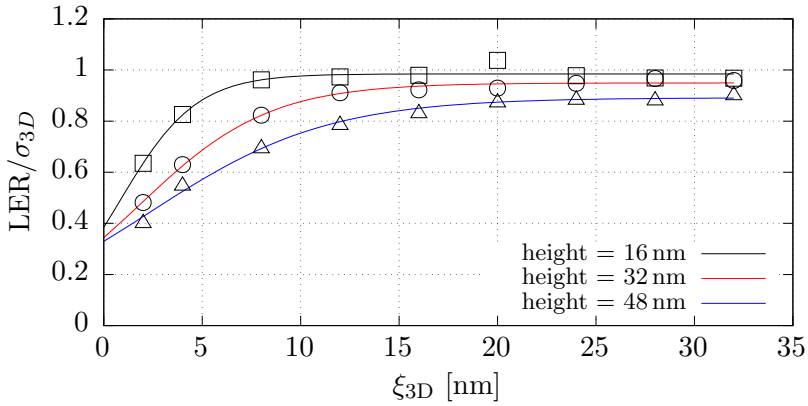


Figure 5.6: The measured LER (one-sigma) versus the correlation length ξ_{3D} of the SWR is shown. The markers are the result of analyzing the PSDs of randomly generated rough lines. The kinetic energy equals 300 eV and the feature height is fixed to the value of 32 nm. The standard deviation σ_{3D} of the SWR is fixed to the value of 1 nm, and the roughness exponent α_{3D} is fixed to the value of 0.75. The solid line corresponds to the best fit for a reciprocal exponential. Details are given in the main text.

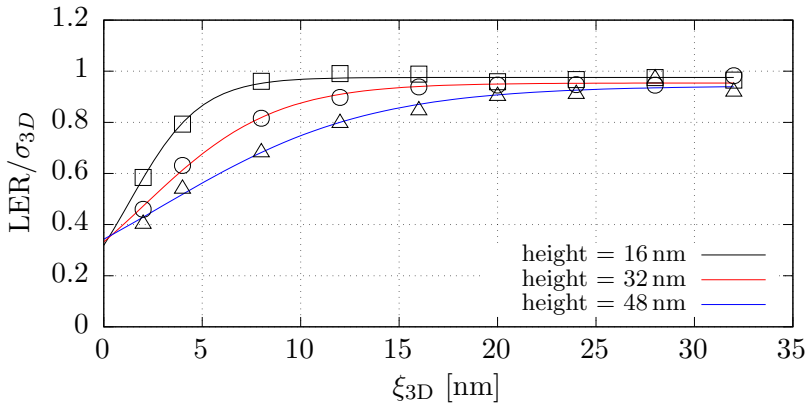
32 nm. The result of that simulation is shown in Fig. 5.7. We observe in Fig. 5.7a that the bias increases dramatically for increasing height. On the other hand, we see in Fig. 5.7b and Fig. 5.7c that the amount of the bias decreases for increasing primary energy. Also notice that the bending point in Fig. 5.7 shifts to the right, i.e. c becomes smaller, as we increase the energy of the primary beam. Moreover, notice for the height of 16 nm that a does not seem to increase further above 100 eV. In other words, it seems that the fit parameter a *saturates* asymptotically for increasing energies. To demonstrate this effect, we show the explicit energy dependence for the case of a fixed feature height of 48 nm in Fig. 5.8. Although the bias appears to decrease asymptotically for increasing energy, we repeat that LER is *always* lower than σ_{3D}. This brings us to the following conclusion: the bias in LER (with respect to σ_{3D}) can be reduced by tuning the kinetic energy of the primary beam with the feature height. The interested reader might point out that we have *only* considered varying the correlation length of the SWR. What about the roughness exponent? In a third simulation



(a) primary energy = 100 eV



(b) primary energy = 300 eV



(c) primary energy = 500 eV

Figure 5.7: The measured LER (one-sigma) versus the correlation length ξ_{3D} of the SWR is shown for various kinetic energies and feature heights. The standard deviation σ_{3D} of the SWR is fixed to the value of 1 nm, and the roughness exponent α_{3D} is fixed to the value of 0.75.

run, we have fixed the kinetic energy of the primary beam to 1 keV. The roughness exponent is then varied from 0.25 to 0.50 and finally to the value of 0.75 for a fixed feature height of 32 nm. The result of that simulation is shown in Fig. 5.9. We inevitably conclude that the fit parameters a, b and c depend on the roughness exponent as well.

We now see that the bias in the determination of the SWR from top-down SEM images can be very significant. But what exactly can we learn from this study? Although we would like that the LER only depends on the geometry of the sample (roughness). This study suggests, however, that the LER can have a strong dependence on (1) the kinetic energy of the beam, (2) feature height, (3) the frequency dependence of the SWR, i.e. correlation length and roughness exponent. In a previous study we proposed to reduce the kinetic energy of the electron beam to avoid resist shrinkage.[28] The current study, however, demonstrates that the bias in the LER measurement is reduced by choosing a larger kinetic energy for the primary beam. In other words, we reduce the risk of resist shrinkage at the cost of increasing the bias in the LER with respect to the SWR. At present, we have no explanation for the dependence of the bias in the determination of the SWR. We suggest to find an explanation in a follow up study.

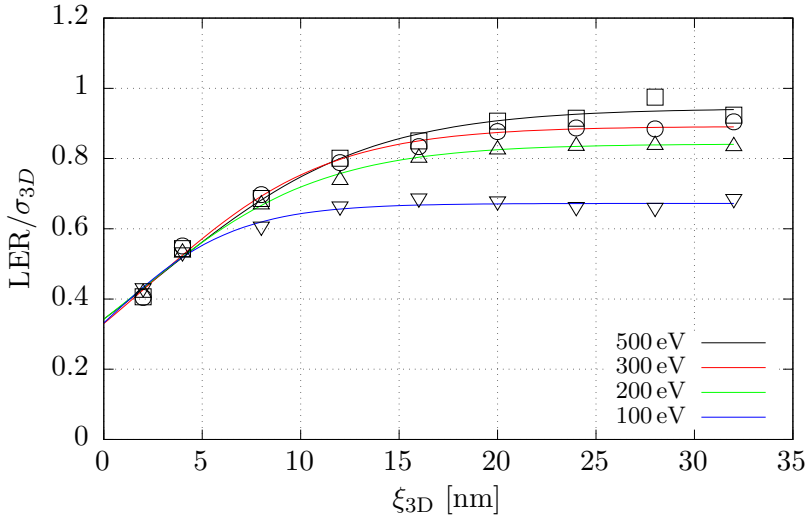


Figure 5.8: The measured LER (one-sigma) versus the correlation length ξ_{3D} of the SWR is shown for different kinetic energies. The feature height is fixed to the value of 48 nm. The standard deviation σ_{3D} of the SWR is fixed to the value of 1 nm, and the roughness exponent α_{3D} is fixed to the value of 0.75.

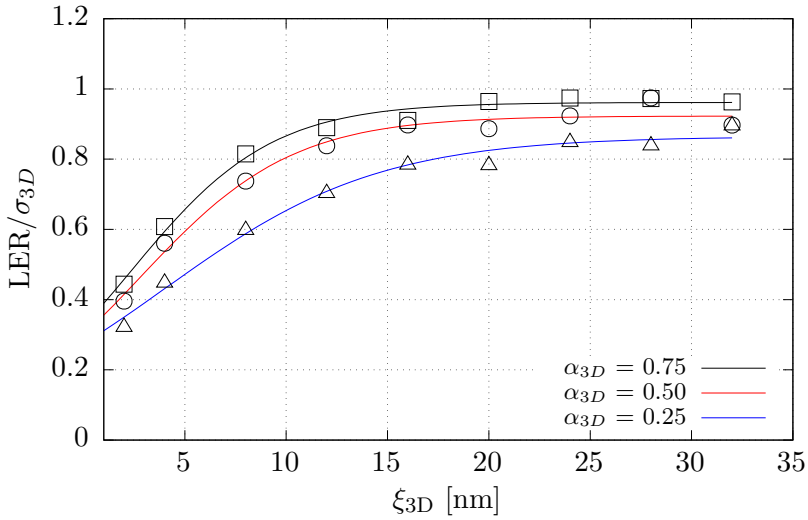


Figure 5.9: The measured LER (one-sigma) versus the correlation length ξ_{3D} of the SWR is shown for different roughness exponents. The kinetic energy equals 1 keV and the feature height is fixed to the value of 32 nm. The standard deviation σ_{3D} of the SWR is fixed to the value of 1 nm.

5.3 Conclusion

We have investigated the determination of the SWR from top-down SEM images by using a numerical study. In particular, we consider patterns of randomly generated rough PMMA lines located on a silicon substrate. Our lines have a length of 2 µm, a width of 32 nm and a height of 32 nm. The SWR is modeled by using the PSD of Palasantzas and generated by using the method of Thorsos. The rough surface is defined at a resolution of 1 nm in both dimensions. The top surface of our PMMA lines and the silicon substrate are modeled as flat.

We generate a SEM-like image of randomly generated rough lines by using a fixed pixel size of 0.43 nm × 2.7 nm (width times length), which is in accordance with our previous study on rough lines. In order to reduce the risk of shrinkage, we use a beam energy of 300 eV and a spot size of 3 nm, which is due to the aberrations caused by the low energy of the beam. Each pixel is exposed with 20 electrons on average, following the Poisson distribution to account for illumination shot noise. The resulting dose is approximately 276 µC/cm^2 on average. We have assumed that the detection of electrons is perfect and does not introduce additional noise to the SEM images.

Finally, we apply the PSD analysis as presented in our previous work while changing the standard deviation and the correlation length of the SWR in a controlled way. The measured LER is then compared against the actual SWR that was used to generate the rough lines. We conclude that the bias in the determination of the SWR is a non-linear function of the correlation length ξ_{3D} of the actual SWR. The bias in the determination of the SWR increases for decreasing correlation length of the actual SWR. We have observed that the LER can have a strong dependence on (1) the kinetic energy, (2) feature height, (3) the frequency dependence of the SWR. We conclude that the bias in the LER measurement can be reduced by choosing a higher kinetic energy, and that the risk of resist shrinkage is reduced at the cost of increasing the bias in the LER with respect to the SWR.

Chapter 6

Sensitivity analysis of scattering models

In the simulation of secondary electron yields (SEY) and secondary electron microscopy (SEM) images, there is always the question: "are we using the correct scattering cross-sections?".

In this chapter we artificially scale the scattering cross-sections, such that the probability for events associated with a particular model is either increased or decreased. We then investigate the influence of this adjustment on the calculated SEYs and simulated SEM images. At first we investigate the influence on the calculated SEY of pure and infinitely thick silicon. Then, we extend the analysis to the simulation of SEM images of three dimensional rough lines made of silicon located on a silicon substrate. Precisely how sensitive is LER with respect to scattering cross-sections?

The content of this chapter is an updated version of an article we have published in the Microelectronic Engineering: *Verduin, T., Lokhorst, S.R., Hagen, C.W., and Kruit, P., "Sensitivity of secondary electron yields and SEM images to scattering parameters in MC simulations," Microelectronic Engineering 155, 114-117 (2016).* This work was presented at the MNE 2015 conference in the Hague.

6.1 Introduction

In nano lithography, scanning electron microscopy (SEM) images of resist features are used for dimensional metrology and inspection. The question remains how to interpret the true size, shape and roughness characterization of the three dimensional resist features from two dimensional SEM images. For that purpose, simulation tools can be of great help. Reliable Monte-Carlo electron-matter interaction simulators exist [44, 79], but are unfortunately notoriously slow for SEM image simulation. The performance has been improved by using a triangulated mesh [24] and voxel based geometries [25]. Nevertheless, computation time can still be a problem. A practical example is the determination of line edge roughness (LER) using the power spectral density (PSD), which requires the simulation of multiple images [28]. Recently, we have reduced the computation time further by rewriting the GEANT4 extension from FEI company, see Ref. 44, for the purpose of SEM imaging and lithography simulations (see chapter 4). The subject of this article is to investigate the sensitivity of (1) calculated secondary electron yields (SEY) and (2) simulated SEM images of three dimensional patterns of lines and spaces to the parameters of the physical models. The idea is to artificially scale the scattering cross-sections, such that the probability for events associated with a particular model are either increased or decreased. The influence of this adjustment on the calculated SEY and simulated SEM images is then evaluated. By doing so, we can determine the importance of the individual scattering processes with respect to the final result.

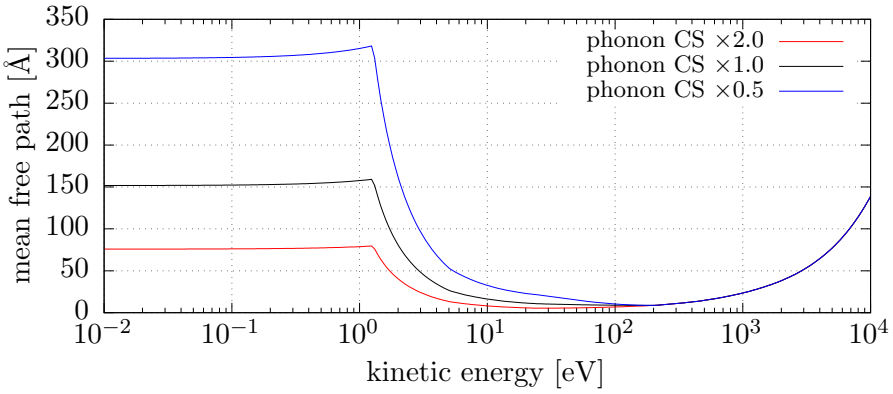
6.2 Model sensitivity analysis

We investigate the cross-section sensitivity by using our own high performance simulation tool (chapter 4) and discriminate between three scattering processes: inelastic scattering, elastic scattering and phonon scattering. For each and per scattering process, we multiply or halve the corresponding cross-section by a factor of two. The factor of two for the rescaling of the scattering cross-sections is argued as follows. In the refinements of Kieft and Bosch, a rescale of 3/2 is introduced to compensate the un-

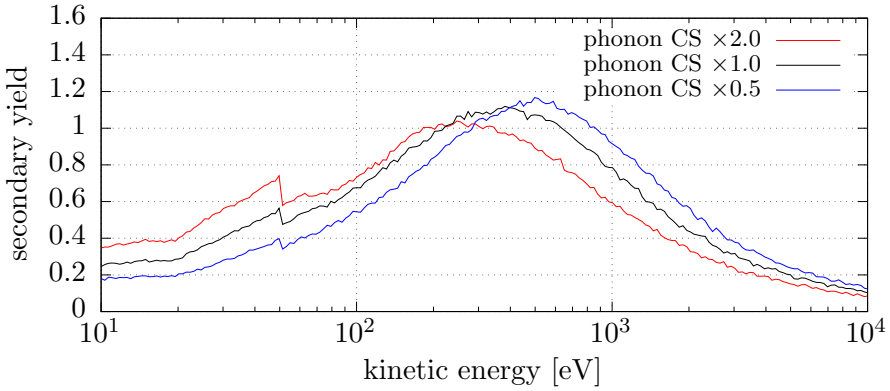
derestimation of Eq. 3.75 in chapter 3 with respect to experimental and theoretical sources. Also notice that the DCS (or inverse MFP) scales (see Fig. 3.5 in chapter 3) roughly with about a factor of two with respect to the components of the optical-model potential. Other contributions that will 'scale' the MFP are, for example, the number density, i.e. $\lambda = 1/(\rho_n \sigma)$ and the parameters in the pre-factor (temperature, sound velocity, deformation potential, ...) of the AC phonon model (Eq. 3.126 in chapter 3). Although nothing prevents the overall factor to be larger than two, the results obtained with a factor of two are still useful as it reveals the trend in the change of the observables.

Before we proceed, let us mention the key differences with respect to our previous publication.[93] First of all, the results shown in Ref. 93 are obtained by using the exact same scattering models as used in the simulator from FEI company.[44]. What is presented here follows the latest developments described in chapter 3. In other words, the physical models have been improved significantly with respect to our previous publication. The model for AC phonons in the present work, for example, is more accurate as it includes *three* acoustic modes (one longitudinal and two transversal modes), which must be compared to the single longitudinal mode used in Ref. 93 and Ref. 44. Furthermore, in comparison to the free electron mass used in Ref. 44, we consider effective masses (density of states and a conductivity mass) in the AC model. Another improvement is the use of the elastic scattering cross-sections from ELSEPA, which includes among the other subtle effects described in chapter 3, the muffin-tin based optical-model potential. In the work of Ref. 93 and Ref. 44, however, the free-atom elastic scattering cross-sections of Czyzewski[37] are used. We finally mention that the FEI simulator considers a peculiar (exponential decaying) energy-dependent particle filter at interfaces, which has as strong tendency to annihilate low energetic electrons. Although no explanation is given in the work of Kieft et al., that filter is *necessary* for SEYs to match with experiment. The energy-dependent filter is fortunately no longer necessary if we follow the (improved) physical models described in chapter 3. The results shown in Figs. 6.1, 6.2 and 6.3 are calculated SEYs of pure and

infinitely thick silicon. Similarly to what is done in chapter 4, we include a small layer of $\Delta = 5\,\text{Å}$ to account for surface effects. Within that layer we exclusively use the surface ELF, instead of the bulk ELF. In each of



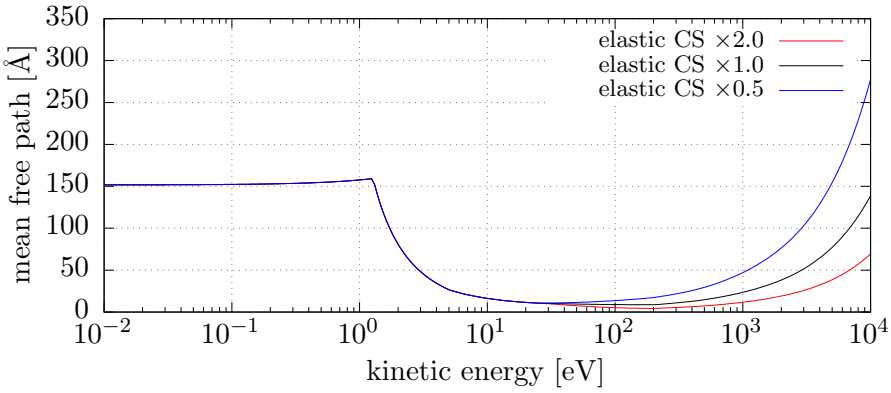
(a)



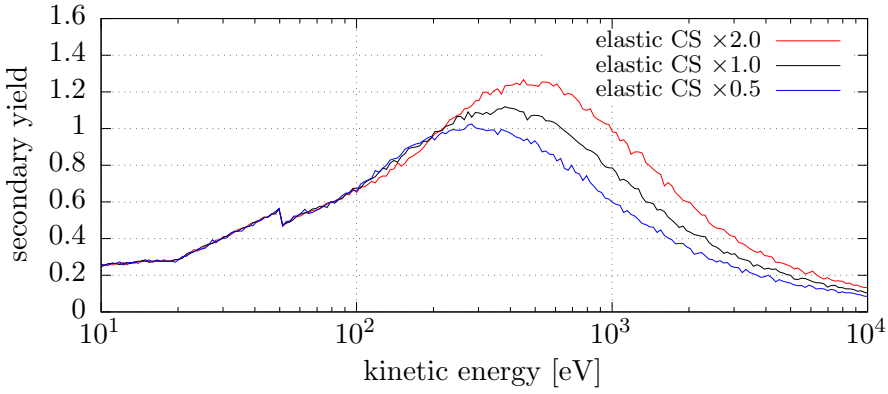
(b)

Figure 6.1: The influence of acoustic phonon scattering cross-sections on the SEY of silicon is shown. Notice that the acoustic phonons are, due to the quasi-elastic nature, part of the elastic scattering cross-sections (E<100 eV).

the three figures, we have examined the influence of a scattering process by scaling the scattering cross-sections associated with that particular process. Much to our surprise, we observe that the influence of the acoustic phonon scattering cross-sections in Fig. 6.1 is seen all the way up to the incident primary electron energy of 10 keV. This is a remarkable effect, because acoustic phonon scattering is a low-energy extension to the elastic scattering cross-sections and is only applied to kinetic energies less than approxi-



(a)

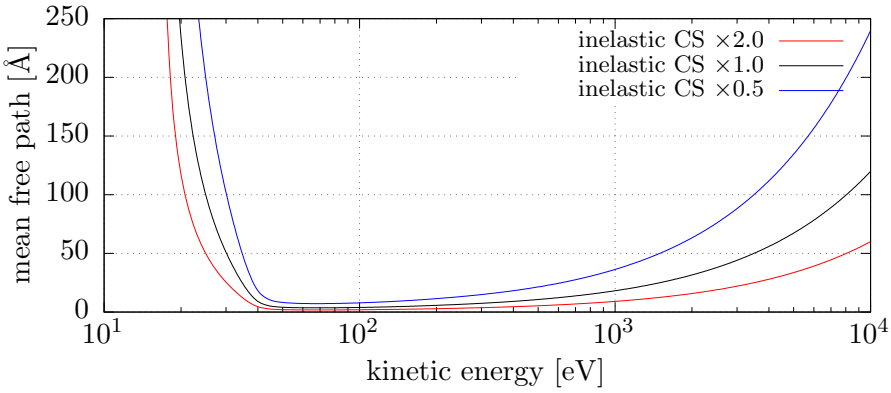


(b)

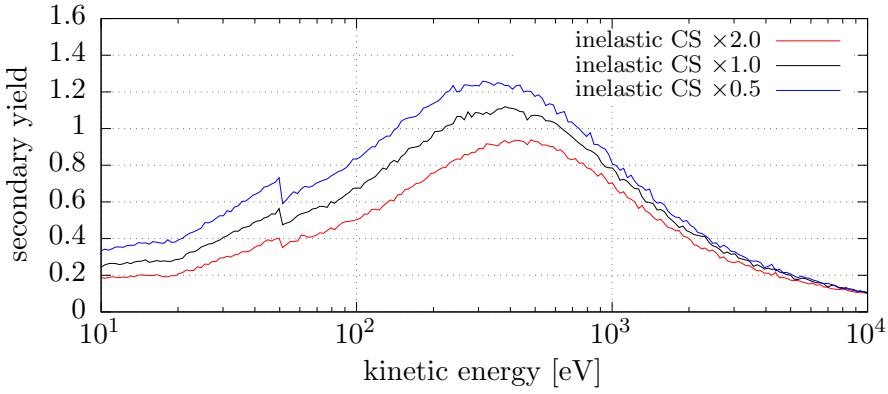
Figure 6.2: The influence of elastic Mott scattering cross-sections on the SEY of silicon is shown. Notice that energies lower than 100 eV in (a) are not scaled because that part is determined by the model for AC phonon scattering instead of Mott scattering.

mately 100 eV. The influence of the acoustic phonon interaction at primary energies higher than 100 eV must stem from the cascading process: electrons with a higher energy ultimately reach, via inelastic scattering events, energy scales at which the coupling to acoustic phonons becomes relevant. In Fig. 6.2, which corresponds to a scaling of Mott scattering cross-sections, we see no observable effect for electrons with an energy less than 200 eV. In fact, there should be no effect at all[1] because, similarly to the work of Ref. 44, Mott scattering cross-sections are only used for primary energies

[1] Differences due to statistics are excluded from this statement.



(a)



(b)

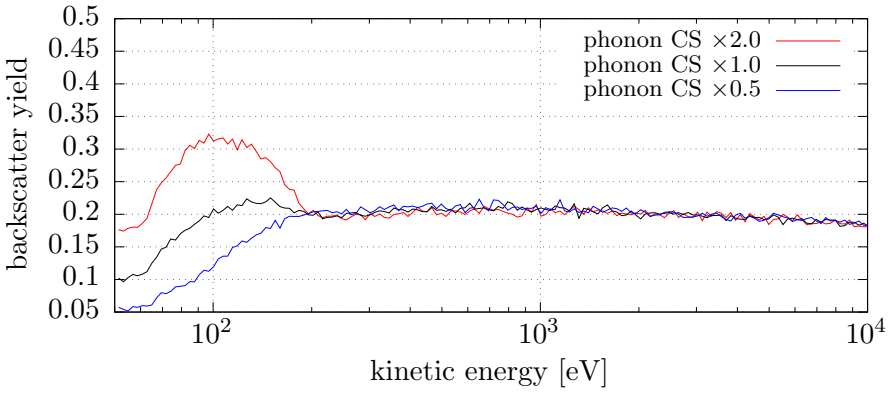
Figure 6.3: The influence of inelastic cross-sections on the SEY of silicon is shown.

ranging from 200 eV and upwards. Notice that in all three figures a sudden decrease in the SEY is observed near 50 eV. The explanation is found in the way that secondaries are distinguished from backscattered electrons. In the simulation, we denote similar to Ref. 44, electrons emitted into the vacuum with an energy less than 50 eV as secondaries and electrons with an energy higher than 50 eV as backscatters. The consequence of that definition is that, below 50 eV every emitted electron into the vacuum is simply counted as a secondary electron. From 50 eV and upwards, we suddenly start to make a distinction between secondary and backscattered electrons based on the kinetic energy. This causes the SEY to slightly decrease because, instead of *all* electrons, a smaller fraction (those with an energy less

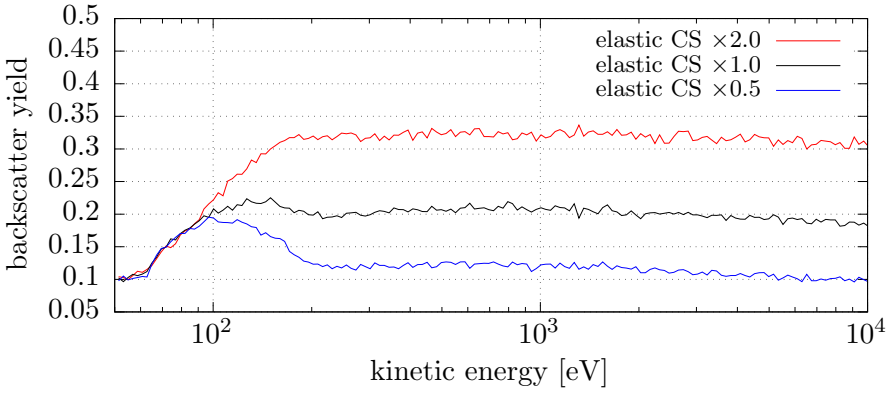
than 50 eV) are now counted as secondary electrons. We remark that the total yield, i.e. the sum of secondary and backscattered electrons, does not have the sudden decrease after 50 eV.

We have calculated for completeness the effect of the scaling of the scattering cross-sections to the backscatter yield (BSY). The result of that for each scattering process is shown in Fig. 6.4. Let us compare the effects in Fig. 6.4 to the influence on the SEY. We already see some interesting effects. First of all, notice that AC phonons play no role in the BSY for energies higher than 200 eV, which coincides with the boundary after which the elastic Mott model is used exclusively.[2] Although this particular energy is an artefact in the modeling of the physics, we can conclude nevertheless that the effect of AC phonons is different for backscatters than for secondary electrons: contrary to the SEY In fact, in the regime where the AC phonon scattering model is applied (less than 200 eV), we see that when the scattering cross-sections for AC phonons are multiplied by two, the corresponding BSY increases, whereas the SEY decreases. The reverse is found when we halve the AC phonon scattering cross-sections. Let us proceed with the the influence of the elastic scattering cross-sections. Notice that the trend on the SEY is comparable to the BSY. A multiplication factor of two results in an increase of the SEY as well as an increase of the BSY for kinetic energies above which we start to include the elastic Mott model (100 eV). However, the increase (or decrease) of the SEY diminishes for higher kinetic energies. Notice on the other hand that the increase (or decrease) of the BSY remains relatively constant for increasing kinetic energy. Finally, we see that the scaling of the inelastic scattering cross-sections affects the BSY on the entire energy scale. Here too we see that both the SEY and BSY increase (or decrease). Observe however that the influence of the inelastic scattering cross-sections is opposite with respect to AC phonon and elastic Mott scattering. When the scattering cross-sections are multiplied by two, the yields (SEY and BSY) increase when applied to AC phonons and elastic Mott, but decrease when applied to the inelastic scattering cross-sections

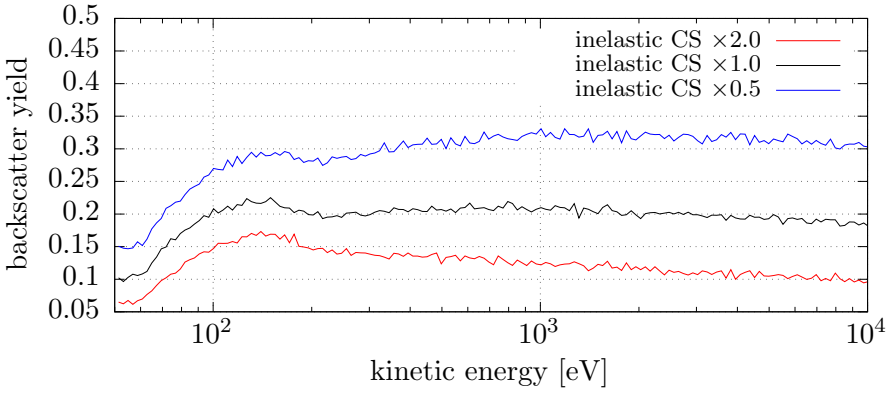
[2] Let us remind the reader that the model for AC phonon scattering is used exclusively below 100 eV and elastic Mott scattering is used exclusively above 200 eV. The models are interpolated in between.



(a)



(b)



(c)

Figure 6.4: The influence of the scaling of the scattering cross-sections on the BSY of silicon is shown. A comparison with the SEY is discussed in the main text.

We now focus on the sensitivity of simulated SEM images. Our approach is to calculate the SEM image of a fixed pattern of rough lines and spaces for four different cases. The lines are made of silicon, located on a silicon substrate, with dimensions 32 nm × 1 µm × 32 nm (width × length × height). The spacing between the lines equals 32 nm. The calculation, in which the pattern is exposed to an electron beam with energy 500 eV, is essentially identical to Ref. 78, except for the fact that for each case, we have scaled the scattering cross-sections of one particular scattering process. Let us emphasize that we have *not* included a surface layer in the calculation of the SEM image. What we have obtained is one SEM image corresponding to the default scattering cross-sections and three SEM images where either the phonon, elastic Mott or inelastic scattering cross-sections are multiplied by a factor of two. The influence on the resulting SEM images is best seen on the SEM signal profile[3] of a single line, which is shown in Fig. 6.5. The primary effect of the scaling of the scattering cross-sections is more or less signal in the SEM image. We expect that, in practice, the roughness characterization of the lines remains unaffected. To demonstrate this, we applied the profile based edge-detection method of Ref. 28 to all four SEM images. For the characterization of the roughness, we use the model of Palasantzas[26]. The result of the roughness characterization (the 3σ LER), including the estimation for the correlation length (ξ) and roughness exponent (α) is given in Table 6.1 for the four different cases.

[3]The SEM signal profile is discussed in detail in chapter 2.

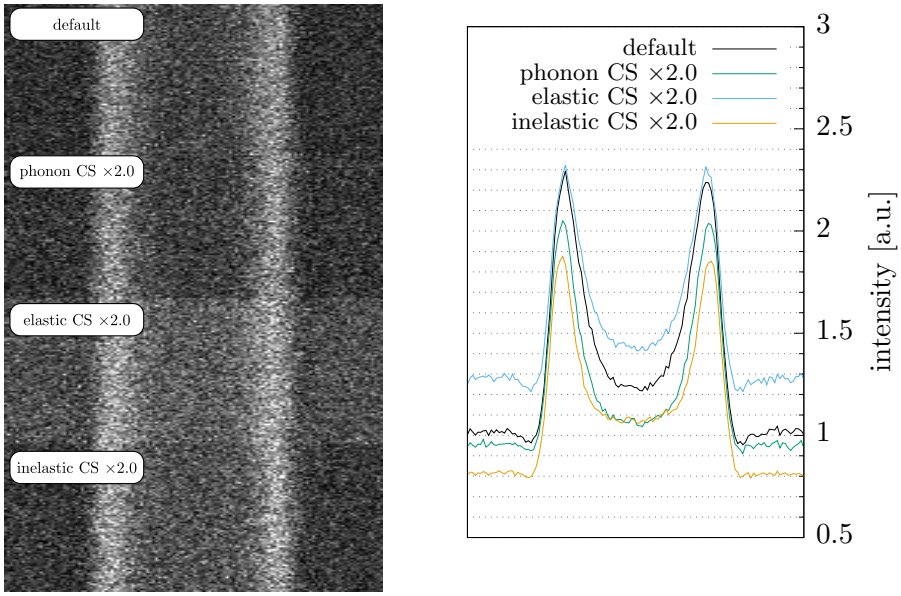


Figure 6.5: The influence of the scattering cross-sections on a SEM image of rough lines and spaces is shown. A composition of the influence of the scattering cross-sections to a small part of the SEM image is shown in the left image. The profile in the figure on the right is obtained by integrating the full SEM image (1 µm in length) from top to bottom.

scattering cross-sections	3σ [nm]	ξ [nm]	α
default	2.68 ± 0.02	30.3 ± 5.2	0.83 ± 0.09
phonon ×2.0	2.72 ± 0.03	34.3 ± 7.4	0.72 ± 0.12
elastic ×2.0	2.75 ± 0.02	32.6 ± 5.8	0.80 ± 0.09
inelastic ×2.0	2.70 ± 0.02	33.4 ± 7.8	0.75 ± 0.12

Table 6.1: The roughness characterization of four simulated SEM images of a fixed pattern of rough lines and spaces is shown. Each row corresponds to a separate simulation, where in each simulation only one scattering cross-section is multiplied by a factor of two.

Notice that the measured LER (3σ) is somewhat lower than the actual SWR (3σ of 3 nm). This is the bias which has been discussed in detail in the previous chapter. Nevertheless, we conclude that the 3σ of the LER is indeed not sensitive to the introduced changes in the scattering cross-sections. This demonstrates that the profile based edge-detection, as explained in Ref. 28, is not sensitive to the scaling of the scattering cross-sections.

6.3 Conclusion

Our sensitivity analysis demonstrates that acoustic phonon scattering plays a significant role in the calculation of SEYs. Although acoustic phonon scattering has a strong coupling to low energetic electrons, its influence on the SEY of pure silicon is seen all the way up to primary electron energy of 10 keV. We have extended the analysis to the simulation of SEM images of three dimensional rough lines made of silicon located on a silicon substrate. The scaling of the scattering cross-sections affects the contrast of the SEM image, but not the roughness characterization of the lines, i.e. the 3σ of the LER, correlation length and roughness exponent. This means that there is no need to increase the accuracy of the scattering cross-sections. SEM image simulation programs could perhaps be simplified, because the exact cross-sections are not that important.

Chapter 7

Simulation of shotnoise effects in lithography

In nanolithography, a resist layer is spin coated on a wafer with a substrate. The resist layer is then exposed optically or with a beam of electrons. In the case of a chemically amplified resist (CAR), the exposure results in the release of acids of in the resist layer. In the post-exposure baking (PEB) phase, the acids start to diffuse and break or cross-link molecular bonds which either become soluble or insoluble. The subject of computational lithography is to predict the effects in detail, such that the lithographic process, for example, can be optimized for throughput while minimizing the errors on feature dimensions to acceptable values.

In this chapter we investigate how quantum noise effects, starting from the initial exposure, affect the resulting feature. The approach is to study the formation of shotnoise induced SWR, for which we have developed a fast three-dimensional Monte-Carlo framework. In comparison to existing studies, where simplified models are used to avoid dramatic computation times, we simulate the full electron-beam resist interaction by using our rigorous GPU-accelerated electron-matter interaction program. After the simulated resist exposure, we then determine the release and diffusion of acids from photo acid generators (PAGs) in three dimensions. We will show the emergence of interesting effects due to the beam-resist interaction.

The content of this chapter is an updated version of an article we have published in the Proceedings of SPIE: *Verduin, T., Lokhorst, S.R., Hagen, C.W., Kruit, P., "Simulation of shotnoise induced side-wall roughness in electron lithography," Proceedings of SPIE 9778, Metrology, Inspection, and Process Control for Microlithography, 97781Z (2016).* The work was presented at the SPIE 2016 conference in San Jose.

7.1 Introduction

The throughput of a lithographic system is an important parameter in a typical production line. This parameter can be increased in several ways. In optical lithography, for example, a more powerful source is one solution. In electron-beam lithography, an increase of the number of parallel beams is another solution. In either case, the use of a more sensitive CAR would result in a reduction of the required exposure dose, and hence a reduction of the exposure time. In order to maximize throughput, it is tempting to choose the most sensitive CAR with the lowest possible illumination dose. In that limit, however, an increase of LER, and hence an insufficient control of critical dimension (CD) is observed.[1] This increase of LER is primarily caused by fundamental quantum noise (shotnoise) effects and becomes the dominant mechanism in the formation of LER.[2, 3, 4, 5, 6, 7, 8] Our attempt, in this theoretical study, is to develop an extended Monte-Carlo framework for the investigation of shotnoise induced SWR formation in CARs. We extend the existing studies in two different ways: (1) by including the interaction of the electron beam with a CAR by using our Monte-Carlo electron-matter interaction simulator and (2) by considering a full three dimensional model which, in contrast to Refs. 4, 5, 6, includes near-surface effects of acid diffusion. We expect that this extended model provides more insight in the effects on SWR of changing parameters such as resist thickness, acid diffusion and dose distribution. We will explain the method of SWR formation and give a striking example with analysis to demonstrate its use.

7.2 Method of SWR formation

At first, a (thin) layer of CAR is defined on top of an infinitely thick substrate. We now use the simulator of Ref. 78 for lithography: the electron beam is scanned over the sample such that a particular pattern is written. During the exposure, we record where the inelastic events take place within the resist layer. Next, we determine the distribution of released acids from the inelastic events under the following three assumptions. We assume that the initial distribution of photo acid generators (PAGs) in the resist layer is homogeneous and isotropic. Furthermore, we assume that every inelastic event in the simulation is associated with the release of an acid from a PAG. We emphasize that, for realistic studies, the probability for the release of an acid as a function of the kinetic energy of the electron must be included. In addition, we assume that a secondary electron is created in the inelastic event. The distribution of the acids is then used to determine the breaking/making of bonds in the resist by considering a diffusion like process in the post exposure baking (PB) phase. The (in)solubility at position (x, y, z) in the resist layer is determined by substituting each acid with a three dimensional Gaussian distribution,

$$S(x,y,z) \sim \sum_{i=0}^{n_{\text{acid}}-1} \exp \frac{-\left(x-\mu_x^{(i)}\right)^2 - \left(y-\mu_y^{(i)}\right)^2}{2r_d^2} \exp \frac{-\left(z-\mu_z^{(i)}\right)^2}{2r_d^2}$$

where n_{acid} is the number of released acids in the resist layer, $\mu_{x,y,z}$ the position of the acid and r_d the diffusion radius. We strive to replace this model by a more sophisticated model in the future. The determination of the (in)solubility close to the substrate and vacuum interface deserves special attention. The problem is that an acid cannot diffuse beyond the interface. Instead, we assume that an acid is reflected. The contribution of an acid to the (in)solubility is determined by using mirror symmetry at the vacuum and substrate interface.

It is achieved by substituting the following expression,

$$\exp \frac{-\left(z-\mu_z^{(i)}\right)^2}{2r_d^2} \longrightarrow$$
$$\exp \frac{-\left(z-\mu_z^{(i)}\right)^2}{2r_d^2} + \exp \frac{-\left(z+\mu_z^{(i)}-2z_s\right)^2}{2r_d^2} + \exp \frac{-\left(z+\mu_z^{(i)}-2z_v\right)^2}{2r_d^2}$$

where z_s defines the position of the substrate interface and z_v defines the position of the vacuum interface. The determination of the (in)solubility at a particular position in the resist layer is shown in Fig. 7.1 schematically. A three dimensional image of the (in)solubility of the exposed resist layer

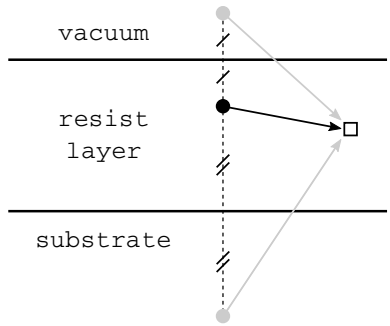


Figure 7.1: Schematic of the determination of the (in)solubility of a CAR at a particular location. The black sphere is the position of an acid in the resist layer, which is obtained with a Monte-Carlo electron-matter interaction simulator. Two mirrored acids (gray spheres) are virtually located in the substrate layer and in the vacuum. The mirrored acids are used for calculating the reflection of the acid diffusion with respect to the vacuum and substrate interface. The (in)solubility at the square marker is determined by evaluating a Gaussian kernel for the diffusion of the acid. The net (in)solubility at the square marker is obtained by accumulating the contributions of all released acids in the resist layer.

is then constructed by evaluating the expression $S(x, y, z)$ for each voxel. The computational complexity of the latter scales with the product of the dimensions of the three dimensional image and the number of acids in the resist layer. Typically, the computational complexity is $> 10^{12}$ and the computation time for the three dimensional image is reduced dramatically by using a graphics processing unit (GPU).

We now proceed similarly to the work of Refs. 4, 5, 6, where a threshold determines the boundary between exposed and unexposed resist. For a positive tone resist (PTR), intensities higher than the threshold in the three dimensional image are dissolvable in a developer. Analogously for a negative tone resist (NTR), intensities lower than the threshold are dissolvable. In reality, there is also a development phase, which we so far have ignored in this study. We acknowledge that this is a simplified view of post lithographic processing.

It is necessary, for realistic cases, that the elastic and inelastic scattering cross-sections of a CAR are available. The inelastic scattering cross-sections, for example, can be determined from optical loss functions, which can be obtained from optical experiments and from electron energy loss spectroscopy (EELS) measurements.[1] Unfortunately, we do not have the optical loss function of any particular CAR at our disposal. Therefore, we can only demonstrate the framework for the investigation of shotnoise induced SWR formation by using artificial scattering cross-sections. In our simulation tool, for example, we do have the scattering cross-sections for the organic resist PMMA. For the sake of demonstration, we will assume in the upcoming example that PMMA behaves as a CAR. We acknowledge that this is not a realistic approximation.

7.3 Example of lithographic exposure

Let us now give an example of the SWR formation. We define a 100 nm thick layer of PMMA on top of an infinitely thick silicon substrate and use the simulator of Ref. 78 for the purpose of lithography. We assume that the resist behaves as a NTR and write the pattern of an isolated line into the resist layer by scanning a beam with 20 keV electrons over an area of 32 nm×1 µm (width and length), see Fig. 7.2. During the exposure, we use a Gaussian spot-size with FWHM=FW50%=20 nm, beam step-size of 4 nm and a Poisson distributed exposure dose of 80 µC/cm^2, 60 µC/cm^2 and 40 µC/cm^2. During the exposure of the sample, we determine the

[1] More details on the sources for the scattering cross-sections, which are used in our Monte-Carlo electron-matter interaction simulator, can be found in Ref. 78.

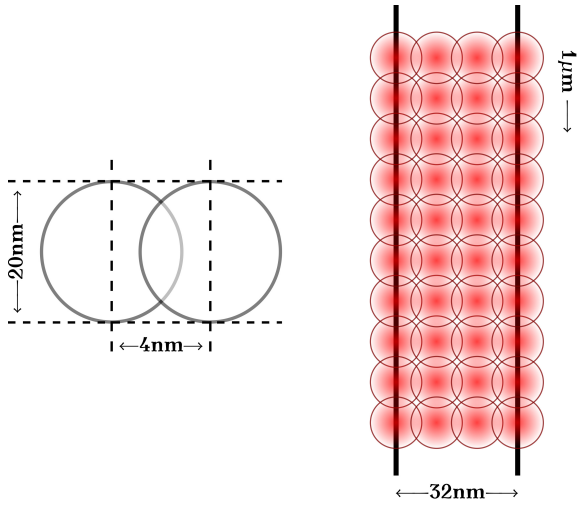


Figure 7.2: Schematic view of the lithographic exposure is shown. The left image shows the spot-size and beam step-size. The image on the right shows how the beam is scanned to form the pattern of an isolated line in the resist layer. Details are given in the main text.

distribution of released acids by recording the locations of the inelastic events within the resist layer. We now construct a three dimensional image of the (in)solubility of the resist layer within a cuboid of 128 nm(256px) wide, 800 nm(1024px) in length and 100 nm(128px) in height. The latter is obtained by evaluating expression $S(x, y, z)$ for each voxel in the three dimensional image, where we have used a three dimensional Gaussian with $\sigma_{x,y,z} = r_d$ =5 nm for the diffusion of the acid. Furthermore, we have normalized the three dimensional image such that the minimum value for the (in)solubility is zero and the maximum value is one. We are now ready to analyze the resulting image of the (in)solubility in different ways. Let us first look at the average (in)solubility in a xz-slice of the three dimensional image. The result of that is shown in Fig. 7.4, where we have marked the boundary between exposed and unexposed resist by contour lines for different values of the threshold. Notice the result of the scattering, which causes the distribution of the acids to broaden at increasing depth. In the remainder of the article, we arbitrarily choose a threshold of 0.5 with respect to the normalized (in)solubility to mark the boundary between exposed and

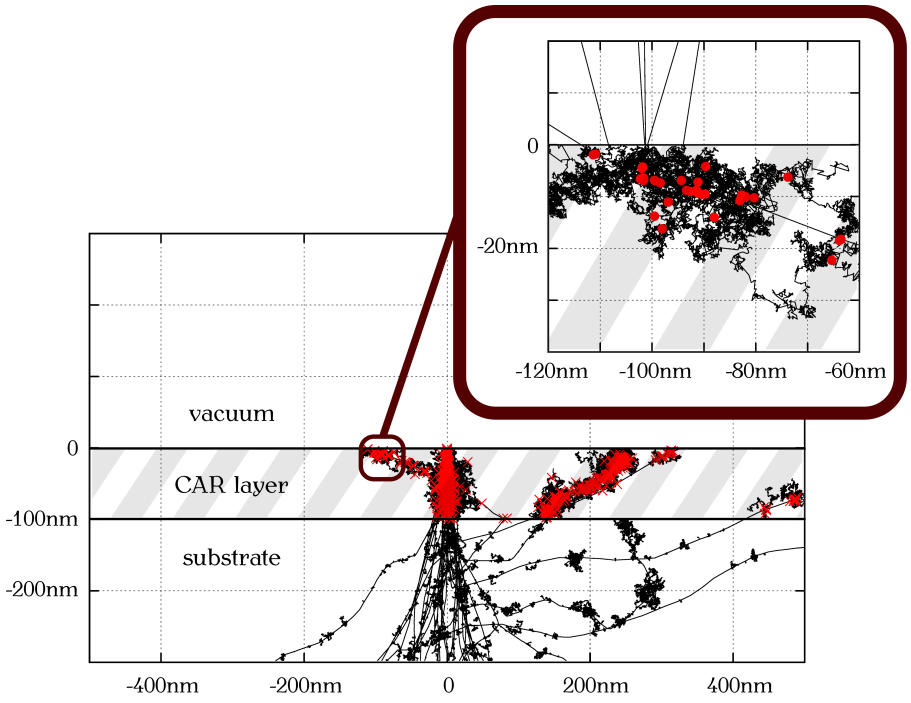


Figure 7.3: Demonstration of the interaction of the electron-beam with the resist layer is shown. A resist layer of 100 nm thickness is exposed to a beam of 20 keV electrons centered at zero. The electrons scatter (black lines) into the resist layer and facilitate the release of acids (red markers) from PAGs. The distribution of the acids is then used to determine the (in)solubility of the resist layer. Notice how electrons backscatter from the silicon substrate back into the resist layer. More details are given in the main text.

unexposed resist. A xy-slice of the three dimensional image, taken at a depth of 50 nm from the top surface of the resist layer, is shown in Fig. 7.5 for three different exposure doses. Notice that the boundary appears to have increasing roughness for decreasing dose (from left to right). We will discuss this effect in detail later on. Instead of a slice at a particular depth, we also construct a three dimensional view of the boundary between exposed and unexposed resist. The result for the left side of the boundary is shown in Fig. 7.6 for the three different exposure doses. Next, we determine the average LER as a function of the depth in the resist layer from the xy-slices by calculating the standard deviation (one-sigma) of the left and right boundary. If we repeat that procedure for all slices, ranging from

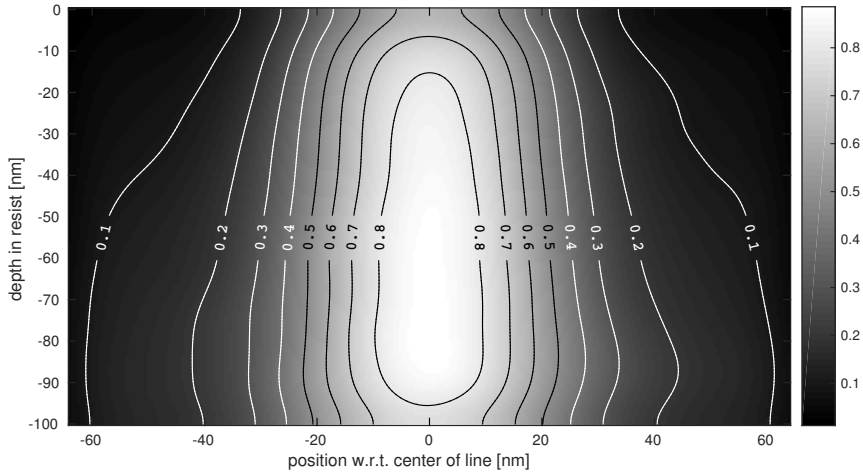
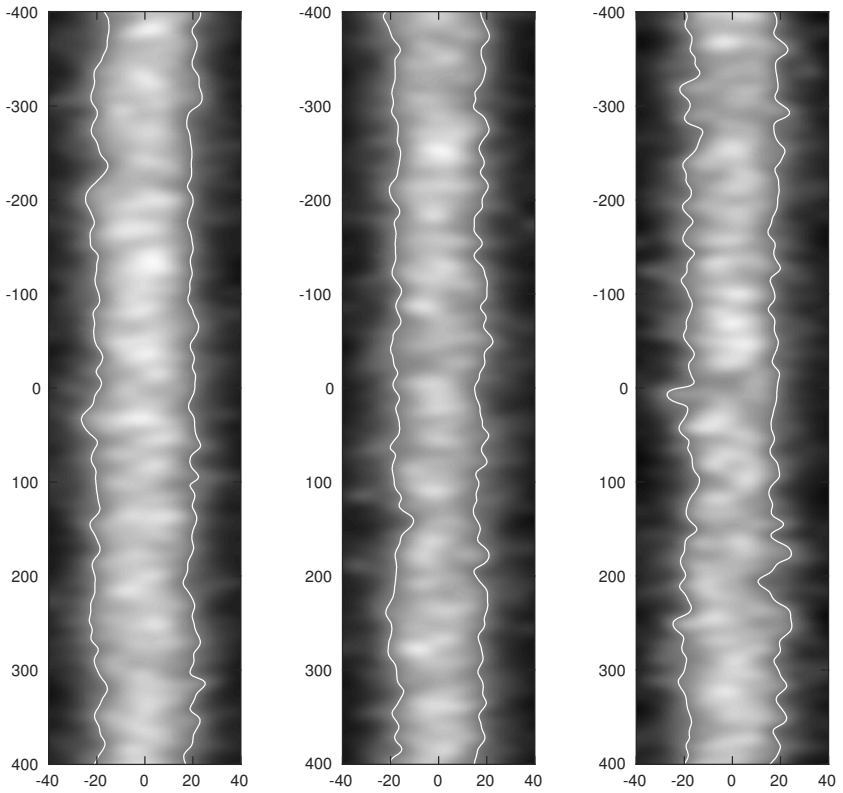


Figure 7.4: The average (in)solubility in a xz-slice of the three dimensional image. The sample consists of a 100 nm thick layer of CAR on top of an infinitely thick silicon substrate. A pattern of an isolated line is written into the resist layer by scanning a beam with 20 keV electrons over an area of 32 nm×1 µm (width and length) by using a Monte-Carlo electron-matter interaction simulator. The (in)solubility is obtained by accumulating the contributions of all acids in the resist layer, where we have used a three dimensional Gaussian with $\sigma_{x,y,z} = r_d$ =5 nm for the diffusion of the acid. The (in)solubility is normalized such that the minimum value for the (in)solubility is zero and the maximum value is one. The boundary between exposed and unexposed resist is shown by contour lines for different values of the threshold.

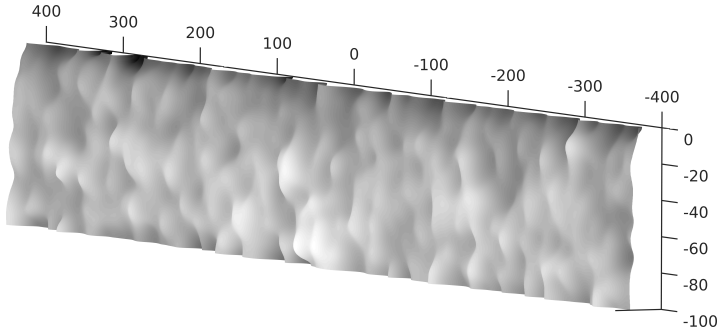
the top of the resist layer to the bottom of the substrate, we obtain the result of Fig. 7.7: the average LER as a function of the depth in the resist layer. We emphasize that in Fig. 7.7 the average LER is shown, and not *not* the boundary between exposed and unexposed resist. The observed increase in roughness for decreasing dose in Fig. 7.5 is confirmed by this calculation. This was also concluded in the work of Refs. 4, 5, 6 for two dimensional cases. In the current study, we observe the very same effect in a more sophisticated three dimensional model. In addition, we observe an interesting effect close to the vacuum and substrate interface: the roughness is increasing as we approach the interface. Let us discuss the observed trends for shotnoise effects in SWR formation. Shotnoise effects arise due to the nature of Poisson statistics: the number of acids varies with the square root of the number of acids, i.e. $n_{\mathrm{acid}} \pm \sqrt{n_{\mathrm{acid}}}$. The ratio between the



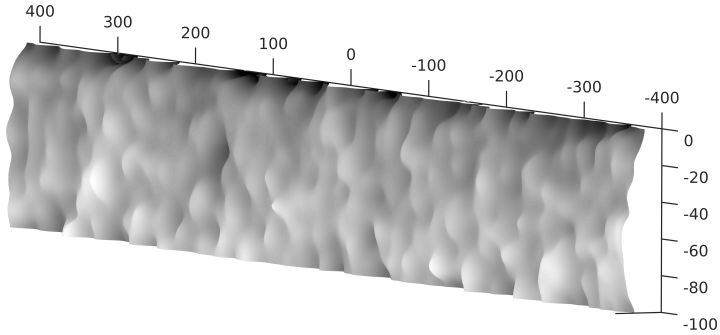
(a) D=80 µC/cm^2 **(b)** D=60 µC/cm^2 **(c)** D=40 µC/cm^2

Figure 7.5: A xy-slice of the simulated three dimensional (in)solubility of an exposed resist layer. Dimensions in the figures are given in units of nanometers. The slice is taken at a depth of 50 nm from the top surface of a 100 nm thick layer of CAR, which is located on a infinitely thick silicon substrate. The white wavy lines mark the boundary between exposed and unexposed resist for a threshold of 0.5. The three subfigures (a)-(c) correspond to a Poisson distributed exposure dose of respectively 80 µC/cm^2, 60 µC/cm^2 and 40 µC/cm^2.

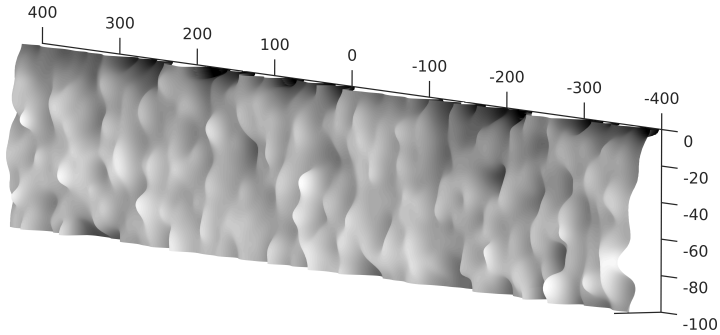
fluctuations and the number of acids is given by $\sqrt{n_{\text{acid}}}/n_{\text{acid}} = 1/\sqrt{n_{\text{acid}}}$. What this really means is that if we decrease the dose, and hence reduce the number of released acids, we get worse statistics. The result of that, as shown in Fig. 7.7, is an increase in roughness for decreasing dose. There are two contributions for the increase of roughness near the vacuum and substrate interface. One contribution relates to the actual number of acids, which is less near the interface than away from the interface. The explanation for this effect is as follows. Secondary electrons at the top of the resist



(a) D=80 µC/cm^2



(b) D=60 µC/cm^2



(c) D=40 µC/cm^2

Figure 7.6: A three dimensional view of the boundary between exposed and unexposed resist. The surfaces are obtained from a simulated exposure of a 100 nm thick layer of CAR, which is located on a infinitely thick silicon substrate. Similar to Fig. 7.5, the boundary between exposed and unexposed resist is determined with threshold of 0.5. The three subfigures (a)-(c) correspond to a Poisson distributed exposure dose of respectively 80 µC/cm^2, 60 µC/cm^2 and 40 µC/cm^2.

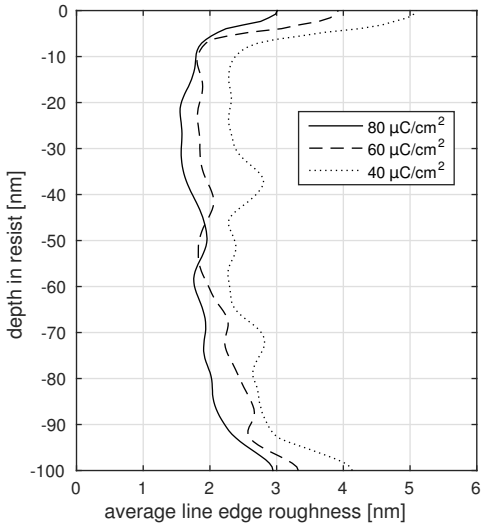


Figure 7.7: The average LER as a function of the depth. The figure is derived from xy-slices of the three dimensional image of the (in)solubility. The boundary between exposed and unexposed resist is determined with a threshold of 0.5. Please note that the average LER is shown, and *not* the boundary between exposed and unexposed resist. The average LER is obtained by calculating the standard deviation (one-sigma) of the left and right boundary. By considering all slices, ranging from the top of the resist layer to the bottom of the substrate, the average LER as a function of the depth from the top surface of the resist layer is obtained.

layer escape into vacuum, and hence cannot contribute any further to acid release within the resist layer. At the bottom we have a similar effect: secondary electrons scatter into the silicon substrate. The difference is (with respect to the vacuum side) that secondary electrons are created in the silicon substrate. There are, however, fewer secondary electrons scattering from the silicon substrate back into the resist layer. Another contribution stems from the fact that no acids are found on the vacuum side nor on the substrate side. We will demonstrate the consequence on the statistics by example. Suppose that the distribution of the acids in the resist layer is homogeneous and isotropic, i.e. the probability to find an acid anywhere in the resist layer is independent of position. The (in)solubility in the center of the resist layer is primarily determined by the sum of the contributions from the $n \pm \sqrt{n}$ acids surrounding that position. There are no acids found

beyond the interface and so near the interface we only have $\frac{1}{2}n \pm \sqrt{\frac{1}{2}n}$ acids in the resist layer to account for the (in)solubility. Remember that the acids in the resist layer are reflected by using mirror symmetry and hence we effectively have $n \pm \sqrt{2}\sqrt{n}$ (mirrored) acids contributing to the (in)solubility, see Fig. 7.8. We conclude that the statistics near an interface

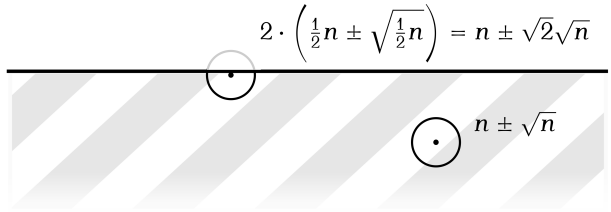


Figure 7.8: Illustration of shotnoise effects in the vicinity of an interface. Notice that the number of acids that contribute to the (in)solubility is less near the interface than away from the interface. In comparison to the bulk of resist layer, only half of the acids can contribute near the vacuum interface. More details are given in the main text.

are worse because (1) the number of acids near the interface is less than away from the interface and (2) because there are no acids found beyond the interface.

7.4 Power spectral density analysis

To complete the analysis, we now proceed by feeding the simulated resist feature into our CD-SEM imaging simulator. The construction of the rough line is analogous to what is done in Chapter 5 (see Fig. 5.1). Let us remind the reader that in Chapter 5 we have used the model of Palasantzas with the method of Thorsos to generate the rough surface for the sidewall. In this study, we use the rough surface obtained from the simulated lithographic exposure (see Fig. 7.6). The idea is to construct a two-dimensional height map of the sidewall and then triangulate the surface in a regular way. An example of the latter is shown in Fig. 7.9 for a sidewall obtained from a simulated lithographic exposure. Next, we add the substrate layer (silicon) and enclose the sample with electron detectors. The resulting triangulated mesh of the isolated line is then fed into the CD-SEM image simulator.

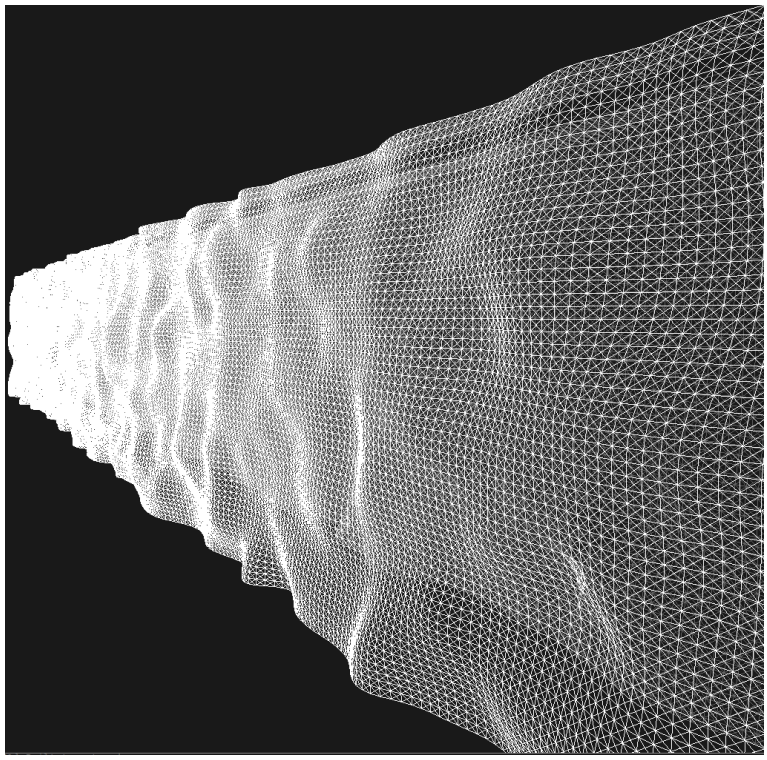
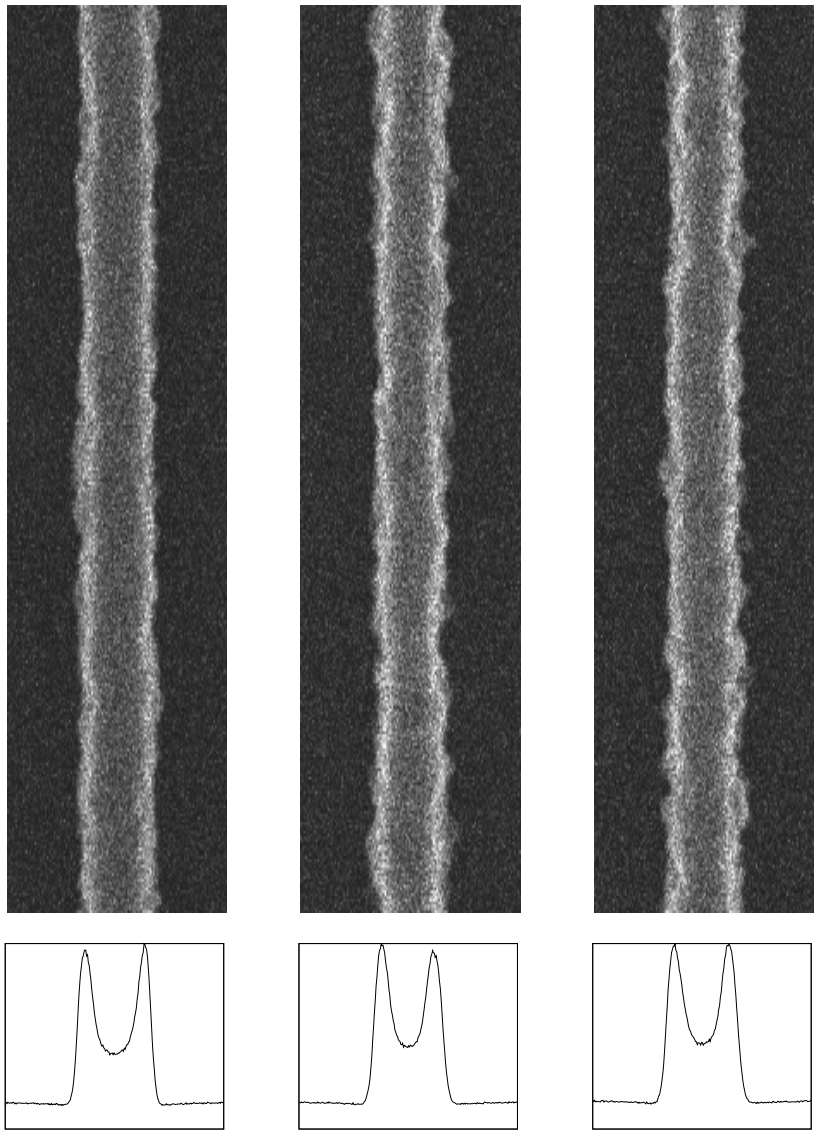


Figure 7.9: An illustration of a triangulated sidewall is shown. The triangulated sidewall is used for defining the geometry of an isolated line. The method of construction of the isolated line is analogous to the procedure explained in Chapter 5 with the exception of a custom generated height map. The length of the line is 650 nm and the height equals 100 nm. More details are given in the main text.

We then generate CD-SEM images by exposing the isolated line with three different primary energies, i.e. 0.5 keV, 1 keV and 5 keV. The resulting CD-SEM images taken with 5 keV are shown in Fig. 7.10. We now would like, similar to chapter 5, determine the PSDs of the simulated CD-SEM images and investigate the influence of the lithographic exposure dose and imaging energy on the LER. To that order we apply the edge detection technique of Chapter 2 and determine the PSDs of the simulated CD-SEM images. An example of a PSD obtained in this way is shown in Fig. 7.11. For the latter we have used a lithographic exposure dose of 80 $\mu C/cm^2$ and imaging energy of 5 keV. Let us make a few remarks about the PSD shown in Fig. 7.11. First of all, by looking at Fig. 2.10d in Chapter 2, we see



(a) D=80 µC/cm^2 **(b)** D=60 µC/cm^2 **(c)** D=40 µC/cm^2

Figure 7.10: The simulated CD-SEM images of an isolated line obtained with various lithographic exposure dose are shown. Each CD-SEM image has a field of view of 128 nm×650 nm (width times length). The images shown in (a), (b) and (c) are all obtained with an electron energy of 5 keV. The integrated profiles are shown at the bottom. More details are given in the main text.

that a single image of 10 lines with an imaging dose of approx. $10\,\mu C/cm^2$ results in an error on the LER of about 10%. This time, however, we have used the extreme case of a single line in the determination of the PSD. The latter is only possible if we compensate for the lack of statistics by using a high imaging dose ($2000\,\mu C/cm^2$). Second, notice that the shape of the PSD fits the description of Palasantzas model. This is to be compared to the results of Chapter 5, where we have explicitly used the Palasantzas model to generate the sidewall roughness. In the present study we made no such assumption on the model of the roughness, and still a Palasantzas-like roughness emerges. This means that the resulting PSD (LER, correlation length and roughness exponent) is a complex mixture of lithographic conditions (spot-size, beam step-size, exposure dose, beam energy, diffusion radius, (in)solubility threshold, ...) and imaging conditions (electron detectors, spot-size, beam step-size, imaging dose, beam energy, ...).

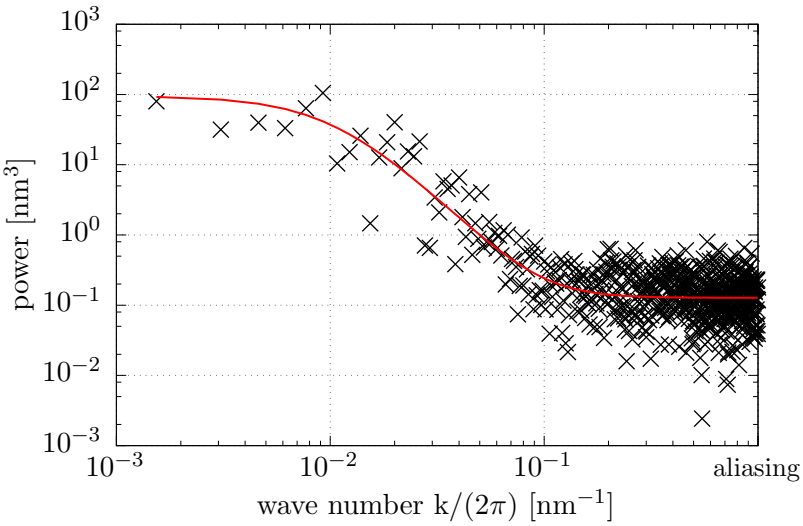


Figure 7.11: An example of a PSD obtained from a simulated CD-SEM image of a simulated isolated line is shown. For this particular example we have used a lithographic exposure dose of $80\,\mu C/cm^2$. The corresponding PSD is obtained by analyzing the simulated CD-SEM image by using 5 keV electrons. The solid line (red) is the best fit of the Palasantzas model, which has been extended with a term to capture the total noise. More details are given in the main text.

exposure dose [µC/cm^2]	imaging energy [keV]	LER [nm]
80	0.5	1.32 ± 0.13
80	1.0	1.24 ± 0.12
80	5.0	0.98 ± 0.10
60	0.5	1.63 ± 0.16
60	1.0	1.56 ± 0.16
60	5.0	1.24 ± 0.12
40	0.5	1.72 ± 0.17
40	1.0	1.56 ± 0.16
40	5.0	1.32 ± 0.13

Table 7.1: The LER and the correlation length resulting from the PSD analysis for various lithographic exposure dose and imaging energies are shown. More details are given in the main text.

We have made Table 7.1 to show the measured LER for various lithographic exposures and imaging energies. We have obtained each LER by fitting the corresponding PSD to the model of Palasantzas extended with a term to capture the total noise. The errors in the LER are estimated by extrapolating Fig. 2.17 (Chapter 2) to a single line with two edges. Although the resulting errors are substantial, we can still see two interesting trends. The first trend is that for a fixed lithographic exposure dose, the LER is decreasing for increasing imaging energy. This is a confirmation of the energy-dependent effect on the LER as described in Chapter 5. Another trend is that for a fixed imaging energy, the LER increases for decreasing lithographic exposure dose. For example, we see for an imaging energy of 500 eV that the LER increases from 1.32 nm at 80 µC/cm^2 to 1.72 nm at 40 µC/cm^2. The same trend is visible for the other imaging energies. This is the expected shotnoise effect caused by dose statistics of the illumination. We would like to end this chapter with a summary outline of the presented methodology for ab-initio SWR analysis:

- Define the initial geometry for the exposure. In the simplest case we have a single resist layer with a certain thickness located on top of a infinitely thick substrate.

- Specify an exposure strategy for a particular pattern by tuning various parameters: spot-size, beam step-size, beam energy and dose.

- Expose the geometry using the GPU accelerated electron-matter interaction simulator (Chapter 4). During the exposure, the distribution of released acids is determined.
- Simulate the PEB phase by allowing the released acids to diffuse with a certain radius. The latter includes reflection at the interfaces by symmetrization of the acid distribution.
- Generate a three dimensional image where each voxel relates to the probability of (in)solubility of the resist.
- Determine the boundary between exposed and unexposed resist by choosing a (in)solubility threshold.
- Construct a triangulated mesh of the resulting resist feature (Chapter 5) and include the remaining components of the geometry (substrate for example).
- Add electron detectors to the geometry and specify imaging conditions: field of view, spot-size, beam step-size, imaging energy and imaging dose.
- Simulate the acquisition of the CD-SEM image using the same GPU electron-matter interaction simulator (Chapter 4).
- Apply the edge detection method of Chapter 2 and determine the PSD.
- Measure the parameters of the PSD by fitting the model of Palasantzas extended with a term to capture the total noise.

The elaborate calculation allows the investigation of trends in the metrology of CD-SEM images by varying lithographic exposure conditions. An example is the minimization of the LER by fine tuning the writing strategy (spot-size, beam step-size) and by controlling the diffusion radius of the electrons. The latter is most interesting for industrial applications when the process is optimized for a reduced lithographic exposure dose, and hence a higher throughput of the lithography process.

7.5 Conclusion

We have succeeded in creating a fast three dimensional Monte-Carlo framework for the investigation of shotnoise induced SWR formation. In comparison with existing studies, we simulate the electron-resist interaction by using an advanced Monte-Carlo program. We have considered a simplified model for the release of an acid from a PAG and used three dimensional Gaussian diffusion for the acids. The latter includes reflection of acids at the vacuum and substrate interface by using mirror symmetry. We can already see interesting effects such as: surface effects, broadening of the acid distribution in the resist layer due to the scattering and roughness as a function of the penetration depth.

The exposed resist gives rise to a three dimensional feature which is then fed into our CD-SEM image simulator, which is the same Monte-Carlo simulator used for the lithographic exposure but now applied for the purpose of imaging. The LER in the resulting two dimensional top-down image is further processed using PSD analysis. What is observed is that the resulting PSD shares the same characteristics of the Palasantzas model and thus can be described effectively by the LER, correlation length, roughness exponent and a (white) noise term. In a case study we have varied the lithographic exposure dose for the pattern of a single isolated line. For each line, we have generated a simulated CD-SEM image by using three different imaging energies. The analysis of the ensemble of PSDs gives us two confirming trends. The first trend is that for a fixed lithographic exposure dose, the LER is decreasing for increasing imaging energy. This is a confirmation of the energy-dependent effect on the LER and is in agreement with the result of Chapter 5. Another trend is that for a fixed imaging energy, the LER increases for decreasing lithographic exposure dose. The increase in LER is the expected shotnoise effect caused by dose statistics in the illumination.

Chapter 8

Conclusions

The following conclusions can be drawn from the thesis:

1. We conclude that low dose CD-SEM image analysis is a necessary prerequisite for reliable characterization of CD and LER. The most important reason is to avoid resist shrinkage.

2. We have shown that the effect of transversal and longitudinal filtering in the determination of LER from CD-SEM images cannot be ignored, even when considering an optimized filter strength. The best solution is to avoid a filter at all.

3. The determination of LER from noisy CD-SEM images *without using image filtering* is made possible by means of profile-based edge detection. We acquire the SEM signal profile by integrating a SEM image of line edges in the direction of the edges. The profile is then matched against the raw (unfiltered) amplitude of edges in the image.

4. We have shown in a simulation study that LER can still be determined from very noisy images with only 2 electrons per pixel on average ($\approx 10\,\mu C/cm^2$): a single image with 12 line edges, produces an estimation of the LER with a relative error of only 10%.

5. We observed that the determination of LER does not significantly improve anymore beyond an image dose of 20 electrons per pixel. There is no point in using higher imaging dose.

6. We have observed in an experimental study that most of the resist shrinkage happens in the very first few frames of the CD-SEM image acquisition and is non-linear of nature.. Furthermore, we have observed that LER decreases for accumulating dose during imaging.

7. We have succeeded in creating a rigorous Monte-Carlo simulation program for CD-SEM images, which runs entirely on a GPU. The GPU simulator is verified by making comparisons to experimental secondary/backscatter yields and EELS spectra.

8. We have determined that the GTX480 from NVIDIA generates CD-SEM images (depending on the primary electron energy) 387 to 894 times faster than a single threaded Intel X5650 CPU.

9. We have observed an averaging of the 3D side wall roughness in the 2D determination of the LER. The averaging is non-linear with respect to the correlation length ξ_{3D} of the SWR. The effect is reduced by increasing the kinetic energy of the primary beam.

10. We have observed that determination of LER from a CD-SEM image has a strong dependence on (1) the kinetic energy of the beam, (2) feature height and (3) the frequency dependence of the SWR.

11. We conclude from a sensitivity analysis that acoustic phonon scattering plays a significant role in the calculation of SEYs. Although acoustic phonon scattering has a strong coupling to low energetic electrons, its influence on the SEY of pure silicon is seen all the way up to primary electron energy of 10 keV.

12. The scaling of the scattering cross-sections affects the contrast of the CD-SEM image, but not the roughness characterization of the lines, i.e. the 3σ of the LER, correlation length and roughness exponent. This means that there is no need to increase the accuracy of the scattering cross-sections. SEM image simulation programs could perhaps be simplified, because the exact cross-sections are not that important.

13. We have succeeded in creating a fast three-dimensional Monte-Carlo framework for the investigation of shotnoise induced SWR formation. In comparison with existing studies, we can simulate the electron-resist interaction by using the GPU accelerated Monte-Carlo program.

14. In a simulation study of SWR formation, we have observed shot-noise effects as we decrease the exposure dose. An increased effect of shotnoise is observed near the vacuum and substrate interface. One contribution relates to the actual number of acids, which due to the scattering is less near the interface than away from the interface. Another contribution stems from the fact that no acids are found on the vacuum side nor on the substrate side.

Curriculum vitae

Personal data:

Date of birth	:	31st of October 1980
Place of birth	:	Rotterdam, the Netherlands
Gender	:	Male
Nationality	:	Dutch
Languages	:	Dutch and English
Driver license	:	Yes (B)
Marital Status	:	Married

Education:

- Master's study at Utrecht University (2009-2013) in Theoretical Physics. Master thesis: "*Computational Charged Particle Optics – Symplectic relativistic ray-tracing of charged particles in electrostatic fields using vector processors*" (final grade 9).

- Bachelor's study at Delft University of Technology (2006-2009) in Applied Physics. Received Bachelor of Science with distinction (Cum Laude). Bachelor thesis: "*On the blind determination of the properties of a ZrO/W electron-emitter*" (final grade 9).

Awards:

- 2014: Karel Urbanek best student award for the paper "*Determination of line edge roughness in low dose top-down scanning electron microscopy images*", which was presented at the SPIE Advanced Lithography conference in San Jose. The article was also published in the BACUS newsletter (volume 30, issue 5, May 2014).

Grants:

- 2016: Pathfinder project titled "*Fast open source simulator of low-energy scattering of charged particles In matter*" funded by the eScience Center in Amsterdam, The Netherlands. The proposal was submitted with six support letters from companies and universities: Dr. F. de Jong (Director of FEI company), Dr. W. van der Zande (Director of research of ASML), Dr. H.R. Poolman (CEO of Amsterdam Scientific Instruments), Dr. M. Asai (Spokesperson of the international Geant4 collaboration) from SLAC National Accelerator Laboratory, Dr. Sébastien Incerti (University of Bordeaux, France) and Prof.Dr. Maria-Roser Valentí (Goethe University, Germany).

List of publications:

- Verduin, T., Cook B., and Kruit P., "*The Influence of Gun Design on Coulomb-interactions in a Field Emission Gun.*" Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 29, 06F605 (2011)

- Verduin, T., Kruit, P., and Hagen, C. W., "*Determination of line edge roughness in low dose top-down scanning electron microscopy images*," Proceedings of SPIE 9424, Metrology, Inspection, and Process Control for Microlithography, 90500L (2014)

- Verduin, T., Kruit, P., and Hagen, C. W., "*Determination of line edge roughness in low dose top-down scanning electron microscopy images*," Journal of Micro/Nanolithography, MEMS, and MOEMS, 13, 033009 (2014)

- Verduin, T., Lokhorst, S.R., Kruit, P., and Hagen, C.W., "*The effect of sidewall roughness on line edge roughness in top-down scanning electron microscopy images*," Proceedings of SPIE 9424, Metrology, Inspection, and Process Control for Microlithography, 942405 (2015)

- Verduin, T., Lokhorst, S.R., Hagen, C.W., and Kruit, P., "*Sensitivity of secondary electron yields and SEM images to scattering parameters in MC simulations*," Microelectronic Engineering 155, 114-117 (2016)

- Verduin, T., Lokhorst, S.R., Kruit, P., and Hagen, C.W., "*GPU accelerated Monte-Carlo simulation of SEM images for metrology*," Proceedings of SPIE 9778, Metrology, Inspection, and Process Control for Microlithography, 97780D (2016)

- Verduin, T., Lokhorst, S.R., Hagen, C.W., Kruit, P., "*Simulation of shotnoise induced side-wall roughness in electron lithography*," Proceedings of SPIE 9778, Metrology, Inspection, and Process Control for Microlithography, 97781Z (2016)

List of co-authored publications:

- Cook B., Verduin T., Hagen C.W., and Kruit P., "*Brightness limitations of cold field emitters caused by Coulomb interactions*," Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 28, C6C74 (2010)

- Cook B., Verduin T., and Kruit P., "*Statistical Coulomb interactions in cold field emitters*," 24th International Vacuum Nanoelectronics Conference (2011)

- Zonnevylle A.C., Verduin T., and Kruit P., "*Deflection properties of an electrostatic electron lens with a shifted electrode*," Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 31, 06F702 (2013)

- Hari S., Hagen C.W., Verduin T., and Kruit P., "*Size and shape control of sub-20 nm patterns fabricated using focused electron beam-induced processing*," Journal of Micro/Nanolithography, MEMS, and MOEMS, 13, 033002-033002 (2014)

Bibliography

[1] van Steenwinckel, D., "Lithographic importance of acid diffusion in chemically amplified resists," *Proceedings of SPIE*, **5753**, 269–280, SPIE (2005).

[2] Gallatin, G. M., "Continuum model of shot noise and line edge roughness," in *Proceedings of SPIE*, **4404**, 123–132, SPIE (2001).

[3] Yuan, L. and Neureuther, A. R., "Investigation of shot-noise-induced line-edge roughness by continuous-model-based simulation," *Proceedings of SPIE*, 312–321 (2004).

[4] Kruit, P., Steenbrink, S., Jager, R., and Wieland, M., "Optimum dose for shot noise limited CD uniformity in electron-beam lithography," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **22**(6), 2948 (2004).

[5] Kruit, P. and Steenbrink, S., "Local critical dimension variation from shot-noise related line edge roughness," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **23**(6), 3033 (2005).

[6] Kruit, P., Steenbrink, S., and Wieland, M., "Predicted effect of shot noise on contact hole dimension in e-beam lithography," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **24**(6), 2931 (2006).

[7] Neureuther, A., Pease, R., Yuan, L., Baghbani Parizi, K., Esfandyarpour, H., Poppe, W., Liddle, J., and Anderson, E., "Shot noise

models for sequential processes and the role of lateral mixing," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **24**(4), 1902 (2006).

[8] Patsis, G., Drygiannakis, D., and Raptis, I., "Detailed resist film modeling in stochastic lithography simulation for line-edge roughness quantification," *Microelectronic Engineering* **87**, 989–992 (2010).

[9] Yamaguchi, A., Tsuchiya, R., Fukuda, H., Komuro, O., Kawada, H., and Iizumi, T., "Characterization of line-edge roughness in resist patterns and estimations of its effect on device performance," *Proceedings of SPIE*, Herr, D. J., ed., **5038**, 689–698 (2003).

[10] Lee, J., Shin, J., Kim, H., Woo, S., Cho, H., Han, W., and Moon, J., "Effect of line-edge roughness (LER) and line-width roughness (LWR) on sub-100-nm device performance," *Proceedings of SPIE*, Sturtevant, J. L., ed., **5376**, 426–433 (2004).

[11] Gustin, C., Leunissen, L., Mercha, A., Decoutere, S., and Lorusso, G., "Impact of line width roughness on the matching performances of next-generation devices," *Thin Solid Films* **516**, 3690–3696 (2008).

[12] Patsis, G. P., Constantoudis, V., Tserepi, A., Gogolides, E., and Grozev, G., "Quantification of line-edge roughness of photoresists. I. A comparison between off-line and on-line analysis of top-down scanning electron microscopy images," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **21**(3), 1008–1018 (2003).

[13] Ohfuji, T., Endo, M., and Morimoto, H., "Theoretical analysis of line-edge roughness using FFT techniques," *Proceedings of SPIE*, Conley, W., ed., **3678**, 732–738 (1999).

[14] Hiraiwa, A. and Nishida, A., "Discrete power spectrum of line width roughness," *Journal of Applied Physics* **106**(7), 074905 (2009).

[15] Nishida, A., "Statistical- and image-noise effects on experimental spectrum of line-edge and line-width roughness," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **9**, 041210 (2010).

[16] Azarnouche, L., Pargon, E., Menguelti, K., Fouchier, M., Fuard, D., Gouraud, P., Verove, C., and Joubert, O., "Unbiased line width roughness measurements with critical dimension scanning electron microscopy and critical dimension atomic force microscopy," *Journal of Applied Physics* **111**, 084318 (2012).

[17] Mack, C. A., "Systematic errors in the measurement of power spectral density," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **12**, 033016 (2013).

[18] Yamaguchi, A., Steffen, R., Kawada, H., and Iizumi, T., "Bias-Free Measurement of LER/LWR with Low Damage by CD-SEM," *Proceedings of SPIE* **6152**, 61522D–61522D–8 (2006).

[19] Yamaguchi, A., Steffen, R., Yamamoto, J., Kawada, H., and Iizumi, T., "Single-shot method for bias-free LER/LWR evaluation with little damage," *Microelectronic Engineering* **84**, 1779–1782 (2007).

[20] Ohashi, T., Sekiguchi, T., Yamaguchi, A., Tanaka, J., and Kawada, H., "Precise measurement of photoresist cross-sectional shape change caused by SEM-induced shrinkage," *Proceedings of SPIE* **8681**, 86810K–86810K–12 (2013).

[21] Yamaguchi, A. and Yamamoto, J., "Influence of Image Processing on Line-Edge Roughness in CD-SEM Measurement," *Proceedings of SPIE* **6922**, 692221–692221–8 (2008).

[22] Hiraiwa, A., "Image-pixel averaging for accurate analysis of line-edge and linewidth roughness," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **10**, 023010 (2011).

[23] Constantoudis, V. and Pargon, E., "Evaluation of methods for noise-free measurement of LER/LWR using synthesized CD-SEM images,"

Proceedings of SPIE, Starikov, A. and Cain, J. P., eds., **8681**, 86812L (2013).

[24] Li, Y. G., Mao, S. F., Li, H. M., Xiao, S. M., and Ding, Z. J., "Monte Carlo simulation study of scanning electron microscopy images of rough surfaces," *Journal of Applied Physics* **104** (2008).

[25] Lawson, R. A. and Henderson, C. L., "Understanding the relationship between true and measured resist feature critical dimension and line edge roughness using a detailed scanning electron microscopy simulator," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **28**, C6H34 (2010).

[26] Palasantzas, G., "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," *Physical Review B* **48**, 472–478 (1993).

[27] Mack, C. A., "Generating random rough edges, surfaces, and volumes.," *Applied Optics* **52**, 1472–1480 (2013).

[28] Verduin, T., "The effect of sidewall roughness on line edge roughness in top-down scanning electron microscopy images," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **9424**, 942405 (2015).

[29] Lawson, R. A. and Henderson, C. L., "Three-dimensional mesoscale model for the simulation of LER in photoresists," *Proceedings of SPIE* **7639**, 76392G–76392G–10 (2010).

[30] Villarrubia, J. S. and Bunday, B. D., "Unbiased estimation of linewidth roughness," *Proceedings of SPIE*, Silver, R. M., ed., **5752**, 480–488 (2005).

[31] Hiraiwa, A. and Nishida, A., "Statistically accurate analysis of line width roughness based on discrete power spectrum," *Proceedings of SPIE*, Raymond, C. J., ed., **7638**, 76380N–76380N–12 (2010).

[32] Kruit, P. and Steenbrink, S., "Shot noise in electron-beam lithography and line-width measurements.," *Scanning* **28**(1), 20–26 (2006).

[33] Coleman, T. F. and Li, Y., "An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIAM Journal on Optimization* **6**, 418–445 (1996).

[34] Kudo, T., Bae, J.-b., Dammel, R. R., Kim, W.-k., McKenzie, D. S., Rahman, M. D., Padmanaban, M., and Ng, W., "CD changes of 193-nm resists during SEM measurement," *Processing*, Houlihan, F. M., ed., **4345**, 179 (2001).

[35] Habermas, A., Hong, D., Ross, M. F., and Livesay, W. R., "193-nm CD shrinkage under SEM: modeling the mechanism," *Proceedings of SPIE*, Herr, D. J. C., ed., **4689**, 92 (2002).

[36] Ayal, G., Andelman, D., and Cohen, Y., "Analytical model for ArF photoresist shrinkage under scanning electron microscopy inspection," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* **27**(4), 1976 (2009).

[37] Czyzewski, Z., MacCallum, D. O., Romig, A., and Joy, D. C., "Calculations of Mott scattering cross section," *Journal of Applied Physics* **68**, 3066 (1990).

[38] Jablonski, a. and Tougaard, S., "Database of relativistic elastic scattering crosssections for calculations of photoelectron and Auger electron transport," *Surface and interface analysis* **22**, 129–133 (1994).

[39] Dapor, M., "Elastic scattering calculations for electrons and positrons in solid targets," *Journal of Applied Physics* **79**(1996), 8406–8411 (1996).

[40] Salvat, F., Jablonski, A., and Powell, C. J., "Elsepa - Dirac partial-wave calculation of elastic scattering of electrons and positrons by atoms, positive ions and molecules," *Computer Physics Communications* **165**, 157–190 (2005).

[41] Shockley, W., *Electrons and Holes in Semiconductors*, Princeton (1950).

[42] Salvat-Pujol, F., *Secondary-electron emission from solids : Coincidence experiments and dielectric formalism*, PhD thesis (2012).

[43] Ashley, J., "Interaction of low-energy electrons with condensed matter: stopping powers and inelastic mean free paths from optical data," *Journal of Electron Spectroscopy and Related Phenomena* **46**(1), 199–214 (1988).

[44] Kieft, E. and Bosch, E., "Refinement of Monte Carlo simulations of electronspecimen interaction in low-voltage SEM," *Journal of Physics D: Applied Physics* **41**(21), 215310 (2008).

[45] Werner, W. S. M., Glantschnig, K., and Ambrosch-Draxl, C., "Optical Constants and Inelastic Electron-Scattering Data for 17 Elemental Metals," *Journal of Physical and Chemical Reference Data* **38**(4), 1013 (2009).

[46] Lucarini, V., Peiponen, K.-E., Saarinen, J. J., and Vartiainen, E., *Kramers-Kronig Relations in Optical Materials Research*, vol. 110 of *Springer Series in Optical Sciences*, Springer-Verlag, Berlin/Heidelberg (2005).

[47] Werner, W. S. M., Went, M. R., Vos, M., Glantschnig, K., and Ambrosch-Draxl, C., "Measurement and density functional calculations of optical constants of Ag and Au from infrared to vacuum ultraviolet wavelengths," *Physical Review B* **77**, 161404 (2008).

[48] Cullen, D., Perkins, S., and Seltzer, S., "Tables and Graphs of Electron Interaction Cross 10 eV to 100 GeV Derived from the LLNL Evaluated Electron Data Library (EEDL), Z = 1 - 100," *Lawrence Livermore National Laboratory Evaluated Electron Data Library UCRL-50400* **31** (1991).

[49] Llovet, X., Powell, C. J., Salvat, F., and Jablonski, A., "Cross Sections for Inner-Shell Ionization by Electron Impact," *Journal of Physical and Chemical Reference Data* **43**, 013102 (2014).

[50] Kuhr, J.-C. and Fitting, H.-J., "Monte Carlo simulation of electron emission from solids," *Journal of Electron Spectroscopy and Related Phenomena* **105**(2-3), 257–273 (1999).

[51] Egerton, R. F., "Electron energy-loss spectroscopy in the TEM," *Reports on Progress in Physics* **72**, 016502 (2009).

[52] Da, B., Shinotsuka, H., Yoshikawa, H., Ding, Z. J., and Tanuma, S., "Extended mermin method for calculating the electron inelastic mean free path," *Physical Review Letters* **113**(6), 1–5 (2014).

[53] Tanuma, S., Powell, C. J., and Penn, D. R., "Use of sum rules on the energy-loss function for the evaluation of experimental optical data," *Journal of Electron Spectroscopy and Related Phenomena* **62**(1-2), 95–109 (1993).

[54] Kuhr, J.-C. and Fitting, H.-J., "Monte-Carlo Simulation of Low Energy Electron Scattering in Solids," *Physica Status Solidi (a)* **172**(2), 433–450 (1999).

[55] Palik, E., *Handbook of Optical Constants of Solids Vol. 1*, Academic Press (1985).

[56] Palik, E., *Handbook of Optical Constants of Solids Vol. 2*, Academic Press (1991).

[57] Palik, E., *Handbook of Optical Constants of Solids Vol. 3*, Academic Press (1998).

[58] Henke, B., Gullikson, E., and Davis, J., "X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at E = 50-30,000 eV, Z = 1-92," *Atomic Data and Nuclear Data Tables* **54**, 181–342 (1993).

[59] Polyanskiy, M., "Refractive index database," (2016).

[60] Penn, D. R., "Electron mean-free-path calculations using a model dielectric function," *Physical Review B* **35**(2), 482–486 (1987).

[61] Bethe, H., "Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie," *Annalen der Physik* **397**(3), 325–400 (1930).

[62] Bethe, H., "Bremsformel fr Elektronen relativistischer Geschwindigkeit," *Zeitschrift fr Physik* **76**, 293–299 (1932).

[63] Sternheimer, R. M., "The Density Effect for the Ionization Loss in Various Materials," *Physical Review* **88**, 851–859 (1952).

[64] Sternheimer, R. M., Seltzer, S. M., and Berger, M. J., "Erratum: Density effect for the ionization loss of charged particles in various substances," *Physical Review B* **27**, 6971–6971 (1983).

[65] Bull, R., "Stopping powers for electrons and positrons," (1986).

[66] Herman, F., Kortum, R. L., and Kuglin, C. D., "Energy band structure of diamond, cubic silicon carbide, silicon, and germanium," *International Journal of Quantum Chemistry* **1**(S1), 533–566 (1967).

[67] Fitting, H. J., Schreiber, E., Kuhr, J. C., and Von Czarnowski, A., "Attenuation and escape depths of low-energy electron emission," *Journal of Electron Spectroscopy and Related Phenomena* **119**(1), 35–47 (2001).

[68] Schreiber, E. and Fitting, H. J., "Monte Carlo simulation of secondary electron emission from the insulator SiO2," *Journal of Electron Spectroscopy and Related Phenomena* **124**(1), 25–37 (2002).

[69] Pop, E., Dutton, R. W., and Goodson, K. E., "Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion," *Journal of Applied Physics* **96**(9), 4998 (2004).

[70] Dolling, G., *Inelastic scattering of neutrons in solids and liquids*, vol. 44 (1963).

[71] Sparks, M., Mills, D. L., Warren, R., Holstein, T., Maradudin, A. A., Sham, L. J., Loh, E., and King, D. F., "Theory of electron-avalanche breakdown in solids," *Physical Review B* **24**, 3519–3536 (1981).

[72] Bradford, J. N. and Woolf, S., "Electron-acoustic phonon scattering in SiO2 determined from a pseudo-potential for energies of E>EBZ," *Journal of Applied Physics* **70**(1), 490–492 (1991).

[73] Green, M. A., "Intrinsic concentration, effective densities of states, and effective mass in silicon," *Journal of Applied Physics* **67**(6), 2944 (1990).

[74] Walker, C. B., "X-Ray Study of Lattice Vibrations in Aluminum," *Physical Review* **103**, 547–557 (1956).

[75] Lynn, J. W., Smith, H. G., and Nicklow, R. M., "Lattice Dynamics of Gold," *Physical Review B* **8**, 3493–3499 (1973).

[76] Tan, G., DeNoyer, L., French, R., Guittet, M., and Gautier-Soyer, M., "KramersKronig transform for the surface energy loss function," *Journal of Electron Spectroscopy and Related Phenomena* **142**, 97–103 (2005).

[77] Ohta, K. and Ishida, H., "Comparison Among Several Numerical Integration Methods for Kramers-Kronig Transformation," *Applied Spectroscopy* **42**, 952–957 (1988).

[78] Verduin, T., Lokhorst, S. R., Kruit, P., and Hagen, C. W., "The effect of sidewall roughness on line edge roughness in top-down scanning electron microscopy images," *Proceedings of SPIE*, Cain, J. P. and Sanchez, M. I., eds., **9424**, 942405 (2015).

[79] Villarrubia, J. S., Vladár, A. E., and Postek, M. T., "3D Monte Carlo modeling of the SEM: Are there applications to photomask metrology?," **9236**, 923602 (2014).

[80] Verduin, T., "The effect of adaptive filtering and edge rotation on line edge roughness determination," tech. rep., Delft University of Technology (2014).

[81] Owens, J., Houston, M., Luebke, D., Green, S., Stone, J., and Phillips, J., "GPU Computing," *Proceedings of the IEEE* **96**, 879–899 (2008).

[82] Barnes, J. E. and Hut, P., "A hierarchical O(N log N) force-calculation algorithm," *Nature* **324**, 446–449 (1986).

[83] Akenine-Möller, T., "Fast 3D Triangle-Box Overlap Testing," *Journal of Graphics Tools* **6**, 29–33 (2001).

[84] Williams, A., Barrus, S., Morley, R. K., and Shirley, P., "An efficient and robust ray-box intersection algorithm," *ACM SIGGRAPH 2005*, **10**(1), 9, ACM Press, New York, New York, USA (2005).

[85] Merril, D. and Grimshaw, A., "High performance and scalable radix sorting: a case study of implementing dynamic parallelism for GPU computing," *Parallel Processing Letters* **21**, 245–272 (2011).

[86] Joy, D. C., "A database on electron-solid interactions," *Scanning* **17**, 270–275 (2006).

[87] Raftari, B., Budko, N. V., and Vuik, C., "Self-consistent drift-diffusion-reaction model for the electron beam interaction with dielectric samples," *Journal of Applied Physics* **118**, 204101 (2015).

[88] Dapor, M., Ciappa, M., and Wolfang, F., "Monte Carlo modeling in the low-energy domain of the secondary electron emission of polymethylmethacrylate for critical-dimension scanning electron microscopy," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **9**, 023001 (2010).

[89] Ganachaud, J. and Mokrani, A., "Theoretical study of the secondary electron emission of insulating targets," *Surface Science* **334**, 329–341 (1995).

[90] Batson, P. E., "Current trends for EELS studies in physics," *Microscopy Microanalysis Microstructures* **2**(2-3), 395–402 (1991).

[91] Cheynet, M. C. and Epicier, T., "Structural and chemical analysis of a model SiSiO 2 interface using spatially resolved electron-energy-loss spectroscopy," *Philosophical Magazine* **84**, 1753–1771 (2004).

[92] Ewels, P., Sikora, T., Serin, V., Ewels, C. P., and Lajaunie, L., "A Complete Overhaul of the Electron Energy-Loss Spectroscopy and X-Ray Absorption Spectroscopy Database: eelsdb.eu," *Microscopy and Microanalysis* **22**, 717–724 (2016).

[93] Verduin, T., Lokhorst, S., Hagen, C., and Kruit, P., "Sensitivity of secondary electron yields and SEM images to scattering parameters in MC simulations," *Microelectronic Engineering* **155**, 114–117 (2016).