Master Thesis

University of Twente

Characterization of segregation in nanometer thin films using hybrid X-ray measurements

Development of Bayesian inference sampling using Hamiltonian Monte Carlo methods for the analysis of the solutions posterior of the combined X-ray reflectivity and X-ray standing wave measurement thin film reconstructions

Cedric Pascal Hendrikx *

October 11, 2020

* Supervised by dr. I Makhotkin & dr. M. Schlottbom

Master Thesis

Disclaimer
This work is done for the completion of the Master of Science of the double program in Applied Physics & Aplied Mathematics at the University of Twente. This research was performed in a collaboration between the Industrial Focus Group XUV Optics & the Mathematics of Computational Science Group at the Univerity of Twente.

Acknowledgements

During my bachelor thesis research I came in touch with dr. Igor Makhotkin who was very motivated to start the development of an XSW laboratory setup in the XUV laboratory. Since my bachelor I have had a part-time position at XUV optics and a role in the development of this setup. The basis of this research stems of my personal interest to the X-ray Standing wave (XSW) method that I have helped implement and improve in the laboratory with my colleagues at the XUV Optics group. Under the supervision of dr. I. Makhotkin I was given the chance to not only help colleges at the XUV optics group to assist in measurements and analysis on their research but also to pursure my own masters research at the XUV optics group. For this I would like extend my gratitude dr. I. Makhotkin for the supervision over the last couple of years which greatly helped in my development, especially over last year during my master thesis.

Under the supervision of dr. Matthias Schlottbom I was able to extend this research to include a mathematical pursuit for the completion of a master in Applied Mathermatics. I would like to extend my sincere thanks to dr. Matthias Schlottbom for the supervision and consultations during this time. Most contact went over video calls due to the coronavirus, but this did not stop dr. Matthias Schlottbom from providing me with the proper supervision and consultation.

I would like to thank ir. Theo van Oijen for the training and assistance with the ADC coater in the XUV depositions laboratory. Dr. Andrey Zameshin for help with free-form analysis and annealing setup. I am grateful to dr. S.N. Yakunin for providing the code for free-form XRR-XSW analysis that was extensively used in this work.

I am also grateful to dr. Maxim Korytov from IMEC for providing me with TEM measurements of my thin films that were used to verify the validity of the reconstruction framework.

In addition, I would like to thank the external members dr. Julio Backhoff and prof.dr.ir. Mark Huijben of the graduation committee for taking the time to read and my thesis.

Furthermore I acknowledge the support of the Industrial Focus Group XUV Optics at the MESA+ Institute for Nanotechnology at the University of Twente, lead by prof.dr.ir Fred Bijkerk, with industrial partners ASML, Carl Zeiss SMT, and Malvern Panalytical for providing me with a position to finish my master in Applied Physics. In addition I would like to thank the Mathematics of Computational Sciences group at the University of Twente, lead by prof.dr.ir. Jaap van der Vecht, for providing me with a position to finish my masters in Applied Mathematics.

Cedric Hendrikx

Contents

1 Introduction

Advanced thin film applications require precise control over the thin film structure, composition and the interfaces. The segregation of solute atoms to the surface and interfaces of a functional layer can affect the multilayer's functional properties. Therefore understanding and, ultimately, control over the segregation process is highly desirable for the development of such structures and its applications. One can imagine that the segregation can be used to passivate interfaces in thin films and 3D nano-structures, by analogy with e.g. metallurgy where the grain-boundary segregation causes passivation of a grain boundary and prevents formation of large crystalline grains.

Until now these segregation effects in nano-meter thin films to buried interfaces have not been the subject of extensive research and the details of these processes remain relatively unknown. In this thesis a research on segregation to buried interfaces of thin film structures is presented. We have made a first step in understanding which material parameters are important for the segregation process in thin films under annealing conditions. This is done by iterating different transition metal combinations in a bi-layer thin film system and studying the behavior of the different material combinations. The material parameters that are of interest in this research are the atomic radius, crystal structure and the surface energy. A schematic example of the thin film systems and the segregation that is the subject of this research is shown in Figure 1.1. The thin film is composed of 2 layers, respectively a bottom layer composed of a solute element and a matrix element and top layer consisting of only one material. Annealing is used to stimulate the segregation process by bringing the system to more mobile state. In this configuration the solute atoms are able, either to remain in the same layer, segregate on one ore more interfaces or diffuse through the entire stack. The final state of the bi-layer after annealing will be cumulative representation of solute mobility in the system and thermodynamics balance.

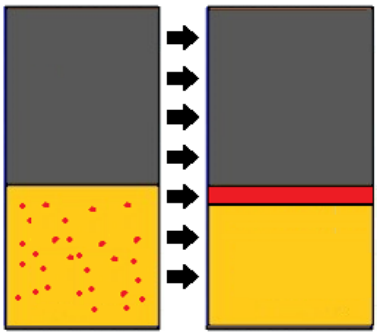


Figure 1.1: Example of the to be researched interface segregation.

A combination of grazing incidence X-ray reflectively (GIXR) measurements with X-ray standing wave fluorescence (XSW-XRF) measurements is used to reconstruct the thin film structures and its atomic depth distributions. This reconstruction is employed to map the changes in atomic distributions before and after the annealing and with sub-nanometer sensitivity, enough for the quantification of potential interface segregation.

To confirm the validity of combined GIXR and XSW structure reconstructions, a Bayesian inference is applied to obtain a stringent confidence bounds on the individual reconstruction parameters. To obtain the posterior distribution used in the Bayesian inference Hamiltonian Monte Carlo methods are used and compared in performance to conventional MCMC methods to test its efficacy. Additionally a transmission electron microscopy (TEM) study is done on 2 samples to verify the accuracy of the thin film reconstructions experimentally.

For predicting grain boundary segregation, a semi-empirical thermodynamical model was proposed for metal-metal systems in the pioneering work of Miedema et al.[1] Here the surface energy and compound formation enthalpy are considered as critical parameters that determine segregation effects in bulk layers. A similar model depending on surface energies and atomic radii was applied for predicting surface segregation on top of solid alloys [2]. This model was later extended in [3] for the analysis of grain boundary segregation in poly-crystalline alloys. Most of recent research into interfacial segregation was

in the field of grain boundaries [4] [5]. Some research has been done for layered systems [6]. A model that predicts formation enthalpies of segregation in thin solid metal films has been proposed by [7] and will be tested for applicability in the solid layer nanometer thin film regime.

To generate an X-ray standing wave in a thin film structure, two oppositely travelling waves of comparable intensity should interfere. This condition can be fulfilled, for example, if the incidence beam and reflected are of comparable intensity. For the generation of standing waves previously 2 methods have been used with success, the waveguide method and the use of a periodic multilayer. The waveguide method has been tried with success in a synchrotron environment [8] but has not yet been successfully deployed in a thin-film laboratory. The multilayer method has already been used in a similar environment [9] but has too many limitations for the current research.

Currently most reconstructions with a large number of degrees of freedom assume a single solution from a data set of measurements [10]. Previously Bayesian inference methods have been applied to GIXR, XSW and related X-ray measurement data to obtain a solutions sets of possible solutions that fit the data, parametrized by 8 degrees of freedom [11]. The usage of multiple data sets in this analysis has shown to reduce the number of solutions that fit to the measured data and therefore decreased the in the width of the confidence interval on individual parameters. With parametrizations that allow for significantly more degrees of freedom, e.g. 50-70, a classical Metropolis-Hastings MCMC (Monte-Carlo Markov Chain) may not show convergence within acceptable time due to the quadratic time scaling per degree freedom [12]. A different promising sample method, the Hamiltonian Monte Carlo method has previously been successfully implemented in Bayesian inference applications on high-dimensional parameter spaces [13] and in optimal situations scales in the power of $\frac{5}{4}$ in time with the number of degrees of freedom [12] and therefore shows potential in the applicability in reconstruction of thin film structures using a free form parametrization.

Research steps

Below the research steps I made are listed, in each step the related education track, namely applied physics (AP) or applied mathematics (AM), is indicated.

- The selection of a set of transition metals for which the material parameters can be isolated to understand their role in interface segregation. (AP)
- Designing a thin film structure that is stable during the annealing and is suitable for XRR-XSW characterization with optimal sensitivity to the segregation process. (AP)
- The adaptation of the free-form parametrization approach [10] for automatic analysis of GIXR-XSW data specifically optimized for the segregation study in thin films. (AP)
- Adaptation of the Hamiltonian Monte Carlo method to the analysis of hybrid X-ray measurements data. (AM)
- Obtaining an accurate approximation of the posterior of reconstructions of GIXR and XSW measurement data to obtain the confidence intervals of the individual reconstruction parameters. (AM)

Methodology

Material Selection A material selection of individual transition metals is made to efficiently investigate the possible parameters important in the quantification of the segregation process. The selection is made in a way that the 3 parameters of investigation, Atomic Radius, Crystal Structure and Surface Energy, are isolated to study the influence of the individual material parameters on the segregation.

Structure Design Thicknesses of the waveguide layers and the to be studied bi-layer system in the thin film structure will be optimized in terms of discerning power to differentiate the fluorescence radiation coming from the different interfaces. The waveguide layers are made of Tungsten due to its high optical contrast and its relatively high $\frac{\delta}{\beta}$ optical constant ratio for the Cu-Kα wavelength.

The optimized sample structure is a bi-layer thinfilm system of pure transitions metals of which the bottom layer has been enriched with a dopant material that is the subject of the segregation research. The layers thicknesses are of the order of $\sim 15\,\text{nm}$ for the top layer and $\sim 10\text{nm}$ for the bottom layer. This bi-layer system deposited between a waveguide consisting of a thick bottom waveguide layer ($\sim$40nm) and thin top waveguide layer($\sim$5nm). The design of the structure with the designated labels for the layers and interfaces that will be used trough out this thesis is shown in Figure 1.2. To verify the practical applicability of the layer optimizations, three samples have been deposited for which the dopant material is individually pre-depositioned to the three expected segregation sites (T,M,B Figure 1.2) successfully verifying the discerning power of the waveguide system in practice.

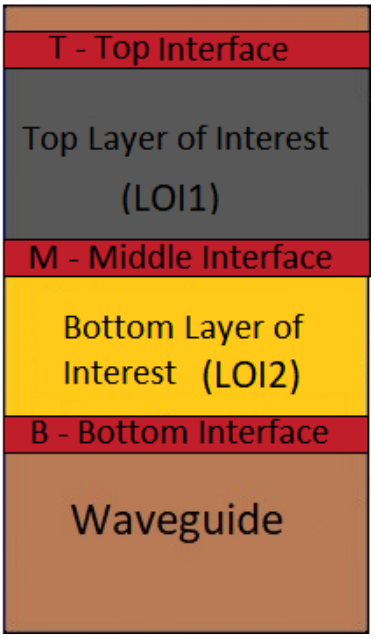


Figure 1.2: Example of the to be measured structure.

Multiple bi-layer material combinations have been deposited with waveguide to study the behavior of the interface segregation under annealing conditions for the different transition metal combinations. For 2 thin films showing significant segregation a Transmission Electron Miscroscopy (TEM) study was done, verifying the results obtained from the GIXR-XSW measurement reconstructions.

Reconstruction Previously sample reconstructions from GIXR and XSW measurements have been done by formulating an inverse problem with a forward map calculating the resulting GIXR and XSW signals from a proposed parametrized structure. A parameter estimation of a set of structure parameters is done by optimizing a forward map with a general optimization method to find the parametrization of the reconstruction that best satisfies the measured GIXR and XSW signals. This method infers a single solution to the data. However due to the ill-posedness of this inverse problem, even with the help of regularization methods these sample reconstructions have not always shown to be unique [10]. The uncertainty in the reconstruction parameters is seen as one of the main criticisms of the usage of the independent reconstructions from the GIXR and XSW measurements.

To address this ill-posedness, A Bayesian inference is used to obtain the confidence intervals on the parameters of the reconstruction methodology. The sample space of the possible parametrizations of the reconstructions that fit the GIXR and XSW measurements is obtained by exploring the parameter space using different Monte-Carlo Markov Chain (MCMC) methods. Due to the high-dimensional nature of the problem (>50 parameters) a conventional Metropolis-Hastings with a transition kernel based on a probability distribution can in practice lead to a low convergence rate. Therefore Hamiltonian

Monte Carlo (HMC) methods are implemented, which are more suitable for Bayesian inference applications in higher dimensions since contrary to the conventional Metropolis-Hastings the convergence rate is less dependent on the number of dimensions in the problem. To confirm the performance of the HMC implementation, the HMC implementation is compared to a conventional Metropolis-Hastings implementation and has shown significantly faster convergence.

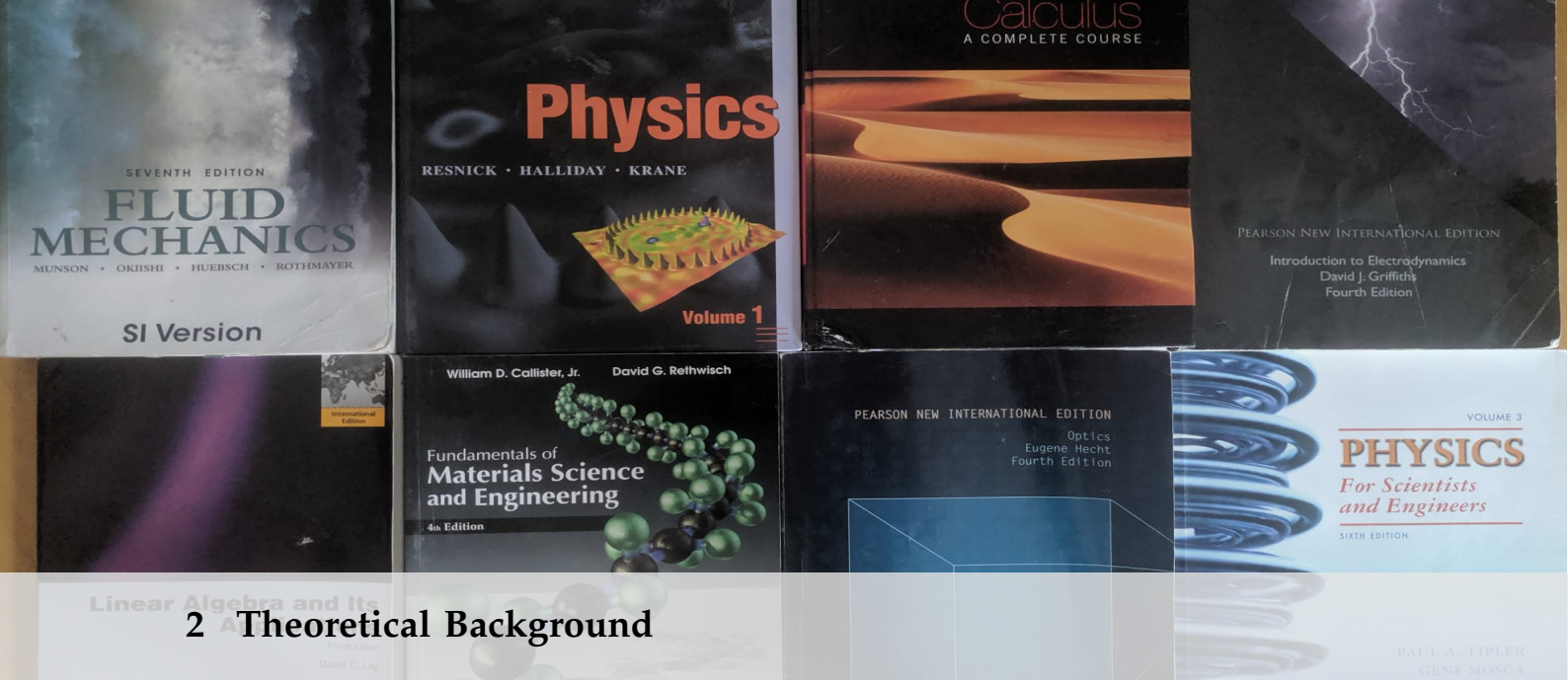
Thesis Overview

In Chapter 2 the theoretical framework is introduced that is necessary to understand the upcoming chapters. First the physical aspects of the segregation are introduced. Following an introduction into the metrological techniques that are used. The chapter follows with a introduction in Bayesian inference, inverse problems and the used MCMC methods. In the end a small example of a MCMC is given to further help the reader understand the Bayesian inference and MCMC methodology.

In Chapter 3 the experimental aspects are discussed. First the technical details of the sample selection and sample manufacturing are discussed. The chapter end with a description of the experimental and data acquisition part of the GIXR and XSW measurements.

In Chapter 4 first the free-form structure parametrization is introduced following the definitions in [10] and the calculations used to calculate a the resulting GIXR and XSW signal from said parametrization. Using these descriptions the forward map of the inverse problem is defined that calculates a loss function based on the mismatch of the resulting GIXR and XSW signals form a proposed structure paramterization and the actual measured GIXR and XSW measurements. Afterwards the Bayesian inference framework is introduced and the details of the MCMC implementations are discussed. The Chapter ends with the implementation of the derivatives that are used in the HMC algorithm.

In Chapter 5 The results are presented. In section 5.1 the thermal stability of the waveguide is presented, in Section 5.2 the confirmation of the sensitivity to the different interfaces using the pre-depositioned test samples is presented. In section 5.3 the stability and efficacy of the HMC applied to the reconstruction methodology is presented. In section 5.4 the reconstruction validity is shown using a TEM microscopy study. In Section 5.5 athe efficacy of the addition of the XSW measurements is presented. In 5.6 the performance of the different MCMC methods is discussed and in Section 5.7 all analyzed transition metal combinations, the segregation enthalpy, the depence on the different atomic parameters and the discovered threshold behavior following the predictions in [7] is presented.



2 Theoretical Background

This chapter contains the theoretical background required to understand the research presented in this thesis. First the physical aspects are discussed, the second part of the chapter is dedicated to the mathematical framework necessary to understand the details of the reconstruction methodology.

2.1 Segregation

Surface Segregation Surface segregation is the enrichment of a surface by an element that segregates from the inside the structure to the surface. This happens when it is more energetically favourable for a certain element to be on the surface due to differences in elemental surface energies. Normally segregation does not occur at ambient temperatures but structures have to be exposed to annealing conditions for the atoms in the structure to achieve mobility to reach an equilibrium state.

Interfacial Segregation Interfacial segregation is a physical process where interfaces between thin-layers or grain boundaries become enriched with a certain element due to equilibrium processes striving for a more energetically favourable state. The behavior of this segregation process in thin films is not well understood and not a lot of literature is yet available on this topic. These processes generally do not occur or occur very slowly at room temperature, therefore an annealing process is often used to achieve this equilibrium.

With surface segregation certain parameters of the materials (i.e. surface energy per m^2) can be used to formulate a model to predict the tendency of a certain compound material to exhibit surface segregation. This has been done in the past with an coincidence of 80% [14]. Previous models based on the interface segregation have been proposed, one of them by Gerkema and Miedema [7] who has made, with success, empirical models predicting mixing and alloying of metals and surface segregation. Their prediction on interface segregation is shown in Equation 2.1 but has not been tested in the thin film regime. The model calculates the difference in total enthalpy in the system when an atom A from matrix B moves to a neighbouring interface between the layer made of matrix element B and some neighbouring layer of element C. The model parameters are based on alloying enthalpies of bi-metal alloys and the individual surface energies of the metals. The metal alloying enthalpies

Equation 2.1

Expression Miedema's proposed interface segregation of sergeant of atom type A present in a matrix of atom type B to the interface of a directly neighbouring a layer of atom type C:

$$\Delta_{B|C}^{segrA} = \Delta H_1^{segr} + \Delta H_2^{segr} + \Delta H_3^{segr} \tag{2.1}$$

Where H_1^{segr} accounts for the change in enthalpy caused by the change in neighbours for segrgant atom A and is given by:

$$\Delta H_1^{segr} = -\frac{1}{3} H_{AinB}^{sol}$$

Where $-H_{AinB}^{sol}$ is the mixing enthalpy of metal A in B as in [1], which is the energy difference in joule per unit of mass to when creating the compound alloy A-B from the pure metals A and B.

H_2^{segr} and H_3^{segr} account for the change in enthalpy due to replacing atoms of material B with material A in the interface between layer B and C and are respectively given by:

$$\Delta H_2^{segr} = 1.33 \cdot 10^{-8} (\gamma_{CA}^{chem} - \gamma_{CB}^{chem}) V_A^{\frac{2}{3}}$$

$$\Delta H_3^{segr} = 1.33 \cdot 10^{-8} \cdot 0.15 (\gamma_A^0 - \gamma_B^0) V_A^{\frac{2}{3}}$$

Where $\gamma_{AB}^{chem} = 2.5 \cdot 10^{-9} H_{AinB}^{sol} / V_A^{\frac{2}{3}}$ and γ_M^0 is the surface energy of solid M at $T = 0°K$

are based on an empirical model containing electron densities, atomic radii and surface energies. When the total enthalpy is negative, segregation is expected to be energetically favourable and predicted to occur when mobility is achieved.

2.2 Metrology

In this section the 2 measurement methods, the Grazing Incendence X-ray Reflectivity and X-ray Standing Wave Fluorescence technique, that are in combination used to reconstruct the thin film structures of interest are discussed.

Grazing Incendence X-ray Reflectivity (GIXR)

GIXR is a metrology method used for obtaining information on the index of refraction of a structure in the depth direction. This method is frequently used to characterize thin film samples uniform in the lateral directions. A beam of X-rays of a single wavelength is directed at a sample and the specular reflection at different angles is measured. The reflected intensity is used to characterize certain properties of the sample ranging from layer and interface thicknesses, densities. Even a complex reconstruction of the index of refraction profile in the depth direction can be obtained from the a GIXR curve.

Index of refraction For light in the X-ray spectrum, a complex number is used to describe the index of refraction, its expression is shown in Equation 2.2. With a transition from air or vacuum to a more dense material the index of refraction generally goes down for wavelengths in the X-ray spectrum.

Equation 2.2

Expression of the index of refraction often used in the X-ray regime:

$$n = 1 - \delta - i\beta \tag{2.2}$$

with real positive parameters δ (dispersion description) and β (absorption description) and i denoting the complex unit.

Scattering Vector In the GIXR curve, each angle corresponds to a length that is studied and the reflected intensity of the specular reflection is proportional to the absolute value of the Fourier component of that respective length in the perpendicular direction in the optical density profile of the sample. This length is called the scattering vector and its expression is shown in Equation 2.3.

Equation 2.3

Scattering vector Q with quantity inverse length:

$$Q = \frac{4\pi \sin(\theta)}{\lambda} \tag{2.3}$$

The GIXR measurements perceive the inverse space in the z direction. The resolution is limited by the largest scattering vector that is measured, therefore the resolution is determined by the largest angle that is measured. From the scattering vector a resolution criteria can be calculated that signifies the smallest length that a measurement is sensitive to which is described by equation 2.4 [10]. This resolution criteria (Equation 2.4) will be used throughout the research since it is the optimal discretization length in modelling thin film structures to account for computational efficiency while being able to accurately describe all aspects of the measurement data.

Equation 2.4

Relation of the angular range that is measured to spatially limited resolution that can be perceived:

$$D_{min} = \frac{\lambda}{4 \sin(\theta_{max})} \tag{2.4}$$

Missing Phase Information The GIXR measurements are done with a classical X-ray tube and is a non-coherent scattering method which only measures the amplitude of the reflection. Therefore obtaining the phase information during these measurements is not possible. This is known as the missing phase information problem which is a general weakness of the GIXR method [10]. A direct reconstruction of a measurement from the GIXR method is therefore not possible and generally only basic information on the thicknesses and densities are obtained from a GIXR curve.

X-ray Standing Wave Fluorescence (XSW-XRF)

The XSW-XRF technique is used to obtain an atomic element-wise sensitive location dependent signal from a structure. To form a standing wave in the structure, a strong reflection is required to create a pair of up- and down-travelling waves. If no dedicated structure is designed to allow for a strong Bragg condition technique method is limited to the total external reflection regime. Multiple other methods are in existence to satisfy the Bragg condition to yield additional angles with different standing wave spatial excitation patterns. By directing monocromatic X-rays at these specific angles one can induce a standing wave with a specific spacial excitation pattern in a structure. The fluorescence emittance of a respective photon energy is proportional to the atomic concentration multiplied by the location specific excitation intensity. A fluorescence signal dependent on this excitation pattern will be emitted from the individual atoms in structure by the general inelastic scattering pathways. Since every atom of a distinct element emits a unique set of wavelengths of fluorescent photons, this signal can be distinguished for every element and is location sensitive due to the specific excitation patterns. By measuring the structure for a set of angles, multiple 'snapshots' can be taken of a thin film structure to aid in the reconstruction.

Multilayer XSW generation By creating a periodic multilayer structure with a high reflectively, a standing wave with a wavelength equal to a multiple of the multilayer period of the sample (Equation 2.5) can be induced in a structure by directing X-rays at the different Bragg angles. By depositing a sample on top of this multilayer this excitation pattern can be induced with a variable phase in a to be analyzed structure. An example of such an excitation pattern that is induced by this method is shown in Figure 2.1.

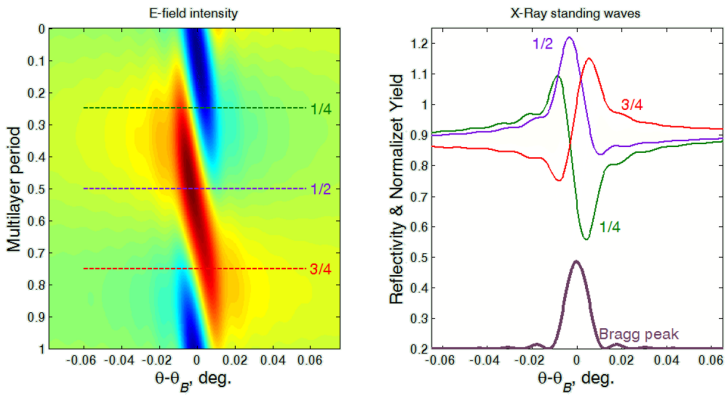


Figure 2.1: XSW excitation pattern induced by a multilayer.

Equation 2.5

Equation expressing the bragg angles:

$$\theta_n = \arcsin \frac{n * \lambda}{2 * d} \tag{2.5}$$

Where d the multilayer period.

Waveguide XSW generation Another method for generating a localized electromagnetic field distribution is by creating a thin film that is surrounded by a waveguide structure as seen in Figure 2.2. By directing X-rays at specific angles one can excite different waveguide modes that satisfy the bragg condition between the waveguide layers. These excitation patterns have a wavelength that follow Equation 2.6. This allows for a more detailled depth reconstruction of the different elements since different regions of the structure can be separately probed. An example of a waveguide structure and excitation pattern for different incidence angles of the incoming X-rays with resulting

Equation 2.6

Expression of the wavelength of the induced standing wave in the waveguide structure

$$\lambda = \frac{2L}{n} \tag{2.6}$$

Where L is the interior width of the waveguide and n is the order of the wave mode.

fluorescence signals is shown in Figure 2.2 with on the right side the resulting relative angular fluorescence signal coming from the different depth positions indicated on the left side.

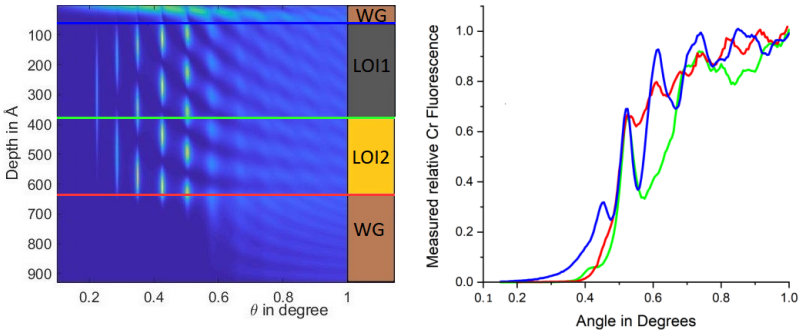


Figure 2.2: Angular dependent 1D excitation pattern induced by directing X-rays at different angles into the waveguide structure with the resulting relative angular fluorescence emmitances from the different depth positions on the right side.

2.3 Inverse Problems

Inverse problems are a class problems in which we try to explain a set of observations by calculating the causal factors that produced these observations. This kind of problem statement is frequently used when a direct calculation of the parameters of interest from a measurement or an experiment is not possible. [15]

Equation 2.7

The forward map of model f is defined as:

$$\boldsymbol{y} = f(\boldsymbol{x}) \tag{2.7}$$

Where $\boldsymbol{y}$ is the modeled observation caused by parameters $\boldsymbol{x}$.

Forward map The forward map f is defined as the model replicating the experiment, yielding the calculated observables $\boldsymbol{y}$ from a set of input parameters $\boldsymbol{x}$, in mathematical terms this is shown in Equation 2.7. The model f is made to replicate reality as accurate as possible. A perfect forward map with input $\boldsymbol{x}$ would yield the observables $\boldsymbol{y_0}$ that in reality would be caused by $\boldsymbol{x_0} = \boldsymbol{x}$.

The inverse problem is defined as the inverse of the forward problem. Instead of calculating the parameters $\boldsymbol{x_0}$ directly from the observables $\boldsymbol{y_0}$, a search is done to a set of parameters $\boldsymbol{x_0}$ that were the cause of $\boldsymbol{y_0}$. The goal is thus to find a parameter set $\boldsymbol{x}$ that when argumented in the forward map, yields the set of observations $\boldsymbol{y_0}$.

Equation 2.8

Loss Function Q is defined as:

$$Q(\boldsymbol{x}) = g(f(\boldsymbol{x}), \boldsymbol{y_0}) \tag{2.8}$$

Where $\boldsymbol{x}$ is the proposed set of parameters and $\boldsymbol{y_0}$ are the set of measured observables. Often for g the mean squared error between measurement and $f(x)$ is taken yielding:

$$Q(\boldsymbol{x}) = \frac{1}{\boldsymbol{\sigma}} \cdot |(f(\boldsymbol{x}) - \boldsymbol{y_0}) \circ (f(\boldsymbol{x}) - \boldsymbol{y_0})|$$

Where $\circ$ is the element wise multiplication (Hadamard product) and $\boldsymbol{\sigma}$ a vector of similar dimension containing the estimated modeling error and measurement error.

Loss function In practice however the construction of the forward map is not always easy and will never exactly replicate reality in a perfect way. This can either be due to modeling imperfections or measurement errors. Therefore a loss function, shown in Equation 2.8, is defined that yields a value depending on the agreement between the measurement $\boldsymbol{y_0}$ and the 'model' $f(\boldsymbol{x})$. With the loss function an optimization process can be used to find a single solution $\boldsymbol{x}$ based on the criteria of the smallest mismatch between measurement and simulation. This method however does not always yield $\boldsymbol{x_o}$ exactly. The distance of the found $\boldsymbol{x}$ and the actual cause $\boldsymbol{x_0}$ can be small or large depending on a the performance of the optimization routine or non-convexity and complexity of the loss function leading to a multi-modal solution space.

Ill-posedness Very often inverse problems suffer from ill-posedness, which is defined in [16] as a problem in which one or more of the following conditions is violated.

- There is a solution
- The solution is unique
- The solution depends continuously on the data

Regularization To address the difficulties of the ill-posedness and non-convexity of the loss function a regularization can be applied, which often is a punishment term of the form of a function of $\boldsymbol{x}$. This punishment term can help limit the number of solutions by assigning non-acceptable loss function values to unrealistic solutions. Also it can help flatten the loss function landscape and thereby easing the optimization of the loss function.

Relation to the GIXR and XSW-XRF Methods A direct reconstruction of the sample structure from a GIXR measurement is not possible due the missing phase information. To obtain more information on the structure than just the densities and thicknesses, a parameter estimation can be done by formulating a parametrization of the structure and creating a model that calculates the measurement that this proposed structure would yield. By formulating an inverse optimization problem with this forward map from the proposed structure parameterization to a measured signal, an optical constant profile can be found from the parameters of the paramterization that fits the measurement data. This can in practice however lead to multiple- but often non-physical solutions to a single measurement curve[10].

By extending the GIXR method with the X-ray standing wave fluorescence (XSW-XRF) technique and simultaneously satisfying both data sets this weakness is addressed since the latter technique is very much sensitive to the phase of the induced standing waves, potentially restricting the set of acceptable solutions.

2.4 Bayesian Inference

Instead of inferring a single solution, another way of approaching an inverse problem is determining how likely a proposed solution $\boldsymbol{x}$ is given a measurement $\boldsymbol{y_0}$. In mathematical terms this is expressed in Equation 2.9 (Bayes Formula). With Bayesian inference one does not specifically determine one solution from given information $\boldsymbol{y_0}$ but infers a probability distribution by assigning a conditional probability to each parameter set $\boldsymbol{x}$ that could have been the source of the information $\boldsymbol{y_0}$. [17]

Equation 2.9

Expression of the posterior Distribution:

$$\pi_*(x) = \pi(x|y_0) = \frac{\pi(x)\,\pi(y_0|x)}{\pi(y_0)} \tag{2.9}$$

Where y_0 is the measured data, $\pi_*(x)$ is the posterior distribution and $\pi(x)$ is the prior distribution.

The prior distribution $\pi(x)$ contains the all prior available knowledge on $\boldsymbol{x}$ which limits the possible range of $\boldsymbol{x}$ and thereby suppress the non-uniqueness. This can be a regularization term or any other information that narrows down the possible regions without excluding any potential values of $\boldsymbol{x}$ that could have been the source of $\boldsymbol{y_0}$. The posterior distribution is the likelihood given information $\boldsymbol{y_0}$ its source was indeed $\boldsymbol{x}$. Obtaining this distribution is the goal of the inference and in theory it contains all the information on $\boldsymbol{x}$ (means, uncertainties, correlation etc) with the given the measurement $\boldsymbol{y_0}$.

Equation 2.10

Expression of the target distribution:

$$\Pi(dx) = Z^{-1} e^{-U(x)} dx \tag{2.10}$$

Where Z is the Normalisation factor.

For the application on inverse problems a likelihood function $\pi(y|x)$ takes the function of the forward map and is defined by assigning a probability to every possible value of y given a target x. A target distribution often used in inverse problems is shown in Equation 2.10. [17] The exact value of Z is not required to be known but is defined as the integral of $e^{U(x)}$ over all values of

Equation 2.11 & 2.12

maximum a posterori estimator:

$$x_{MAP} = \arg_x \max \pi_*(x) \quad (2.11)$$

Conditional Mean:

$$x_{CM} = \mathbb{E}\langle x|y_0\rangle = \int_{-\infty}^{\infty} x\pi_*(x)dx \quad (2.12)$$

Equation 2.13

The symmetric Baysian credibility set is defined as $I_k(A) = [a,b]$ satisfying:

$$\int_{-\infty}^{a} \pi_k(x_k)dx_k = \int_{b}^{\infty} \pi_k(x_k)dx_k = \frac{A}{200} \quad (2.13)$$

Where $\pi_k(x_k)$ is the marginal posterior density of the k-th parameter.

Equation 2.14

$$\widetilde{F}_N = \frac{1}{N}\sum_{i=1}^{N} F(X_i) \quad (2.14)$$

Equation 2.15

The detailed balance equation:

$$\pi_*(x)K(x,y) = \pi_*(y)K(y,x) \quad (2.15)$$

Where K is defined as:

$$K(x,y) = \Phi(y|x)\alpha(y|x)$$

Where $\Phi(y|x)$ is the proposal distribution at sample x to proposal y and $\alpha(y|x)$ its acceptance probability.

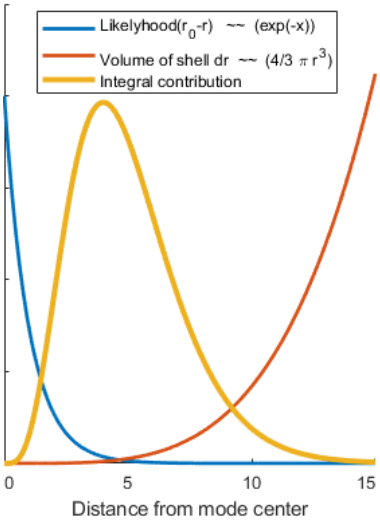


Figure 2.3: Visualisation of the typical set integral.

x. Here $U(x)$ is a mapping $U : \mathbb{R}^d \to \mathbb{R}$ and represents a loss function between the forward map from x to y and the actual measured information y_0 and is generally tuned to the problem at hand.

Estimators Since the complete posterior visualisation and analysis is often impossible due to the high number of dimensions, estimators can be calculated which contain more concise information on the posterior. The measure x_{MAP} in Equation 2.11 [17] gives the most likely parametrization of x given a certain measurement y_0, which is essentially the same parametrization that is searched for using a general optimization method looking for the x that results in the lowest value of the loss function. x_{CM}, described in Equation 2.12 [17], is the mean of probability distribution of x given a measurement y_0, and just as with normal expectation values does not necessarily represent a probable value itself.

Confidence Intervals A measure exploiting the whole set of probable values of x that can be calculated using the posterior distribution is the symmetric Baysian credibility set. This set contains the marginal values of x that lie within the A% confidence interval of the marginal distribution of x_k. The expression of this set is given in Equation 2.13 [17]. This measure yields a confidence interval of the individual parameters x_k within a chosen percentage A.

2.5 Monte Carlo Markov Chains (MCMC)

The challenging part of the Bayesian inference is in obtaining an accurate approximation of the posterior distribution. Multiple methods exist to obtain the posterior with all methods having as the limiting factor time. As is expected, most of these methods scale super-linearly in time with the number of dimensions that are explored. The MCMC is a Monte Carlo method that can be used to obtain a sequence of samples that converge to a target probability distribution $F(x)$ where $x \in \mathbb{R}^d$ from which direct sampling is unfeasible.

In a MCMC a chain of sequential samples is drawn using a transition kernel that is only dependent on the previous sample that was accepted (Markov Property). By determining a starting point X_0 and successively proposing samples that are either rejected or accepted and then appended to the chain, a chain of samples is accumulated. This process is shown in Algorithm 2.16 for the specific case of the Metropolis-Hastings algorithm. If the samples are drawn in a way that leaves chain invariant to the target distribution invariant, meaning that the sampled Markov chain is distributed according to the target distribution $F(x)$, an accurate approximation of $F(x)$ can be obtained by averaging the samples in the chain. This is expressed in Equation 2.14. The invariance to the target distribution is ensured when the sampled posterior distribution together with the transition probability K forms a non-periodic, reversible ergodic Markov chain that satisfies the detailed balance equation which is shown in Equation 2.15.

Typical Set The size of a spherical (hyper-)volume element scales with its radius to the power of the dimension of the space (See the red line in Figure 2.3). Therefore when exploring high-dimensional spaces with associated probability measures, the thin neighbourhood around a high-probability point x_0

generally has a larger integration volume contribution than the neighbourhood containing point itself. For volume fractions far away from these points of high probability the contribution from the likelihood term (See the blue line in Figure 2.3), which scales with a negative exponent, becomes so small that these regions also do not pose a significant contribution to the probability distribution integral. In Figure 2.3 the integral contribution, which is the product of the volume of shell at radius r times the likelihood, is shown in yellow. This thin neighbourhood around the high-probability modes that dominates the integral is called the typical set and in practice is the only region that needs to be explored since the rest of the parameter space of x does not pose an important contribution in high dimensions.

Metropolis-Hastings A commonly used MCMC algorithm is the Metropolis-Hastings algorithm which follows a very simple procedure shown in Algorithm 2.16. Generating a starting point $\widetilde{X}_0$ can be done by e.g. taking an arbitrary X that satisfies the boundary conditions. This can however lead to a long burn-in time, which is the time needed for the chain to reach equilibrium, which should be ignored in the posterior approximation. One can also find a starting point by a local optimization algorithm to reduce this burn-in time since this point will be closer to the typical set. Starting from X_0 random samples are drawn from a probability distribution $\Phi(\widetilde{X}_{n+1}|\widetilde{X}_n)$. Since the proposal X_{n+1} depends only on X_n the generated chain has the Markov properties. Once a proposal has been made a specific accept/reject probability α is used that keeps the chain of accepted samples invariant to the target distribution by satisfying the detailed balance equation. (See Equation 2.17)

Often used distributions for the proposals are of the form $\Phi(\widetilde{X}_{n+1}|\widetilde{X}_n) = \mathcal{N}(\widetilde{X}_n, \sigma^2)$ where σ is tuned depending on the problem at hand. A larger σ reduces the correlation between the samples and therefore leads to a faster convergence and more efficient exploration of the typical set. But if the parameter space of $F(X)$ is high-dimensional or has a complex probability landscape many proposals have to be generated before a region of high-probability is found therefore greatly increasing the time needed to generate an accepted sample.

Convergence rate The accuracy of the MCMC can be estimated under ideal behavior with Equation 2.18 [18], which is equivalent to the central limit theorem. This Equaution implies that the standard deviation of the difference between an approximated measure $\widetilde{g}$ of x and the actual $g(X)$ goes down with the inverse square root of the number of effective uncorrelated samples of which the approximation consists of.

2.6 Hamiltonian Monte Carlo (HMC)

The HMC is a special type of Metropolis-Hastings sampling method that does not rely on random guesses or random walks to explore the parameter space but finds a direction to samples of similar likelihood by using the derivatives of the vector field of the loss function $U(x)$. This is done with the help of an auxiliary randomly generated vector and a deterministic integration process evolving the system over time using the derivatives for determining the proposal direction.

The idea is to treat the parameters x as a spatial position, start in a low point in the potential energy landscape shaped by the loss function $U(x)$ and introduce an auxiliary momentum parameter set (of similar dimension) that acts like

Algorithm 2.16

(**Step 0:** Generate $\widetilde{X}_0$)
Step 1: Draw proposal $\widetilde{X}_{n+1}$ with $\Phi(\widetilde{X}_{n+1}|\widetilde{X}_n)$.
Step 2: Accept $\widetilde{X}_{n+1}$ with probability α.

Where alpha is defined as:

$$\alpha(\widetilde{X}_{n+1}|\widetilde{X}_n) = \min\{1, \frac{\pi_*(\widetilde{X}_{n+1})\Phi(\widetilde{X}_n|\widetilde{X}_{n+1})}{\pi_*(\widetilde{X}_n)\Phi(\widetilde{X}_{n+1}|\widetilde{X}_n)}\} \quad (2.16)$$

With this acceptance probability $\pi_*(x)$ satisfies the detailed balance equation from Equation 2.15:

$$\begin{aligned}
&\pi_*(x)K(x,y) \\
&\quad = \pi_*(x)\Phi(y|x)\alpha(y|x) \\
&\quad = \pi_*(x)\Phi(y|x)\frac{\pi_*(y)\Phi(x|y)}{\pi_*(x)\Phi(y|x)} \\
&\quad = \pi_*(y)\Phi(x|y) \\
&\text{w.l.o.g } \alpha(x|y) = 1 \text{ since it is} \\
&\text{assumed that } \alpha(y|x) <= 1 \\
&\quad = \pi_*(y)\Phi(x|y)\alpha(x|y) \\
&\quad = \pi_*(y)K(y,x) \quad (2.17)
\end{aligned}$$

Equation 2.18

The convergence behavior of a Monte-Carlo Markov chain depending on the number of samples $N \xrightarrow{\infty}$ can be expressed using the central limit theorem:

$$(\widetilde{g}_N - g(x)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \frac{\sigma(F)^2}{N}) \quad (2.18)$$

Where N is the number of samples, $g(x)$ is a real valued function with finite variance and σ is:

$$\sigma(g)^2 = \sigma_0(g)^2 + 2\sum_{i>1}\mathrm{cov}(g(X_1), g(X_i))$$

Where chain is assumed to be at stationary.

a small kick of added momentum with a comparable energy to the current position. From this point a path can be traced to a new point in the mechanical phase space of similar probability by evolving the system from Equation 2.19 over time for a duration of λ. This mechanical phase space is a construct in which all states of x and p are represented and have a unique position, similarly to normal space in which only x is represented. When a new point has been reached after integrating for time λ the marginal parameter set p is discarded and x_λ either declined or accepted added to the chain of samples depending on the acceptance criteria.

The advantage of the HMC over the Metropolis-Hastings is clearly visible in high-dimensional spaces. The acceptance rate of random sampled proposals exponentially decreases with the number of dimensions while for the HMC the gradient of the loss function is exploited to be almost guaranteed an accepted sample is found on the first try. The second advantage of the HMC is the inherent directionality of the added momentum parameters in which the integration time can be increased to de-correlate successive samples. Where the distance between two consecutive samples in the Metropolis-Hastings chain is limited by the acceptance probability and complexity of the parameter space leading to random walk behavior with very small step-sizes, the HMC integration time can in be chosen to sample points having much larger distance and less correlation. The expected distance traveled by a random walk is the square root of the number of samples while for the HMC a path can be traced efficiently through space where each successive step can be made in a correlated direction [19].

Equation 2.19

Expression of the Hamiltonian:

$$H(x,p) = U(x) + T(p) = U(x) + \frac{1}{2}p^T M^{-1} p \quad (2.19)$$

Where $U(x)$ is the potential energy(often the loss function of an inverse problem), T is the kinetic Energy ,x and $p \in \mathbb{R}^d$ are the corresponding position and momentum vectors, M is a positive definite matrix $\in \mathbb{R}^{dxd}$ and d is the number of degrees freedom.

Where the system $H(x,p)$ satisfies the Hamiltonian conditions (2.20 and 2.21) defined as: [18]

$$\frac{d}{dt}\begin{bmatrix} x \\ p \end{bmatrix} = J^{-1}\nabla H = J^{-1}\begin{bmatrix} \frac{\delta H}{dx} \\ \frac{\delta H}{dp} \end{bmatrix} \quad (2.20)$$

Where J is defined as:

$$J = \begin{bmatrix} 0_{dxd} & -I_{dxd} \\ I_{dxd} & 0_{dxd} \end{bmatrix} \quad (2.21)$$

For the Equations of motion:

$$\frac{d}{dt}x = M^{-1}p \quad \& \quad \frac{d}{dt}p = -\nabla U(x) \quad (2.22)$$

For these conditions (2.20 and 2.21) and Equations of motion (2.22) the Hamiltonian value is conserved:

$$\frac{dH}{dt} = \nabla H J^{-1}\frac{d}{dt}\begin{bmatrix} x \\ p \end{bmatrix} = \nabla H J^{-1}\nabla H = 0$$

Hamiltonian mechanics First a vector $p \in \mathbb{R}^d$ containing random numbers is generated that complements the parameter set of interest $\boldsymbol{x} \in \mathbb{R}^d$. With the (x,p) parameter pair a Hamiltonian system is created with the phase space of the position vector $\boldsymbol{x}$ and momentum vector $\boldsymbol{p}$ that fulfills the general criteria for a Hamiltonian system shown in Equation 2.19. New samples along the chain can be generated by evolving the Hamiltonian vector field, containing the potential energy term and the kinetic energy term, over time.

The mass matrix M in Equation 2.19 is a positive definite matrices and can be chosen based on the specific structure of the problem to give a weight to certain parameters. This weight acts like a natural mass and parameters with a higher mass will require more energy to change their momentum. $U(x)$ is the potential energy and contains the function on which the parameter space will be explored. (i.e. the a loss function between a measurement and simulation.) An arbitrary starting point for $\boldsymbol{x}$ can be found by using a simple local optimization or by drawing a random vector that satisfies the boundary conditions of the prior.

The Hamiltonian has 4 properties that are key to keeping the sampled Markov chain invariant to the target distribution:

- Conservation of the energy. (under flow ϕ: $H \circ \phi_t = H$)
- Symplectic flow. (By definition, see Equation 2.25)
- Reversibility. ($H(x,p) = H(x,-p)$)
- Preservation of the $dx \cdot dp$ phase space volume. (direct consequence of symplectic flow)

The last property is in practice the most important one. This property ensures the invariance of the sampled Markov chain to the target distribution. Because due to the volume preservation of $dx \cdot dp$, the probability measure in Equation 2.23 is preserved by the flow of the Hamiltonian system. Meaning that the

Equations 2.23 & 2.24

The invariant preserved probability measure of the Hamiltonian:

$$\mu(x,p) = \frac{e^{\beta H(x,p)}}{Z} \quad (2.23)$$

For a Hamiltonian of the form in Equation 2.19 this equation becomes the Boltzmann-Gibbs distribution:

$$\mu(x,p) = (2\pi)^{-\frac{d}{2}}|\det(M)|e^{-\frac{\beta}{2}p^T M^{-1}p}\cdot \frac{e^{\beta H(x,p)}}{Z}dxdp \quad (2.24)$$

Where Z is a normalization constant and β as the inverse of the temperature T.

Implying that for each t:

$$\mu(x,p)) = \mu(H(x,p)\circ\phi_t)$$

volume occupied by a certain element $dx \cdot dp$ in phase space by this measure is preserved when integrated over time.

Since the integration generally is done numerically, to ensure that these properties are conserved and not too large truncation errors are accumulated during the successive steps, a careful consideration of the integrators should be done. Different integrators exist that conserve several of these properties [18] but no numerical integrator conserves all of these properties simultaneously. Therefore in practice the best choice of numerical integrator is one that conserves the 3 bottom properties, which is the symplectic integrator. A symplectic integrator is defined as any integrator which for the determinant of the jacobian of its mapping is one, meaning the volume element in phase space of the canocial coordinates x and p between different times is conserved [18]. Mathematically this means that for each integration step its mapping satisfies the condition shown in Equation 2.25.

Equation 2.25

A mapping Φ is symplectic or canonical if at each point $(q, p) \in \mathbb{R}^{2d}$:

$$\Phi'(x,p)^T J \Phi'(x,p)^T = J \qquad (2.25)$$

Where $\Phi'(x,p)$ is the Jacobian matrix of Φ.)

Depending on the step-size a 'shadow' Hamiltonian is followed when integrating with a symplectic update scheme which diverges from the original Hamiltonian energy-wise with $\mathcal{O}(h^n)$ given an integrator of order n. Integrating this way multiple successive steps can be made sequentially without diverging into low-probability regions of significant higher energy (which have negligible contribution to the actual target distribution).

The most commonly used integrator for the HMC is the Verlet velocity integrator. This is a splitted symmetric explicit integrator in which the position/-momentum parameters are splitted and updated in separated steps using the intermediate obtained available results. This is possible (At the cost of an extra truncation error) since the equations of motion are separated in two Hamiltonian systems of variables for which the flow can be evaluated individually (See Equation 2.22). The update scheme of this integrator is shown in Algorithm 2.26 with its symplecticness shown in Example 2.27. Due to its palindromic structure the reversibility property of the Hamiltonian is also conserved. The advantage of the Verlet scheme is although it is a second order scheme the gradient only has to be calculated once, since the gradient from the previous step can be reused for the evaluation of $p_{n+\frac{1}{2}}$. In practice the Verlet scheme generally has shown the best performance in terms of truncation error versus time spend per step in HMC sampling [18]. Its complement, the position Verlet scheme $(\phi^{(x)}_{\frac{1}{2}} \circ \phi^{(p)}_{1} \circ \phi^{(x)}_{\frac{1}{2}})$ is also commonly used and their individual performance can be problem dependent.

Algorithm 2.26

Verlet scheme: $(\phi^{(p)}_{\frac{1}{2}} \circ \phi^{(x)}_{1} \circ \phi^{(p)}_{\frac{1}{2}})$,

Where $\phi^{(x)}_{s}$ is defined as the flow of x over time period s following the equations of motion (Equation 2.22).

$$p_{n+\frac{1}{2}} = \phi^{(p)}_{\frac{1}{2}}(x_n, p_n) = p_n + \frac{h}{2}F(x_n)$$

$$x_{n+1} = \phi^{(x)}_{\frac{1}{2}}(x_n, p_{n+\frac{1}{2}}) = x_n + hM^{-1}p_{n+\frac{1}{2}}$$

$$\begin{aligned} p_{n+1} &= \phi^{(p)}_{\frac{1}{2}}(x_{n+1}, p_{n+\frac{1}{2}}) \\ &= p_{n+\frac{1}{2}} + \frac{h}{2}F(x_{n+1}) \end{aligned} \qquad (2.26)$$

Where $F(q_n)$ is $-\Delta U(q)$ following from the equations of motion. (Equation 2.22).

Acceptance In Algorithm 2.28 the most simple form of the Hamiltonian Monte Carlo is outlined. In absence of truncation errors, because the energy of the system is not altered during integration and space phase volume is conserved in the proposals $(x_0, p_0) \longleftarrow (x_\lambda, p_\lambda)$, is the chain invariant to the Gibbs-Boltzmann distribution since every step leads to an equally probable state. Therefore a chain sampled using this algorithm is also invariant to the marginal $F(x)$ which is independently distributed from the momentum parameters. Since it is impossible to chose a numerical integrator that conserves both the energy and the phase space of the Hamiltonian, a correction has to be made for the bias that is induced due to the truncation errors. When using a symplectic integrator, the phase space is preserved, only leaving the need to correct for the energy error between the samples. This correction is exactly the classic Metropolis-Hastings correction, but instead correcting for the difference in likelihood, the correction is applied to the energy difference between the steps.

Example 2.27

By substitution direct expressions can be found:

$$\begin{aligned} p_{n+1} &= p_n - \frac{h}{2}(U(x_n) \\ &+ U(x_n + hM^{-1}(p_n - \frac{h}{2}U(x_n)))) \end{aligned}$$

$$x_{n+1} = x_n + hM^{-1}(p_n - \frac{h}{2}U(x_n))$$

Substituting Equation 2.25 for the 2 dimensional case:

$$\begin{aligned} &\frac{x_{n+1}}{x_n}\frac{p_{n+1}}{p_n} - \frac{x_{n+1}}{p_n}\frac{p_{n+1}}{x_n} \\ &= (1 - \frac{h}{2})^2 - h(\frac{h^3}{4} - h) = 1 \end{aligned} \qquad (2.27)$$

By tuning the acceptance rate α to the energy error, this method can be reduced to a well understood Metropolis-Hastings chain with a particular way of proposing samples. Because of the reversibility properties and the preserved probability measure in Equation 2.24, the proposal distribution $\Phi(\widetilde{X}|\widetilde{Y})$ is identical in both directions. Therefore the acceptance rate of the proposal (x_λ, p_λ) from algorithm 2.16 reduces to the natural exponent of the negative difference in the Hamiltonian energy between the sequential samples that occurs due to the accumulation of truncation errors in the numerical integration (See Algorithm 2.28).

Multiple variations have been proposed that in specific cases allow for a faster exploration of the target distribution. e.g. One can use a variable λ as the integration time. Other variations have been proposed where in the last couple of steps of the numerical integrator the q and p combination of the lowest energy is chosen to increase the acceptance probability. Also a dynamic step-size can be used to increase the exploration speed for highly non-convex energy landscapes in which the curvature varies from region to region, although this modification can break the invariance of the chain but in practice can result in a faster useful exploration. [18]

Algorithm 2.28

Let λ be the time duration parameter, which determines the integration time between successive samples.
Step 1: Draw a d-dimensional vector from the distribution chosen for p.
Step 2: Evolve the Hamiltonian over time for a duration of λ starting from (q_0, p_0)
Step 3: Accept proposal $(x_\lambda,\ p_\lambda)$ with probability α.

$$\begin{aligned}\alpha &= \min\{1, \frac{\pi_*(\widetilde{X}_\lambda)\Phi(\widetilde{X}_0|\widetilde{X}_\lambda)}{\pi_*(\widetilde{X}_0)\Phi(\widetilde{X}_\lambda|\widetilde{X}_n)}\}\\ &= \min\{1, \frac{\mu(\widetilde{X}_\lambda)}{\mu(\widetilde{X}_0)}\}\\ &= \min\{1, e^{\beta(H(x_0,p_0)-H(x_\lambda,p_\lambda))}\}\end{aligned} \tag{2.28}$$

Where μ is taken from Equation 2.24.

Theoretical Comparison of Metropolis-Hastings to the HMC The HMC in theory should converge faster than conventional Monte Carlo methods since it does not suffer so much from the random walk phenomena and the curse of dimensionality in finding a direction for a proposed next step, leading to a faster exploration of the typical space. In practice the costliest part of the HMC is the calculation of the derivative $\nabla U(x)$. Since target distributions that are approximated in these problems almost always have a high dimensionality ($d \gg 1$), an efficient gradient calculation is not always possible and can cost several forward evaluations. When calculating the gradient with a finite difference for every dimension of x, every step will be d times as expensive.

The parameter space of interest that will be explored is high-dimensional and for the conventional guess and check Metropolis-Hastings in theory to reach a specified error margin has a computational cost scaling as $\mathcal{O}(d^2)$ [12]. For the HMC chain this scales as $\mathcal{O}(d^{\frac{5}{4}})$ [12], if the derivatives are calculated for each direction individually by evaluating the loss function, the performance of both methods is on paper very comparable so a smart implementation of obtaining the derivatives of the loss function $U(x)$ is critical.

Equation 2.29

Potential energy function of example:

$$\begin{aligned}F(x,y) &= (\frac{200000}{1+x^2+y^2})^{0.12}\\ &+ (0.003(x^8+y^8) - 10(x^2+y^2))^{0.3}\end{aligned} \tag{2.29}$$

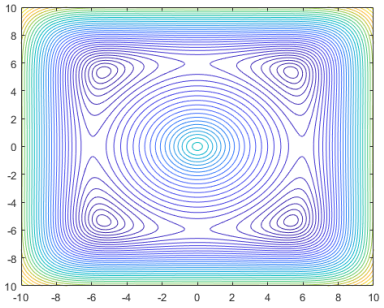


Figure 2.4: Contour lines of the energy landscape of Equation 2.29 that is used in the example of the MCMC.

2.7 MC Example

A small example will be given for the 2D case of the MCMC techniques that will be applied to the thin film inverse problem to provide an introduction to the detailed implementation to the different methods.

Energy Function The example potential energy function in given in Equation 2.29. Designed to be a semi-challenging energy landscape with multiple minima for the MCMC to explore and is shown in Figure 2.4.

Metropolis-Hastings Using a proposal distribution of the form of Equation 2.30 and an acceptance probability following Equation 2.31 with a $\sigma = 1$ and $T = \frac{1}{10}$. The random walk Guess and Check metropolis-Hastings after roughly 700 steps (1000 proposed samples) is only able to properly reconstruct the part of the energy functional close to the starting point and leaves parts of the

high-probability areas unexplored. The sampling process for the Metropolis-Hastings chain is shown in Figure 2.5

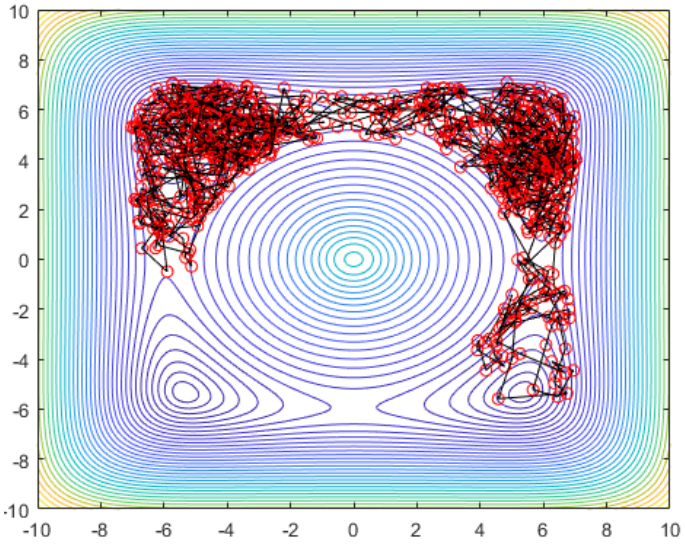


Figure 2.5: The path and samples tracked with the Metropolis-Hastings algorithm for 1000 proposed steps.

Equations 2.30 & 2.31

Proposal distribution:

$$\Pi(\widetilde{X}_{n+1}|\widetilde{X}_n) = \mathcal{N}(\widetilde{X}_{n+1} - \widetilde{X}_n, \sigma) \quad (2.30)$$

Acceptance probability:

$$\alpha = \min\{e^{\frac{-F(x_1,y_1)+F(x_0,y_0)}{T}}, 1\} \quad (2.31)$$

HMC The HMC paths are are as expected following a single direction bent by the energy landscape thereby overcoming the random walk behavior between the intermediate time steps in the same sample. There is a significant visible improvement in de-correlating the intermediate sample. The cost is, of course, that only 50 accepted samples can be drawn while at the same time the using the Guess and check method 700+ accepted samples can be drawn. The sampling process for the Hamiltonian Monte Carlo chain is shown in Figure 2.6

Settings 2.33

h: 0.2
Steps per Sample: 20
Integration time: 4

$$p = \mathcal{N}(0,1) \quad (2.32)$$

$$\frac{dF}{dx} = \frac{-\frac{3}{10}\frac{20x+160000x}{(x^2+y^2+1)^{\frac{7}{5}}} + \frac{3}{10}\frac{x^7}{125}}{\left(\frac{200000}{(x^2+y^2+1)^{\frac{2}{5}}} - \frac{x^2+y^2}{0.1} + \frac{x^8+y^8}{333.3} + 100\right)^{\frac{7}{10}}} \quad (2.33)$$

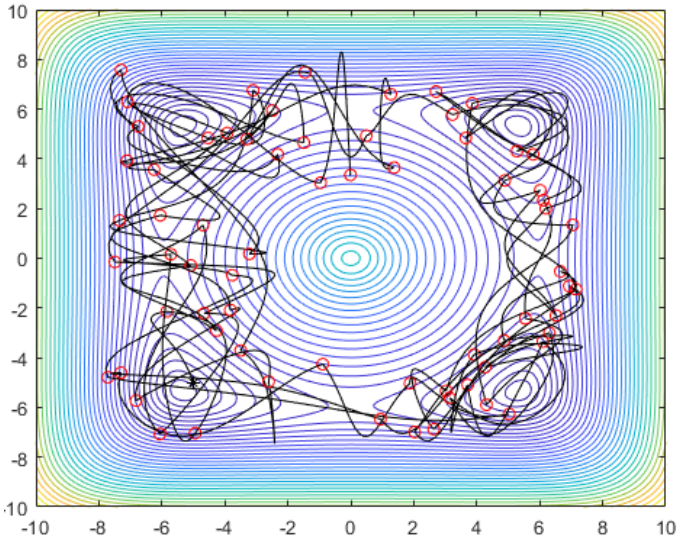
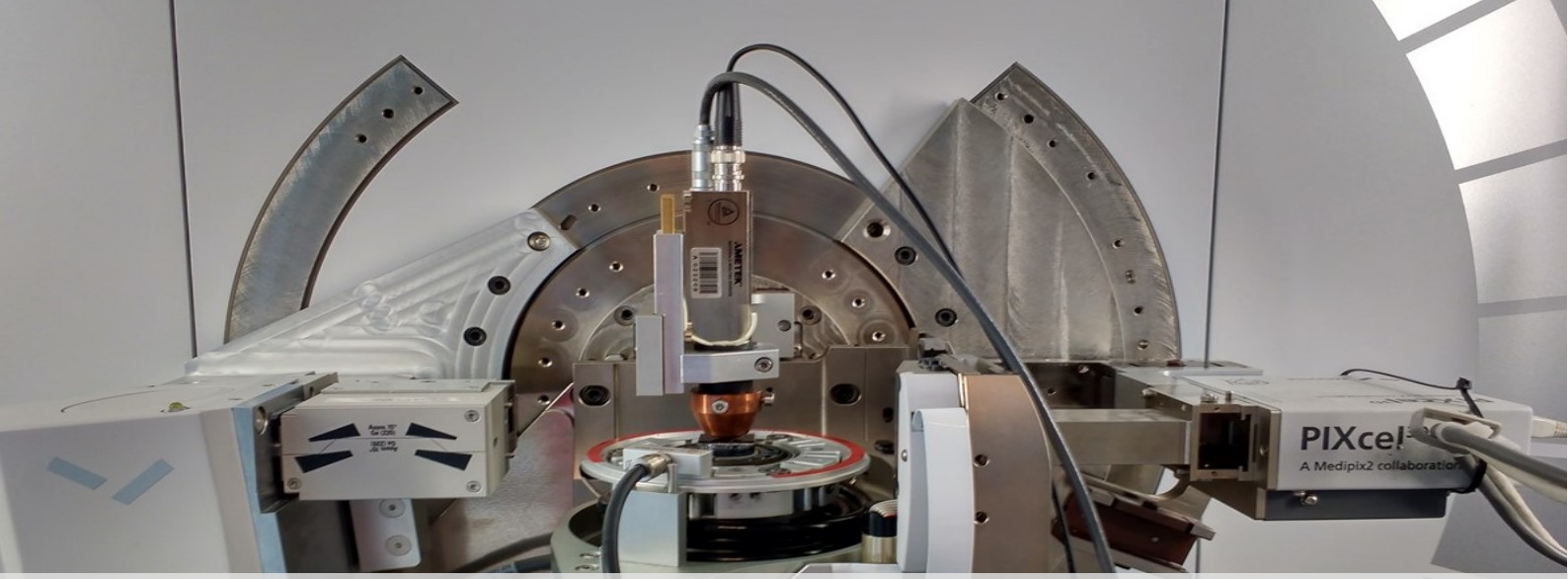


Figure 2.6: The path and samples tracked with the HMC algorithm for 50 accepted samples.

One can imagine, when increasing the numbers of dimensions and minima, the HMC becomes more and more preferable to use when exploring parameter spaces.



3 Experimental Methods

In this chapter the experimental aspects are introduced. First the sample design and engineering is discussed and afterwards the data acquisition is discussed.

3.1 Sample Design

In the design of the samples it has to be taken in to account that only 4 material mounts are present in the deposition machine. The waveguide only takes up a single material slot in the ADC machine used for the depositions, in comparison with the multilayer excitation methods, which requires 2 materials to be deposited for manufacturing. The waveguide also in theory allows for more different orthogonal components in the resulting signals due to the availability of multiple wavelength excitation patterns from a single structure. Also the waveguide structure is expected to be more resistant to annealing treatments. Therefore the waveguide is chosen as the XSW excitation method of choice. The basic sample design is schematically presented in Figure 3.1. The dimensions of the different layers are optimized to obtain the best experimental performance in terms of discerning power and thermal resistance.

Layer dimensions The criteria to take into account for determining the optimal layer dimensions are thermal stability, layer grotwh behavior and the change in the emitted XSW signal when the dopant element changes in distribution (segregation). e.g. if the layers are too thin they might not grow properly or partly intermix when they are annealed. The as deposited structure should clearly resemble 2 separate layers, of which one is enriched with a dopant, to keep the conditions of the experiment in line with the research goals.

Material Choice The material combinations in this research are chosen based on multiple criteria. The primary goal is to isolate a set of fundamental atomic parameters to investigate whether they play an important role in the tendency for a system to exhibit interfacial segregation. The 3 main parameters of investigation are:

- Atomic Radius
- Crystal Structure
- Surface Energy

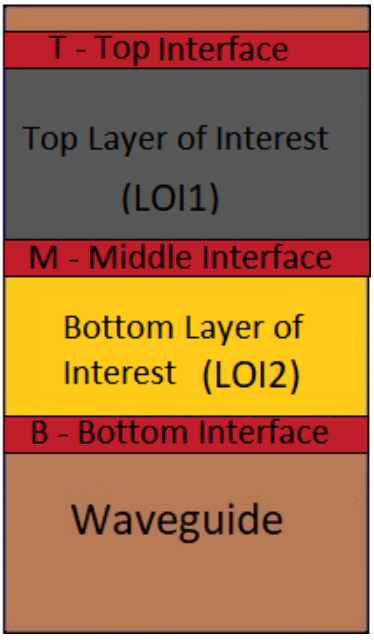


Figure 3.1: Schematic illustration of the to be deposited standardized structure. Throughout this thesis the notation in terms of materials for these samples will be 'Primary'/'Secondary'/'Ternary'. In the example of Table 3.2 this reduces to Cr/Sc/Co

The materials should also be chosen in a way that the layers grow properly as will be explained in the engineering paragraph below.

Engineering limitations The deposition apparatus is limited to 4 magnetrons and we also have a waveguide material which occupies a slot, the selection is very limited. A suitable material combination of 3 materials is defined as a dopant (Primary), a matrix material (Secondary) and the Ternary material for the top layerm, satisfying: when the matrix material and the ternary material are interchanged, only 1 of these fundamental parameters changes with respect to the dopant matrix combination. Once deposition apparatus has had a bakeout it is very easy just to deposit any structure combination of the inserted materials, therefore when a material selection has been made all remaining combinations will be deposited regardless of the criteria. To refer to the samples with the specific materials, the notation Primary/Secondary/Ternary will be used. For the Cr/Sc/Co sample, an example of this parameter isolation the Atomic radius is shown in Figure 3.2.

Element A	Element B	Structure	Atom size difference (A-B)	Surface energy difference (A-B)
Cr (small atom)	Sc (large atom)	BCC-HCP	-0.34 Å	0.11 eV/atom
Cr (small atom)	Co (small atom)	BCC-HCP	0.03 Å	-0.07 eV/atom
		SAME	DIFFERENT	SIMILAR

Figure 3.2: A table showing the process of isolating material parameters. In this case atomic radius is varied while keeping the other parameters roughly similar.

The structure feasibility also has to be taken into account. Since the structures are assumed to be 1D grown thin films the layers must exhibit nearly ideal growth. To achieve this some materials cannot be deposited on one another. By depositing materials with a very high surface energy on a low surface energy material, 2D or 3D growth can occur which is unfavourable.

To be able to take into account these material combinations and cross checking them with availability of the materials in the laboratory and feasibility of the deposition(i.e. some materials will not grow properly when grown on a materials with low surface energy.). A MATLAB program was developed to incorporate all criteria to be able to make an educated choice of materials to allow for an efficient research method. The following criteria were put into the selection program.

- Per material combination 1 parameter of interest should be isolated. (Atomic Radius, Crystal Structure and Surface Energy)
- All elements should emit simultaneously measurable and discernible Fluorescence signals excitable with the Cu-Kα radiation used in the laboratory and being measurable in an ambient enviroment.
- The materials should be selected in a way that only 1D growth occurs.
- A target should available in the laboratory for deposition of that material.

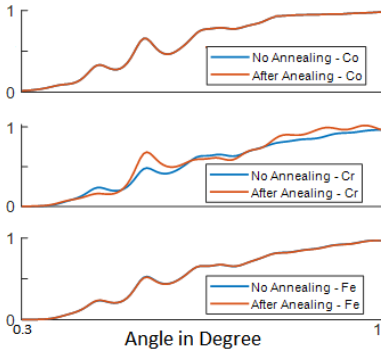


Figure 3.3: Simulated signal of a Cr/Fe/Co simulated structure where 15% of the dopant material has segregated to the bottom after annealing.

Regarding layer thicknesses, the signal that will be measured from the interface can be drowned out by the signal from the dopant that is left in the as deposited layer so we can only measure the segregation when a substantial relative change has taken place in the distribution of the dopant. If the layers are too thick the relative change due to the segregation is not large enough to change the fluorescence signal substantially to be measured since only a very small (mono)layer is expected to from between the interface. Therefore the dopant concentration should be as low as possible and the layers sufficiently thin.

The bottom waveguide layer should be as thick as possible since the function is to reflect as much of the incoming X-rays as possible. A problem was that Tungsten tended to become rough when grown thick. Simulations showed

that more than 40nm did not contribute to significantly more reflection while still being decently flat when grown on silicon.

The top waveguide layer was subject to a trade-off between the reflectance from the inside to the potential of the x-rays to penetrate trough the layer below the critical angle. Simulations show than a thickness of 5-6nm led to the highest contrast depending on the density of the layers inside the waveguide. A growing width of 6 nm was found to yield the optimal results since in practice a layer rarely has a sub-angstrom interface width.

All samples used in this research follow the optimized criteria proposed above. A 6nm top layer of tungsten, A 15nm ternary layer of material C , a 10nm layer of matrix material B(95%) and dopand Material A(5%) and a 40nm layer of Tungsten. An illustration of the system is is shown in Figure 3.1. A non symmetric layer system of 10nm and 15nm was determined, looking at the excitation pattern in a waveguide it is obvious that this leads to better discerning power between the different potential segregation sites.

To put this in numbers if we start out with concentration of 5% dopant in a layer of 10nm and the interface become enriched with 50% of the dopant (taking the atomic radius to be 150pm) about 15% of the dopant will have segregated. This leads to a substantial relative change in the atomic distribution which in the simulations lead to a discernable difference in angular fluorescence signal as shown in Figure 3.3.

The criteria for which the sample structure is optimized are as follows:

- Maximize sensitivity to dopant movement.
- Inducing a unique XSW signal for all potential segregation sites.
- Thermally stable thin-film structure under the annealing condition.
- Consisting of at maximum of 4 different elements.
- Having clear separated layers that survive the annealing process.

Segregation to interfaces and layers No segregation trough or mixing with the Tungsten is expected to be observed due to the thermal stability of the material. In the case that this does happen this will be very observable since the density will change drastically and the fluorescence signal from any material passing trough or in the tungsten will no longer be suppressed below the critical angle of the incident X-rays and therefore be very different from any signal coming from within the waveguide. In practice there are 3 potential interfaces and 2 solid layers of interest. The terminology used to refer to these layers and interfaces is introduced in Ffigure 3.1. The material specific Samples are referred to as Primary/Secondary/Ternary. (i.e. (V/Sc/Nb).). An illustrations of the different dimensions and the segregation process is give in Figure 3.6.

3.2 Depositions

All depositions were performed using the Advanced Deposition Coater in the XUV depositions laboratory. Only magnetron depositions were done to manufacture the samples. A bakeout was done after every material sequence which required opening the machine. A picture of the ADC deposition machine that was used is shown in Figure 3.4.

Figure 3.4: A picture of the ADC Deposition machine that was used to deposit these samples.

Co-Depositions To create an alloyed layer with a dopant element co-deposition is used. This is a process 2 magnetrons are simultaneously ignited during the sputtering process, resulting in a deposition of a mixed layer on the structure.

To control the concentration ratio between the 2 materials that are deposed the power output to the individual magnetrons is varied.The target concentration is around 5% but anything between 3% and 7% would be acceptable. The accuracy needed for the concentrations is not extremely high since we do not yet know what the optimal concentration is for the interface segregation and observation thereof. Also the XSW method is not very sensitive to the absolute concentration but more sensitive to the spacial relative distribution of the element in the layered structure. To calibrate the system, a single magnetron deposition with both magetrons is done to determine its deposition speed at a certain angle.

Figure 3.5: Steel tube in which the samples are annealed in a vacuum atmosphere.

3.3 Annealing

The annealing setup that is used is a thin tube vacuum chamber as shown in Figure 3.5), which for temperature control in turn is placed in a furnace which is shown in Figure3.8. In the tube a sample holder is placed on which the samples are mounted as shown in Figure 3.7, thermocouples are mounted to backside of the sample-holder to monitor the temperature. After testing it was concluded the insulation of the thermocouples started to break down when raising the temperature of the system above 400 degrees. Since also segregation started to be observed from 300-350 degrees onward the temperature of 400 degrees was established as the annealing temperature. Multiple segregation rounds have been ran for different time scales and the optimum was established for 24 hours. This length of time did not yet lead to structural changes in the waveguide but did show significant segregation for a part of the samples. Longer annealing times led to a measurable breakdown of the waveguide structure but not leading to significantly more segregation.

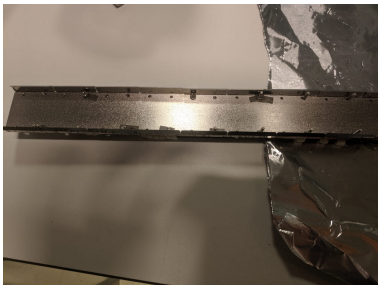


Figure 3.7: Sample holder that is put in the Tube.

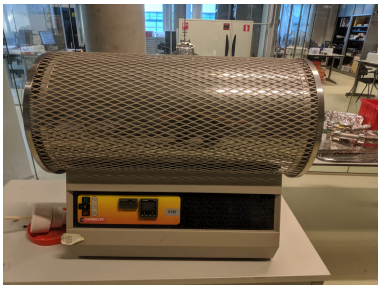


Figure 3.8: Furnace in where the Sample tube is placed to heat up.

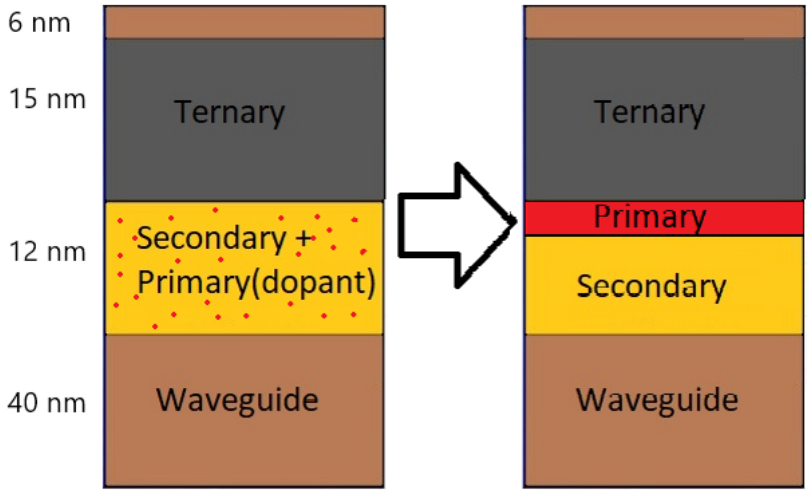


Figure 3.6: An illustration of the Standardized sample used in this research with the materials assigned.

3.4 Metrology

GIXR The GIXR measurements are performed using the Malvern Panalytical Empyrean X-ray diffractometer with a Cu-K α source with a wavelength of 1.5406 . For all measurements a hybrid monochromator was used. The XRR measurements were done using steps sizes of 0.03 for angles less than 1.5° and 0.05 for angles more than 1.5°. The accumulation time that was used was a minimum of 1 second per point and up to 10 seconds to have a minimum statistic of 100 counts(1-10 seconds per point). Before measuring a standard alignment procedure is done.

XSW-Fluorescence To capture the X-ray fluorescence signals, the diffractometer was custom equipped with an Amptek Energy dispersive XR-100SDD Silicon Drift Detector. All measurements were performed with the 4xGe 220 monochromator or hybrid-monochromator. Fluorescence energies between 100 eV and 10 Kev were measured and processed. The excitation and GIXR beam that is used is Cu-k alpha radiation with an energy of 8.047 keV.

To get the angular dependent fluorescence yield, an angular range from 0.2 up to 1 degrees was measured with an angular step of 0.005 degrees To accumulate sufficient number of fluorescent photons the accumulation time is between 1 to 6 minutes per angular step depending on the materials that were used. This difference is due to the different type of atoms shells and the overlapping energies. Capturing significant photons for K lines takes significantly less time than for example L lines; The minimal number of accumulated counts per measurement per angle is 1000 but depending on the overlap with other peaks of comparable energy 10000 or more can be needed to make a good deconvolution and fit. All samples are measured both before and after annealing. The measurements after annealing are done in the same week as the annealing to minimize any non-related effects that might happen after annealing due to the potential structure degradation induced by the annealing.

3.5 Fitting Fluorescence Spectra

An example of a measured fluorescence spectrum from a single angle is shown below in Figure 3.9. The obtained fluorescence spectra are fitted using PyMca. This is a Open-Source software package provided by the ESRF institude in Grenoble [20].

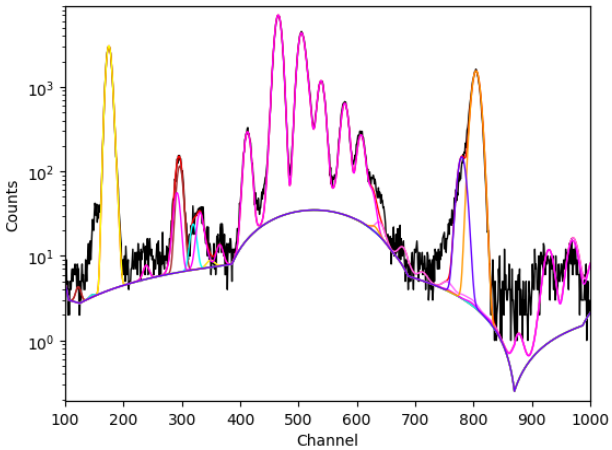


Figure 3.9: A fluorescence spectrum with fit from a sample of a Lanthanum and Silicon multilayer.

To fit the spectra a fitting configuration has been written with details as escape peaks, resolution and other setup parameters to tweak the fitting procedure to the measurement data. When such a configuration has been written for a certain sample, all angular scans measurements can be batch fitted with an automated script. Corrections can be made for secondary excitations, escape peaks, tails and matrix effects which can affect the fundamental ratio's between peaks of the same element.

The spectrum in Figure 3.9 is fitted easily due to the large separation between the peaks and high photon count. In some cases the materials have overlapping fluorescence spectra which can complicate the fitting procedure. In Figure 3.10 an example is given of such a spectrum with fit. In this case more attention has to be taken to properly extract the information of the photons coming from the different periodic elements.

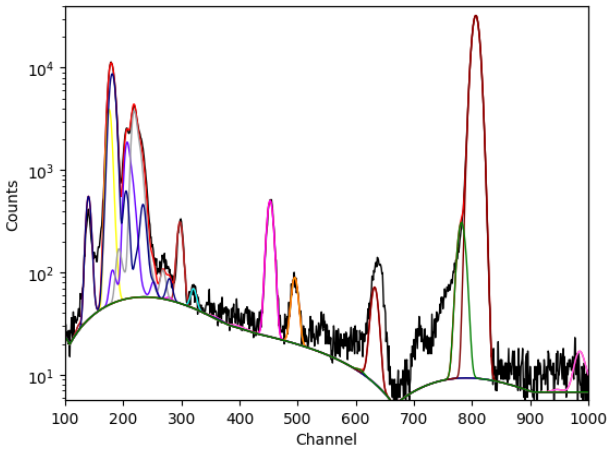


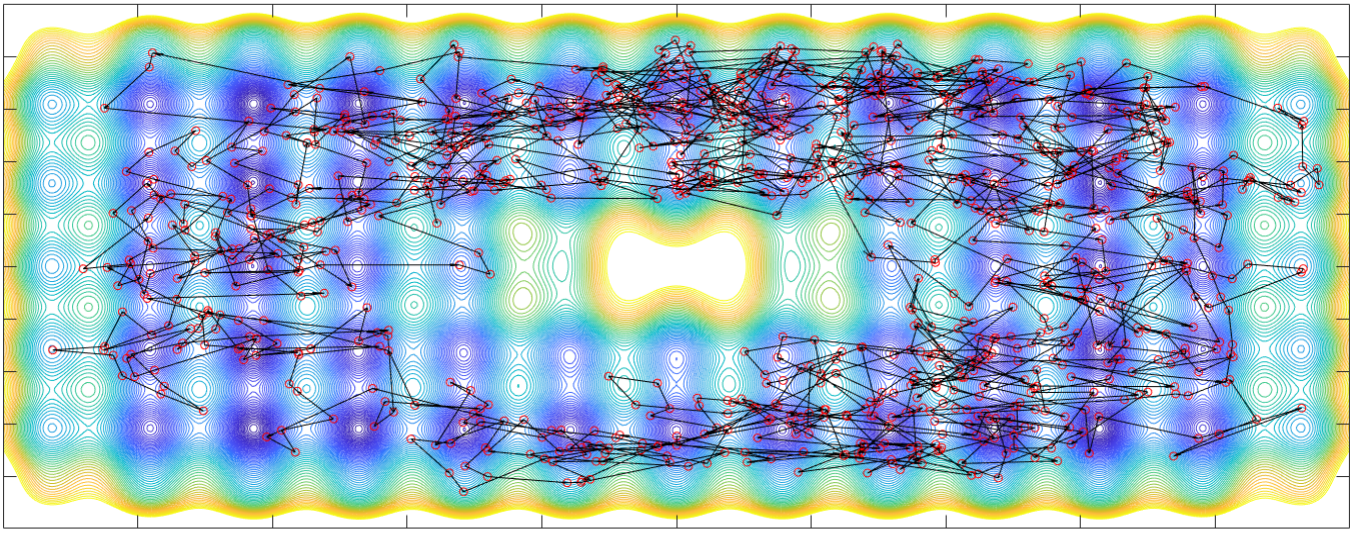
Figure 3.10: A fluorescence spectrum with fit from a sample as deposited as (Si wafer/40nm W/10nm Ti in Zr/15nm Nb/6nm W)

The main challenge is to get a significant signal from the dopant material which is present with a concentration of about 5% in a 10 nm layer. This translates to a layer of half a nanometer of material to detect using these non-destructive methods. For example for a Ti dopant: in the the spectrum in Figure 3.10) the titanium peaks are at 4.5 Kev and 5 Kev (channel 450 and 500) which are properly detectable at 1k counts for the highest peak. Great care has to be applied in choosing materials so that all elements can be properly discerned.

Every angular fluorescence measurement consists of 150+ of these scans for a set of different angles. the measurement procedure and fitting is atomized using matlab, powershell and PyMca since it is too time consuming to do this manually for re-ocurring measurements.

The measurements from the GIXR and XSW are from 2 different measurement systems that have been coupled. Matlab, powershell and C++ programs has been written which is able to process all measurement data automatically by coupling the timestamps to the correct angles and coupling the GIXR and XSW measurements of the same sample. Also a batch measurement feature was developed allowing to measure multiple samples overnight in which the program can isolate the measurements of the different samples.

Using the PyMca command line feature and the timestamps of the measurements all data is automatically ordered and the fluorescence spectra fitted and combined. Missing measurements, which sometimes happens when an error in the accumulation occurs are automatically skipped and ignored. The end result is a GIXR curve and the per element angular XSW curves.



4 Computational Modelling

In this chapter the computational implementation is introduced. First the forward map (loss function) is defined using the Abeles matrix formalism (Section 4.1) to calculate a resulting GIXR and XSW signal from a discretized structure, then the free form parametrization method is introduced (Section 4.2) to map physical structures into these discretized structures that serve as the input for the forward map. With the forward map in Section 4.3 the problem statement is given. In Sections 4.4 and 4.5 respectively the Hamiltonian Monte Carlo and Metropolis-Hastings implementations are discussed and in Section 4.6 the calculation of the derivatives of the loss function used in the HMC are given.

4.1 Formulating the forward map

Sublayers To accurately represent a thinfilm structures in the computer, the thin films that are simulated are discretized into uniform sublayers. Each sublayers discretization width in the z direction follows the resolution criteria proposed in Equation 2.4. Each sublayers properties are uniform thoughout this discrete space and therefore changes only occur at the boundaries of these sublayers. The structure is considered to be uniform in the lateral directions. An illustration of a stack of discretized sublayers representing a thin film is given in Figure 4.1.

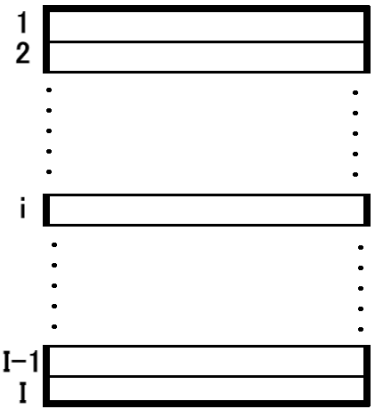


Figure 4.1: Discretized structure with index labels that is used throughout the research to accurately model thin films in the computer.

GIXR Curve Calculation

A resulting reflectively curve from a proposed sublayer structure can be calculated effectively using the abeles matrix formalism [21]. The formalism relies of the multiplication of a set of matrices, each matrix representing the transfer trough a sublayer (refraction and propagation). The matrix representing a single sublayer is shown is Equation 4.1. The variables of this equation are shown in Variables Equation 4.1 and are dependent on the the angle of incidence, the optical constants and the width of the respective sublayers.

$$M_{i,\theta} = \begin{pmatrix} \cos(k_0 \cdot q_{i,\theta} \cdot D_i) & \frac{j}{q_{i,\theta}} \sin(k_0 \cdot q_{i,\theta} \cdot D_i) \\ j \cdot q_{i,\theta} \cdot \sin(k_0 \cdot q_{i,\theta} \cdot D_i) & \cos(k_0 \cdot q_{i,\theta} \cdot D_i) \end{pmatrix} \tag{4.1}$$

Variables Equation 4.1

$$k_0 = \frac{2\pi}{\lambda}$$

$$q_{i,\theta} = \sqrt{\epsilon_i - \epsilon_0 * \cos\theta^2}$$

Where i is the i-th sublayer counted from the top, D_i is the thickness of the i-th sublayer, θ is the angle of incidence, j is the complex unit and ϵ_i is the permitivity of sublayer i calculated as:

$$\epsilon_i = (\delta_i + j\beta_i)^2$$

Where δ_i and β_i are the optical constants of sublayer i.

Equation 4.4

The electric field intensity in sublayer i at angle θ is given by:

$$I_{i,\theta} = \frac{|E_{i,\theta}|^2}{C_\theta} \tag{4.4}$$

Where the 0 index represents the upper part of the air interface and C is given by:

$$C_\theta = \left| \frac{F_{0,\theta} + \sqrt{\epsilon_0}\frac{E_{0,\theta}}{\sin(\theta)}}{2\epsilon_0} \right|$$

And where $\begin{bmatrix} E_{i,\theta} \\ F_{i,\theta} \end{bmatrix}$ is given by:

$$\begin{bmatrix} E_{i,\theta} \\ F_{i,\theta} \end{bmatrix} = M_i * M_{i+1,\theta} ... * M_{N,\theta} * A_N$$

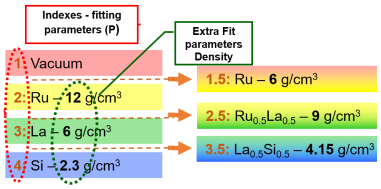


Figure 4.2: A visual explanation of the P vector with material stack of K = 4.

Equation 4.5

Equation expressing the index of refraction in terms of parameter P.

$$n_i = n_A + P_i \bmod 1(n_B - n_A) \tag{4.5}$$

Where n_A is the index of refraction at $\lambda_{Cu-k\alpha}$ for $\lfloor P_i \rfloor$ (floor) and n_B for $\lceil P_i \rceil$ (ceil).
The relation of n_i in terms of optical constants δ and β is given below:

$$n = 1 - \delta + j\beta$$

Where j is the complex unit.

1 - Air
2 - Waveguide
3 - Ternary
4 - 95% Secondary 5% Primary
5 - Waveguide
6 - Structure

Figure 4.3: Parametrization assignment of material sequence of the as as deposited samples with $K = 6$.

Each angle θ and sublayer i with width D_i has a unique assigned matrix, for each angle a string of matrices is calculated to get an expression of the up (E) and down going wave (F) in the structure of that specific angle as shown in Equation 4.2. Where A_n is the unit vector e_1 since at the bottom of the structure it is assumed only a down travelling wave is present.

$$\begin{bmatrix} E_{0,\theta} \\ F_{0,\theta} \end{bmatrix} = M_{1,\theta} * M_{2,\theta} ... M_{i-1,\theta} * M_i * M_{i+1,\theta} ..... * M_{N,\theta} * A_N \tag{4.2}$$

The reflectively constant $r_t heta$ at the top of the structure and the resulting absolute reflectively can be calculated using Equation set 4.3 below:

$$r_\theta = \frac{\sqrt{(\epsilon_0)}}{\sin(\theta)}\frac{F_{0,\theta}}{E_{0,\theta}} \quad \& \quad R_\theta = \frac{1 - r_\theta}{1 + r_\theta} \tag{4.3}$$

Angular Fluorescence Yield Calculations

The angular fluorescence signal from the proposed structure can be calculated by taking the dotproduct of the calculated electric field intensity $I_{i,\theta}$ for each separate angle with the concentration of the respective materials in those sublayers. To get a description of the electric field in the structure at every depth position, the structure is further discretized in set of sublayers of 1 . This is done to take into account a gradual absorption and excitation of the electric field throughout the structure to get an accurate representation of the excitation pattern that will induce the fluorescence radiation that is measured from the structure. The electric field intensity is calculated using Equation 4.4.

The result is a vector with the amplitude for the up traveling wave and down traveling wave at every sublayer. By normalizing the electric field to the outgoing component as seen in Equation 4.4, the electric field intensity is obtained which will be used for the atomic profile reconstruction.

4.2 Free-Form Parametrization

Parameter *P* With the discretization a parametrization of the structure can be formulated in the following way. Assuming each deposited physical layer only mixes with neighbouring layers, each deposited physical layer can be assigned an integer number representing the material it is made of with its respective material density. For example the material stack with an Air($k = 1$) layer on top, following Ruthenium($k = 2$), Lanthanum($k = 3$) on a Silicon($k = 4$) substrate (See Figure 4.2 for an illustration). The structure is divided into a number of sub-layers following the minimum resolution criteria. Each individual sublayer is assigned a number representing the material mix in that sublayer[10]. An Example of parameters P and their material representation are shown below:

- 1 ⟶ 100% Air/Vacuum
- 1.7 ⟶ 30% Air/Vacuum 70% Ruthenium
- 2.4 ⟶ 60% Ruthenium 40% Lanthanum
- 3.3 ⟶ 70% Lanthanum 30% Silicon
- 4 ⟶ 100% Silicon

with the interval of the acceptable parametrization values: $P \in [1, 4]$.

The assigned number P between 1 and K, K being the number of material layers in the parametrization, corresponds to an index of refraction[10] shown in Equation 4.5. Not only does this parametrization greatly reduce the number of parameters in this optimization problem, it is also an efficient way of keeping the ratios between δ and β in line with the materials in the structure and thereby accurately describing the absorbance.

Since the samples that will be analyzed contain compound layers existing of 2 materials we can assign a compound to a integer constant of the parametrization. This will allow for example the mixture more than 2 materials. The parametrization stack used for samples not showing segregation behavior is shown in figure 4.3. For example 3.5 represents a mixture of 50% Ternary, 47.5% Secondary and 2.5% Dopant. Using this parametrization an effective description of the system can be found by optimizing a parametrized profile of discretized sublayers.

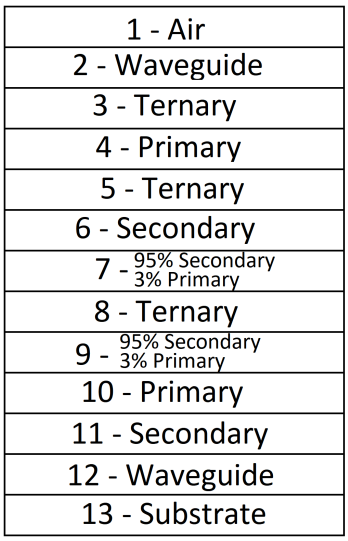


Figure 4.4: Parametrization assignment of the material sequence of the as segregated samples.

Segregation Parametrization The only mixing neighbouring layers condition breaks down in the segregation experiments. (Segregation can occur to non-neighbouring interfaces.) A different initial layer structure has to be developed to allow for this phenomenon. The main parametrization structure that is used to describe the samples that show segregation is shown in Figure 4.4. By using this more complex layered model, every layer and interface is allowed to contain all different material combinations present inside the layer of interest.

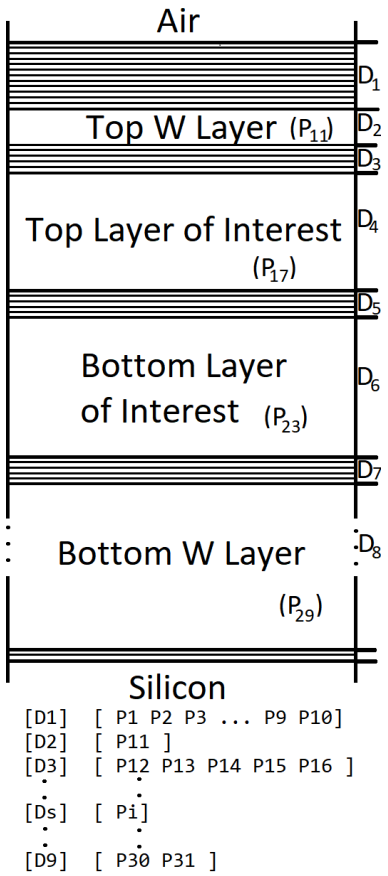


Figure 4.5: Layer/Interface structure that is used to effectively describe layered structures in a free form model. Below is an example of the parametrization matrix. The left column contains the thickness of the segment and the right column contains the parameterization P

Regularization

For the application of optimization and serving as the prior Bayesian inference a regularizing term is defined in the form of a stricter layer/interface parametrization and a punishment term. The system of sublayers is divided up into S segments in which each segment is either a solid thick layer or an interface between solid thick layers and is assumed to have a width of D_s. This regularized system is a stack of solid layers with in between every layer an interface that is modeled to exist of a sequence of equally sized sublayers following the earlier proposed resolution criteria. A structural illustration of the layered system in terms of thicknesses and parameterization $\boldsymbol{P}$ is shown in Figure 4.5 with also an example of a MATLAB cell containing a paramterization of this said structure. Each segment has a thickness(first column) and the assigned parametrization $\boldsymbol{P}$ for each sublayer (second column).

To reduce the number of redundant solutions and to accelerate the process of obtaining solution a regularization term is also included. In previous similar methods Tichinov regularization was used[10]. A punishment term was included proportional to the sum of the total difference squared between the parameters values of the adjacent sublayers. Due to the particular parametrized layered structure a slightly different method was chosen where only the negative finite difference between the sequential sublayers are included in the punishment term for the regulatization. This is done not to punish the large steps needed to describe the multicomponent interfaces present in these parametrizations. The expression for the regularization is stated in 4.7. The weight term in practice has a value proportional to the loss function value of the measurement/theory mismatch. This is to make sure that it does not dominate the optimization term but also is not negligible.

Quantifying segregation

The depth resolution of the XSW signal is in the simulations about half a nano-meter, the same amount of dopant material in a narrower region does not lead to significant changes in the fluorescence signal and we are therefore insensitive to changes on scales smaller than this and therefore also cannot explicitly state the concentrations of dopant in the interface. Also the GIXR and XSW methods are in the regime of low concentrations of dopant material insensitive to the quantitative total amount of dopant present in the whole structure. A measure of segregation has to be made that is invariant to changes in this sub-nanometer regime and to the total quantitative amount of dopant material. The measure that is developed and used is the Segregation Vector (SegVector), which contains in the sequence [AirWaveguideInt(interface), TopWaveguide, TopInt, LOI1, MiddleInt, LOI2, BottomInt, BotWaveguide] the relative fractions of dopant element present in the respective segments of the structure (See Figure 3.1). For the fractions in the Middle Interface & Bottom Interface the concentration present in LOI2 is first subtracted to correct the background signal from the as deposited structure. Quantitatively interfaces are of the order of 0-2nm depending on the proposed parametrization. A fraction of 0.2 of the dopant element in an interface means a considerably higher concentration(5-10x) than a fraction of 0.5 a 12nm layer.

Equations 4.6 & 4.7

Expression of the forward map Q:

$$Q(P, D, \rho, I_0, \theta_R, \theta_F) = \frac{1}{N_1}\sum_{n=1}^{N_1} L(I_0 X(f(P, D, \rho), \theta_R), y_1, \sigma) + \frac{1}{N_2}\sum_{n=1}^{N_2}\sum_{e=1}^{N_e} L(X(g_e(P, D, \rho), \theta_F), y_{2,e}, \sigma) \tag{4.6}$$

Where $L(Y_0, Y_1, \sigma)$ is defined as:

$$L(Y_0, Y_1, \sigma) = \frac{1}{\sigma} \cdot ((Y_0 - Y_1) \circ (Y_0 - Y_1))$$

Where $\circ$ is the element wise multiplication.

$f(\boldsymbol{P}, \boldsymbol{D}, \boldsymbol{\rho})$ is the absolute simulated reflectivity curve $\boldsymbol{R}$ from in Equation 4.3 for all measured angles θ and structure parameters $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{\rho}$.

$g_e(\boldsymbol{P}, \boldsymbol{D}, \boldsymbol{\rho})$ are the normalized to $\max(g_e(\boldsymbol{P}, \boldsymbol{D}, \boldsymbol{\rho}))$ simulated angular XSW fluorescence signals from the individual elements e calculated from the structure parameters $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{\rho}$ according to Equation 4.4.

y_1 and $y_{2,e}$ are the respective GIXR and XSW measurements, θ_R and θ_F are the respective perceived beam divergences of the GIXR and XSW measurements, I_0 is the modelled beam intensity of the X-ray tube and X is a convolution with a Gaussian defied as:

$$X(y, \theta) = (y * \mathcal{N}(0, \theta))$$

N_1 and N_2 the number of measurement points per respective dataset and N_e are the number of elemental fluorescence signals measured and σ is defined as:

$$\sigma = \sqrt{y_0} + (\sigma_s * y_0)^2$$

Where σ_s is the standard error and y_i is the respective intensity of each measurement point.

The regularization term C_R is given by:

$$C_R = W \sum_{n=1}^{N-1} (P_n - P_{n+1})^2 \mathbb{1}(P_{n+1} - P_n) \tag{4.7}$$

Where $\boldsymbol{P}$ is parametrization vector of the structure, W is a predetermined weight of 100 and $\mathbb{1}$ is the indicator function.

4.3 Problem Statement

To reconstruct the measured structure using the GIXR and XSW measurements the forward map defined in Equation 4.6. The forward map is a function of the difference between the measured data and the proposed parametrized structure with parameters $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{\rho}$. The measurement error σ takes into account the general Poisson statistics of incoming photons, Gaussian distributed variance in the beam power and the slight uncertainty of incidence angle determination. We have also introduced the σ_s of 0.002 which is the variance of the beam intensity I_0 during a 1 second measurement. With this σ_s the loss function is defined as $F_{loss} = Q + C_R$ and can be minimized to find a satisfactory parametrization. In practice this minimization reaches in to the range of 50-200 where 50 is considered a near perfect fit (See appendix Figures .1 and .2 for an example).

Additional fitting parameters Not only the parameter P and the thicknesses D_s of the layers and interfaces are free-fitting parameters in this forward map. Also the density of the materials present in the layered system are fitting parameters since it is unknown whether the materials behave identical to their bulk behavior in thin layered system. Also the resolution of the beam and its intensity are used as fitting parameters. The GIXR and XSW measurements although having as source the same beam due to the slits used in the GIXR measurement their perceived resolutions can differ and are therefore separately modelled.

Bayesian Inference

For the estimation of the posterior and its measures the prior and the likelihood function for this problem statement are given by Equations 4.9 and 4.10 respectively. The prior contains the regularization punishment and is independent on any information from measurement y. The likelihood function

contains the forward map based on Equation 4.6. The constants C are empirically determined to take into account the acceptable values of the objective function and regularization term.

Optimization To obtain the parametrization that fits the measurement data the best, x_{MAP}, the most probable parametrization is used. multiple optimization methods are used for minimizing $F_{loss} = Q(P, D, \rho, I_0, \theta_R, \theta_F) + C_R$. (Equivalent to maximizing $\pi(x)\pi(y_0|x)$). A general optimization round is run sequentially until a satisfactory solution has been found. One optimization round is defined a routine starting with a global search algorithm for 30 minutes (on a regular i5 desktop PC) and afterwards this solution is refined by running a short local optimization algorithm.

Bounds To keep the parametrization P within the bounds of $[1, K]$ both in the optimization and in sampling process, a punishment term is used to keep the solutions within the bounds of allowed parametrization values using the expression shown in Equation 4.8

Equation 4.8

$$B = W * \sum_{n=1}^{N-1} \mathbb{1}(P_n < 1) * (1 - P_n)^2 + \mathbb{1}(P_n > K) * (P_n - K)^2 \quad (4.8)$$

Where P_n is parametrization vector of the structure, W is a predetermined weight of 100 and $\mathbb{1}$ is the indicator function.

Local Optimization For the local optimization the BOBYQA algorithm was chosen from the NLopt optimization library [22], this algorithm seeks to minimize the Frobenius norm of the change of the second order derivative of the objective. To implement this algorithm the NLopt library for MATLAB was used[23]. It was chosen since this algorithm showed the best performance in optimizing structures of this kind.

Global Optimization For the Global optimization routine the 'Controlled Random Search (CRS) with local mutation'[24] was used which is a randomized quasi simplex method. Also implemented using the NLopt libary for Matlab.

Confidence Intervals Using Equation 2.13 the symmetric Baysian credibility set can be calculated and is used for the calculation of the confidence bounds for the parameters of interest after a complete posterior estimation has been obtained.

Equations 4.9

x is defined as

$$x = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{D} & \boldsymbol{\rho} & \boldsymbol{I_0} & \boldsymbol{\theta} \end{bmatrix}$$

The prior $\pi(x)$ is given by:

$$\pi(x) = e^{-C_r(x)/C} \quad (4.9)$$

The likelihood function $\pi(y_0|x)$ is given by

$$\pi(y_0|x) = e^{-Q(x)/C} \quad (4.10)$$

Where C is 20 for the combined GIXR & XSW data sets and 10 the case for one of the single datasets. The total Loss function can now be defined as:

$$U(x) = Q(x)C_r(x) \quad (4.11)$$

Gaussian Fitting Procedure

Up to now the dopant distribution is assumed to have a very limited localizations in terms of the distribution. It is assumed to be uniform in the layers and only complete freedom has been given in the interfaces of the thinfilm parametrization. To confirm the correctness of these assumptions, another way of interpreting the fluorescence signal of the dopant material is developed by doing an independent analysis where more freedom is given to the possible distributions. By first fitting the measurements without the dopant XSW signal, a structure can be found which resembles the measured structure, since the effect of the dopant on the index of refraction is negligible and obviously has no effect on the XSW measurements of the other materials.

To interpret the dopant XSW signal a solution single solution is proposed in the form of Gaussian peaks resembling the dopant distribution. These Gaussian peaks can be extended by allowing a flat top peak to better imitate a layer containing a uniform concentration of the dopant. The position, flattop-width and height are chosen as the fitting parameters. A single peak can act as the

distribution in a certain layer by imitating its width and interface. To also allow for the description of more complex segregation multiple peaks are allowed simultaneously where the same fitting parameters are used except for the flattop which is 0 for the extra peaks. The standard deviation is taken to be 5 to resemble roughness of the layers and interfaces. This is kept constant since it does not considerable change the resulting fluorescence signal but over-complicates the analysis of the results.

4.4 Hamiltonian Monte Carlo Implementation

Non-linearity For the mass matrix the general consensus is to take the co-variance matrix of the different parameters. This however due to the non-lineary in the problem at hand does not make much sense since most parameters do not have a consistent co-variance over prior range and the derivative shows large differences even within steps made in the same sample.

Step-Size h The step-size h used in the integration scheme determines the size of the time step taken in the individual steps.

Equation 4.12

Mass Matrix M is defined as:

$$M = \text{diag}\begin{bmatrix} V_P & V_D & V_\rho & V_r \end{bmatrix} \tag{4.12}$$

Where:

$$V_P = \frac{\boldsymbol{D}^{\frac{3}{4}}}{30}$$

$$V_D = \frac{\boldsymbol{D}^{\frac{3}{4}}}{15}$$

$$V_\rho = \frac{\sqrt{\rho}}{3}$$

$$V_r = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Stability For 1 dimensional datasets the stability for the verlet scheme is under the condition that $h < 2$. However for datasets having more dimensions this condition only serves as an absolute upper bound and the stability condition is often much stronger scaling with the number of parameters in the problem and the curvature of the potential energy landscape [18].

Mass-Matrix The Mass-Matrix M acts as the weight of the individual phase space parameter pairs. By taking the naive approach and using the identity matrix the step-size will be limited to the curvature caused by the most influential parameter in the phase space. (changing the thicknesses of the solid layers represent a larger change than the composition.) To optimize the sampling a different weight can be given to parameters to couple the curvature of the individual directions of x to p to make the per step distance travelled in parameter space larger without sacrificing the stability.

Heuristic 4.13

A step is considered stable when the following conditions are met:
Condition 1: $\Delta E_x * \Delta E_p < 0$
Condition 2: $\Delta H < 0.02 H_n$
Where Δ is the difference between the steps n & n+1.

(4.13)

Emphasis Vector All parameters of x in the parametrization are assigned an emphasis-value forming the emphasis vector. These values are empirically determined to maximize the stable step-size in the Monte Carlo sampler and are highly correlated to the average second order derivatives of x in the high-probability regions. These values are used to construct the mass-matrix for this problem. The mass matrix is shown in Equation 4.12.

Stability measure A measure to determine the stability of the integrator is necessary to determine the step-size for the time integration. The Heuristic 4.13 uses the energies calculated in the stepping process and therefore does not lead to a more costly evaluation of the integration. This heuristic proved very effective in practise for determining the stability of the integration process and is used throughout the sampling.

Algorithm 4.14

If Stable: multiply h by 1.1
If Unstable : divide h by $1.1 * S^2$
Where S is the number of successive unstable steps made.

(4.14)

Dynamic step-size The curvature of the energy landscape of the position parameter in the Hamiltonian very different from region to region. Having constant h, the step-size will be limited to the maximum curvature regions in this Hamiltonian. Since a heuristic can be used to determine if a step diverges too far from the actual Hamiltonian a dynamic h can be implemented. Using Algorithm 4.14 for the dynamically sized h the sampling can be drastically sped up since the regions of lower curvature can be explored much faster. Stable values of h have found to be ranging in the interval $[10^{-3}, 10^{-5}]$.

Acceptance rate and Temperature The optimal acceptance rate for the fastest convergence to the target distribution empirically has been found to be around 60-70%[19]. To achieve this a temperature of $T = \frac{1}{\beta} = 2$ has been established in the tempering of the acceptance rate α.

Determining the proper integration time In the case of a 1D convex system, an obvious choice would be to integrate along time until the momentum direction has turned $\frac{\pi}{2}$. Currently in literature is not yet a definite agreement on the optimal integration time for higher dimensional problems. There have been several ideas, like the No-U turn sampler that assumes almost convex energy levels and samples until half on this energy level has been explored. For the problem at hand the energy levels are not convex and therefore it is chosen to use a measure of direction shown in Equation 4.15. S_{DR} gives the ratio of the fraction current step that is in the direction of the sum of directions of the previous steps. When S_{DR} becomes smaller then 0.1 it means that 90% of direction is no longer in the direction of the momentum generation. Therefore this is a good point to propose an end point since from this point intermediate information might be lost when not adding the point to the chain while continuing does not increase the de-correlation more than generating a new set of momentum values. This is a similar approach to the No-U turn sampler criteria, only here the value of S_{DR} can be averaged over 5 steps to counteract the effects of the high frequency curvature and the cut-off value can be tuned to the problem at hand.

Algorithm 4.15

The Direction Ratio S_{DR} is defined as:

$$S_{DR} = \frac{\Delta X_m}{\Delta X_m^*} \tag{4.15}$$

Where ΔX_m is defined as:

$$\Delta X_m = \frac{1}{N} |(x_m - x_0) \cdot (x_m - x_0)| - \frac{1}{N} |(x_{m-1} - x_0) \cdot (x_{m-1} - x_0)|$$

And ΔX_m^* is defined as.

$$\Delta X_m^* = \frac{1}{N} |(x_m - x_{m-1}) \cdot (x_m - x_{m-1})|$$

Where N is the number of parameters in the problem and m is the m-th step in the time integration λ.

Momentum Generation The momentum generation is an important parameter that can be tuned to customize the behavior of the samples in the chain. Since the total energy of the Hamiltonian is constant or very slowly changing due to accumulating truncation errors. The energy of the generated momentum vector gives the user a customization to what possible potential energies are accessible to the chain. In relation to the typical space, this tell the sampler how far away from the high-probability singularities it is supposed to sample. It defines initial kick and so the extra energy that the system has available serves as the maximum that the position parameter can take in terms of its potential energy in that integration time. In the grand scheme, of course no limit exists, since at the end of an integration time the potential energy parameter can be higher than at the beginning and the generation of the new momentum in the next round the momentum energy will be reset again. The momentum sampling is tuned that its momentum parameters have an energy of 50. In practice it was found that higher values lead to considerable time wasted in sampling regions of low probability. Important is that the momentum value is not too low so that a good mixing is achieved for the Markov chain. If the energy of the momentum generation is significantly less than the expected different in energy between the states of interest, it can take a significant number of samples to travel between these states, possibly slowing down the exploration.

4.5 Metropolis-Hastings

A classical Metropolis-Hastings guess and check Monte Carlo Markov chain is be implemented to compare the performance of the Hamiltonian Monte Carlo method to.

Equation 4.14

Proposal distribution used for the Metropolis-Hastings algorithm.

$$\Pi(\widetilde{X}_{n+1}|\widetilde{X}_n) = \mathcal{N}(\widetilde{X}_n, \sigma^2) \tag{4.16}$$

Where σ is defined as:

$$\sigma^2 = \frac{\begin{bmatrix} V_P & V_D & V_\rho & V_r \end{bmatrix}}{1000 * \sqrt{S}}$$

Where S is the number of previously successive declined samples.

Proposal Distribution The proposal distribution $\Pi(\widetilde{X}_{n+1}|\widetilde{X}_n)$ is shown in Equation 4.16. Because of the curvature in the loss functions landscape is highly variable a dynamic σ is chosen since after some trial runs this showed the most consistent performance.

Acceptance For the Metropolis Hastings a temperature of 1 degree is chosen. This is lower than for the HMC since the a proposal in not necessarily made to a direction of similar or higher likelihood. The temperature was determined empirically by raising the temperature until the MCMC starts to sample outside of accepted error margin. The acceptance for the Metropolis-Hastings of sample x_{n+1} from sample x_n is shown in equation 4.17.

Equation 4.17

$$\alpha(x_n, x_{n+1}) = e^{(F_{loss}(x_n) - F_{loss}(x_{n+1}))\frac{1}{T}} \tag{4.17}$$

4.6 Derivatives for the HMC implementation

In the Hamiltonian Monte Carlo sampler the derivatives with respect to the all parameters of the loss function is used to evolve the Hamiltonian over time. For parameters as resolution and beam intensity these derivatives are trivial and are therefore assumed to be known from the loss function definition in Equation 4.6 and are omitted from the text. The calculation of the derivatives of f and g with respect to all indexed parameters are shown below. The vectorized MATLAB implementations are shown in the Appendix in Section .4.

GIXR curve derivatives ($f(P_i, D_s, \rho_k)$)

The gradient of the absolute reflectively (Equation 4.3) with respect to the parameters of interest (P_i,D_s and ρ_k) is computed by using the chain and the product rule.

Derivative with respect to $q_{i,\theta}$ The term $\frac{dR_\theta}{dq_{i,\theta}}$ is calculated first since this term is present in both the derivatives $\frac{dR_\theta}{dP_i}$ (Equation 4.24) and $\frac{dR_\theta}{d\rho_k}$ (Equation 4.27). The derivative $\frac{dR_\theta}{dq_{i,\theta}}$ expressed in Equation 4.18 using partial derivatives of the known expressions from Section 4.1.

Equation 4.18

$$\frac{dR_\theta}{dq_{i,\theta}} = \frac{\partial R_\theta}{\partial r_\theta}\left(\frac{\partial r_\theta}{\partial E_\theta}\frac{\partial E_\theta}{\partial q_{i,\theta}} + \frac{\partial r_\theta}{\partial F_\theta}\frac{\partial F_\theta}{\partial q_{i,\theta}}\right) \tag{4.18}$$

The derived expressions for the partial derivatives with respect to E_θ, F_θ and r_θ of equation 4.18 are shown in Equation 4.19 and derivatives of E_θ and F_θ with respect to $q_{i,\theta}$ are shown in Equation 4.20.

Equation 4.19

$$\frac{\partial r_\theta}{\partial E_\theta} = \frac{-\sqrt{\epsilon_0}}{\sin(\theta)}\frac{F_\theta}{E_\theta^2}$$

$$\frac{\partial r_\theta}{\partial F_\theta} = \frac{-\sqrt{\epsilon_0}}{\sin(\theta)}\frac{1}{E_\theta}$$

$$\frac{\partial R_\theta}{\partial r_\theta} = 2 \cdot \mathbf{Real}\left(\frac{1-r_\theta}{1+r_\theta} \cdot \left(\frac{-2}{(1+r_\theta)^2}\right)^+\right) \tag{4.19}$$

$$\frac{\partial \begin{bmatrix} E_\theta \\ F_\theta \end{bmatrix}}{\partial q_{i,\theta}} = M_{1,\theta} * M_{2,\theta} \ldots M_{i-1,\theta} * \frac{\partial M_{i,\theta}}{\partial q_{i,\theta}} * M_{i+1,\theta} \ldots\ldots * M_{N,\theta} * A_N \tag{4.20}$$

With $\frac{\partial M_i}{\partial q}$ as shown in Equation 4.21 below:

$$\frac{\partial M_i}{\partial q_{i,\theta}} = \begin{pmatrix} -k_0 D_i \sin(k_0 q_{i,\theta} D_i) & \frac{jk_0 D_i}{q_{i,\theta}} \cos(k_0 q_{i,\theta} D_i) - \frac{j \sin(k_0 q_{i,\theta} D_i)}{q_{i,\theta}^2} \\ ik_0 D_i q_{i,\theta} \cos(k_0 q_{i,\theta} D_i) + j \sin(k_0 q_{i,\theta} D_i) & -k_0 D_i \sin(k_0 q_{i,\theta} D_i) \end{pmatrix} \tag{4.21}$$

The structure of this derivative can be exploited to quickly calculate the value of this derivative for every direction $q_{i,\theta}$. By calculating all matrices $M_{i,\theta}$ and its derivative with respect to q, a simple loop can be used to calculate the whole derivative by first calculating the strings $M_{1\to n,\theta} = M_{1,\theta} * M_{2,\theta} ... * M_{n,\theta}$ & $M_{n->N,\theta} = M_{n,\theta} * M_{n+1,\theta} ... * M_{N-1,\theta} * M_{N,\theta}$ sequentially. Now the derivative in each direction can be calculated with Equation 4.23. The cost of computing the derivative in each direction using this methodology is 3 times that of the evaluation of the loss function once. Which is much more efficient in comparison to treating each direction individually which would lead to a quadratic scaling cost with respect to the number of sublayers.

$$\frac{\partial \begin{bmatrix} E_\theta \\ F_\theta \end{bmatrix}}{\partial q_{i,\theta}} = M_{1\to(n-1),\theta} * \frac{\partial M_{i,\theta}}{\partial q_{i,\theta}} * M_{(n+1)\to N,\theta} \tag{4.23}$$

Derivative with respect to P_i The derivative $\frac{dR_\theta}{dP_i}$ expressed in partial derivatives of the known expressions from Section 4.1 is shown in Equation 4.24.

$\frac{\partial q_{i,\theta}}{\partial P_i}$ is evaluated using a finite difference method (Equation 4.22) since the finite difference evaluation is accurate and fast so analytically deriving this does not yield any considerable benefit in term of accuracy or speed. On the integer values of P_i the derivative shows a discontinuity, in practice however the value of P_i is never exactly integer therefore this has not posed any issues.

Derivative with respect to density ρ_k The derivative $\frac{dR_\theta}{d\rho_k}$ expressed in partial derivatives of the known expressions from Section 4.1 is shown in Equation 4.25. The first 5 partial derivatives in this expression are the same expression as the ones calculated for Equation 4.24. $\frac{\partial q_{i,\theta}}{\partial P_i}$ is for the same reasons as before evaluated using a finite difference method and is shown in (Equation 4.26).

A change in the density of a material A assigned a number k changes all sublayers assigned a parameter P in the interval $(k-1, k+1)$. Therefore all the separate derivatives towards q in each affected sublayer have to be summed to obtain the derivative with respect to the change in ρ_k.

Derivative with respect to thickness D_s The derivative $\frac{dR_\theta}{dD_s}$ expressed in partial derivatives of the known expressions from Section 4.1 is shown in Equation 4.27. The expressions of the partial derivatives with respect to E_θ, F_θ and r_θ of equation 4.18 are shown in Equation 4.19 and derivatives of E and F with respect to D_s are shown in Equation 4.28.

$$\frac{\partial \begin{bmatrix} E_\theta \\ F_\theta \end{bmatrix}}{\partial D_i} = M_{1,\theta} * M_2 ... M_{i-1,\theta} * \frac{\partial M_{i,\theta}}{\partial D_i} * M_{i+1,\theta} ..... * M_{N,\theta} * A_N \tag{4.28}$$

Equation 4.22

$$\frac{\partial q_{i,\theta}}{\partial P_i} = \frac{q_{i,\theta}(P_i, \rho_k) + q_{i,\theta}(P_i + dP, \rho_k)}{dP_i} \tag{4.22}$$

Where dP is the finite difference step size

Equation 4.24

$$\frac{dR_\theta}{dP_i} = \frac{dR_\theta}{dq_{i,\theta}} \frac{\partial q_{i,\theta}}{\partial P_i} \tag{4.24}$$

Equation 4.25

$$\frac{dR_\theta}{d\rho_k} \sum_{\forall i | P_i \in (k-1, k+1)} \frac{dR_\theta}{dq_{i,\theta}} \frac{\partial q_{i,\theta}}{\partial \rho_k} \tag{4.25}$$

In which the sum runs over all sublayers that have a value of P that is affected by a change in density of material k.

Equation 4.26

$$\frac{\partial q_{i,\theta}}{\partial \rho_k} = \frac{q_{i,\theta}(P_i, \rho_k) + q(P_i, \rho_k + d\rho))}{d\rho} \tag{4.26}$$

Where $d\rho$ is the finite difference step size.

Equation 4.27

$$\frac{dR_\theta}{dD_s} = \sum_{i \in s} \frac{dR_\theta}{dD_i} \tag{4.27}$$

Where D_i is the thickness of the sublayer i which has a thickness of D_S divided by the number of sublayers present in segment s.

$\frac{dR_\theta}{dD_i}$ is given by:

$$\frac{dR_\theta}{dD_i} = \frac{\partial R_\theta}{\partial r_\theta} \left(\frac{\partial r_\theta}{\partial E_\theta} \frac{\partial E_\theta}{\partial D_i} + \frac{\partial r_\theta}{\partial F_\theta} \frac{\partial F_\theta}{\partial D_i} \right)$$

With matrix system $\frac{\partial M_i}{\partial D}$ is shown in 4.29

$$\frac{\partial M_i}{\partial D_i} = \begin{pmatrix} -k_0 \cdot q_{i,\theta} \cdot \sin(k_0 \cdot q_{i,\theta} \cdot D_i) & k_0 \cdot j \cdot \cos(k_0 \cdot q_{i,\theta} \cdot D_i) \\ k_0 \cdot j \cdot q^2 \cdot \cos(k_0 \cdot q_{i,\theta} \cdot D_i) & -k_0 \cdot q_{i,\theta} \cdot \sin(k_0 \cdot q_{i,\theta} \cdot D_i) \end{pmatrix} \quad (4.29)$$

The same trick as in Equation 4.23 that be applied here to reuse the calculated strings of matrices.

Since some segments are compromised of multiple lamella, all separate derivatives towards these lamella have to be summed to obtain the derivative towards the change of length of this segment.

XSW signal derivatives ($g_e(P_i, D_s, \rho_k)$)

In constrast to the reflectivity where only the absolute value at the top of the structure is of interest, the XSW calculations require the knowledge of the absolute value of the derivate of the electric field intensity from Equation 4.4 at every depth point i. Due to the nature of the fine discretization of the XSW calculations, a finite difference approach will lead to faster evaluation speed of the derivative. This is because to before the calculations of the electric field intensity are done, the sublayers are locked onto a finer grid of 1 Å sublayers to get a better approximation of the resulting XSW signal that is coming from the sample by better modelling the absorption. Due to this refinement process, calculating the analytical derivative would lead to a longer calculation time since for each refined sublayer a separate string of calculations has to be done to calculate the analytical expression of the partial derivative towards each individual sublayer in this new refined grid of sublayers. Not each of these separate derivatives is needed since one parameter P_i and D_i now maps to multiple sublayers. By using finite differences only one string of matrices has to be perturbed per parameter P_i and D_i, instead of each individual sublayer individually.

Since the wavefield calculation is calculated from bottom to top, the calculation of the derivative only differs from the unperturbed state starting from the lowest sublayer that is affected by the perturbation. Therefore for sublayers beneath the finite difference pertubation, the unpertubed values can be used which can later be re-normalized. In order speed up the process first the unperturbed wavefield is calculated as shown in Algorithm 1.

Algorithm 1: Wavefield

$M_{n,1:\theta} \leftarrow \forall n, \theta$ from Equation (4.1);
$A_{N,1:\theta} \leftarrow \text{repmat}([1,0], N_\theta)'$;
for $n \leftarrow N$ **to** 1 **do**
    $A_n[1] \leftarrow M_{n,1:\theta}[1,1] \cdot A_{n+1}[1] + M_{n,1:\theta}[1,2] \cdot A_{n+1}[2]$;
    $A_n[2] \leftarrow M_{n,1:\theta}[2,1] \cdot A_{n+1}[1] + M_{n,1:\theta}[2,2] \cdot A_{n+1}[2]$;
$E^0_{1:N,1:\theta} \leftarrow |A[1]_{1:N,1:\theta}|^2$;
$Z_\theta \leftarrow |\left(A(1)_{1,\theta} + \sqrt{\epsilon_0}\frac{A(2)_{1,\theta}}{\sin\theta}\right)\frac{1}{\epsilon_0}|$;
$W_{n,\theta} \leftarrow E_{1,1:n}./Z_\theta$;

Using Algorithm 2 the derivatives in the directions of $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{\rho}$ can be calculated by substituting any of the parameters for K. The MATLAB implementation can be seen in the Appendix in Section .4.

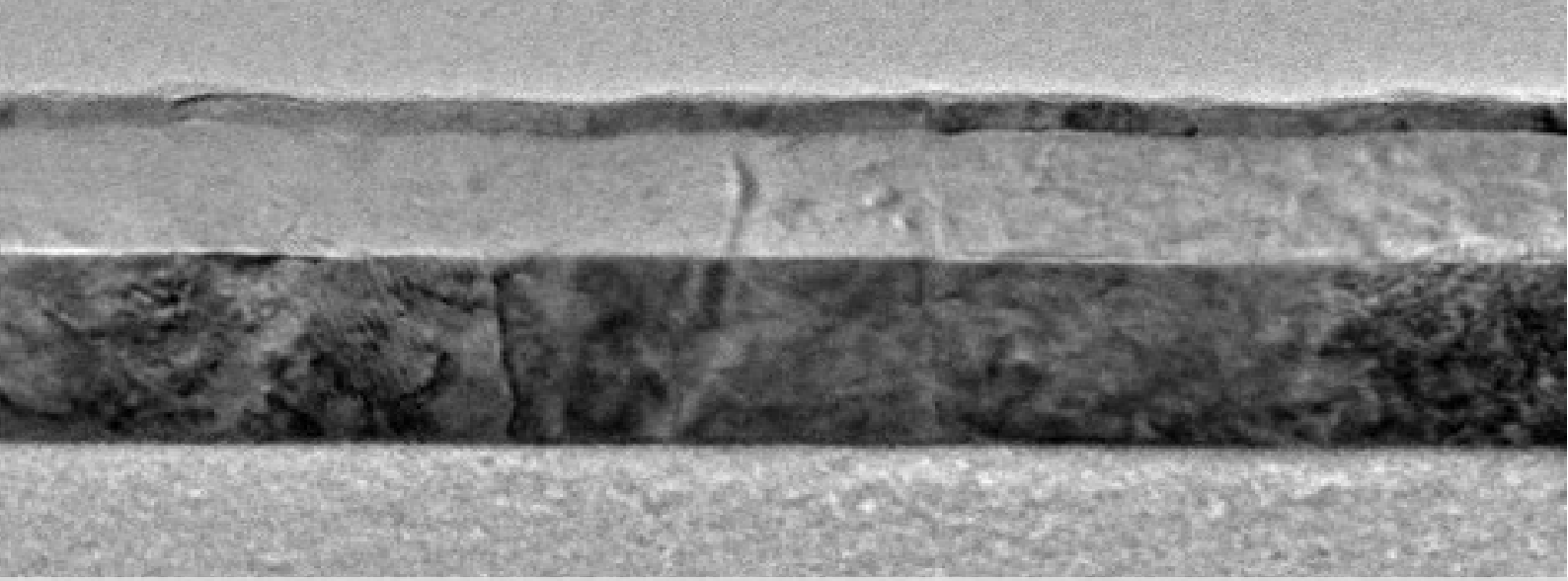
Algorithm 2: Wavefield Derivatives

for $k \leftarrow 1$ **to** K **do**
  $\widetilde{M}_{n,1:\theta} \leftarrow \forall n, \theta$ from (4.1) with perturbation dk;
  $\widetilde{A}_{N,1:\theta} \leftarrow \text{repmat}([1,0], N_\theta)$;
  $N_* \leftarrow \max(n) | n \in \text{perturbation}$;
  for $n \leftarrow N_*$ **to** 1 **do**
    if $n \in$ *perturbation* **then**
      $\widetilde{A}_n[1] \leftarrow \widetilde{M}_{n,1:\theta}[1,1] \cdot \widetilde{A}_{n+1}[1] + \widetilde{M}_{n,1:\theta}[1,2] \cdot \widetilde{A}_{n+1}[2]$;
      $\widetilde{A}_n[2] \leftarrow \widetilde{M}_{n,1:\theta}[2,1] \cdot \widetilde{A}_{n+1}[1] + \widetilde{M}_{n,1:\theta}[2,2] \cdot \widetilde{A}_{n+1}[2]$;
    else
      $\widetilde{A}_n[1] \leftarrow M_{n,1:\theta}[1,1] \cdot \widetilde{A}_{n+1}[1] + M_{n,1:\theta}[1,2] \cdot \widetilde{A}_{n+1}[2]$;
      $\widetilde{A}_n[2] \leftarrow M_{n,1:\theta}[2,1] \cdot \widetilde{A}_{n+1}[1] + M_{n,1:\theta}[2,2] \cdot \widetilde{A}_{n+1}[2]$;
  $\widetilde{E}_{1:N,1:\theta} \leftarrow [|\widetilde{A}[1]_{1:\widetilde{N},1:\theta}|^2; E_{\widetilde{N}+1:N,1:\theta}]$;
  $\widetilde{Z}_\theta \leftarrow |\left(\widetilde{A}[1]_{1,1:\theta} + \sqrt{\epsilon_0}\frac{\widetilde{A}[2]_{1,1:\theta}}{\sin\theta}\right)\frac{1}{\epsilon_0}|$;
  $\widetilde{W}_{1:N,1:\theta,k} \leftarrow \widetilde{E}_{1:N,1:\theta} \cdot / Z^o_{1:\theta}$;

Now one has to follow again the summation procedure from the GIXR derivatives to sum the derivatives of the individual perturbations that relates to one of the parameters.

With the perturbed wavefield intensity $\widetilde{W}_{1:N,1:\theta,k}$ the perturbed XSW signal can be calculated by multiplying with the perturbed atomic distribution of the respective materials.

Using the vectorized and optimized derivatives, the evaluation of the gradient is in the order of a tenth of a second for the samples that were deposited for this thesis. Although this is about 5x the cost of a normal forward iteration of the forward map, the dimension of x is >50 meaning a performance improvement of about 1 order of magnitude has been made by optimizing the derivatives.



5 Results

In Section 5.1 the thermal stability of the waveguide is presented and the reconstruction results are discussed. In Section 5.2 the discerning power of the XSW signal from the different interfaces is presented. In Section 5.3 the stability and applicability of the HMC methodology to the thin-film reconstructions is discussed. In section 5.4 the verification of the results with a TEM study is presented. In section 5.5 the efficacy of the addition of the XSW data is shown. In Section 5.6 the comparison between the different MCMC methods is presented. The chapter ends with Section 5.7 in which the segregation behavior is shown to follow some of Miedema's predictions.

5.1 Waveguide Performance

Thermal stability The thermal stability and the data-analysis aspects of the waveguide sample design described in section 3.1 were tested on the waveguide with the Vanadium dopant, Scandium solvent and Niobium Ternary layer. (denoted as V/Sc/Nb system). This waveguide structure is coated and annealed for 24 hours at 400 °C. The reconstruction of GIXR & XSW data collected from the pristine and annealed samples are shown in Figure 5.2. (reconstructions obtained according to x_{MAP} from Equation 2.11).

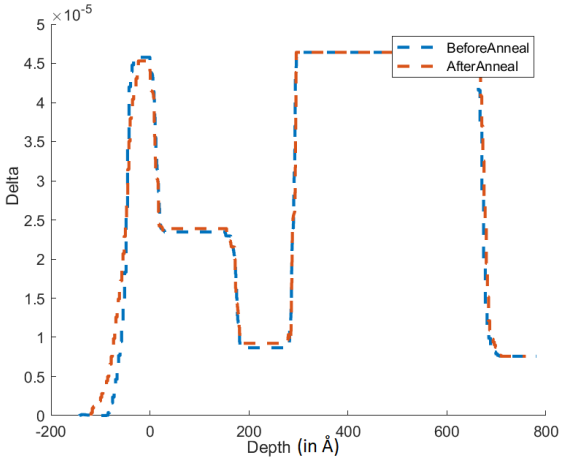


Figure 5.2: Optical constant profiles before and after annealing $@Cu - k\alpha$ reconstructed from the GIXR & XSW measurements on the V/Sc/Nb Sample.

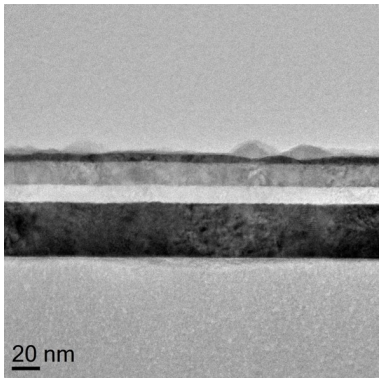


Figure 5.1: Close up TEM image of the V/Sc/Nb Sample

These reconstructions show very minor structural changes in the optical constant profile. The top layer of the waveguide seems to be minimally oxidized and therefore does not compromise the function of the waveguide.

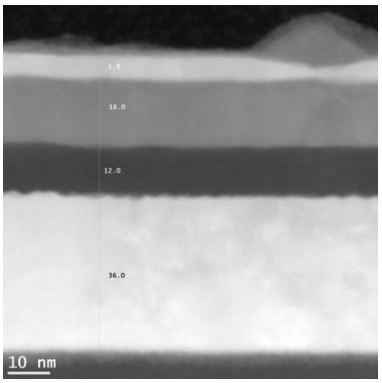


Figure 5.3: Spatial HAADF spectrum accumulated from the V/Sc/Nb Sample

The TEM and HAADF images of this V/Sc/Nb sample after annealing can be seen in Figures 5.1 & 5.3. This data is in good agreement with GIXR and XSW analysis. The layers are properly grown and the interfaces are sharp (1.5 <nm). Some oxidation is visible on the top. This oxidation is likely the result of aging after the annealing process since these TEM images were taken 4 months after depositions and these oxidation hills did not shown up in the GIXR & XSW analysis.

Reconstruction For all measured samples reconstructions parametrized according to section 4.2 resulting from the combined GIXR and XSW analysis are found. The reconstructions are simultaneously in good agreement with both respective measurement sets. The GIXR fit of the annealed V/Sc/Nb sample reconstruction is shown in Figure 5.4. The fitting process of the annealed sample yields almost exactly the as deposited thin film system and only minimal roughness is observed on the top layer.

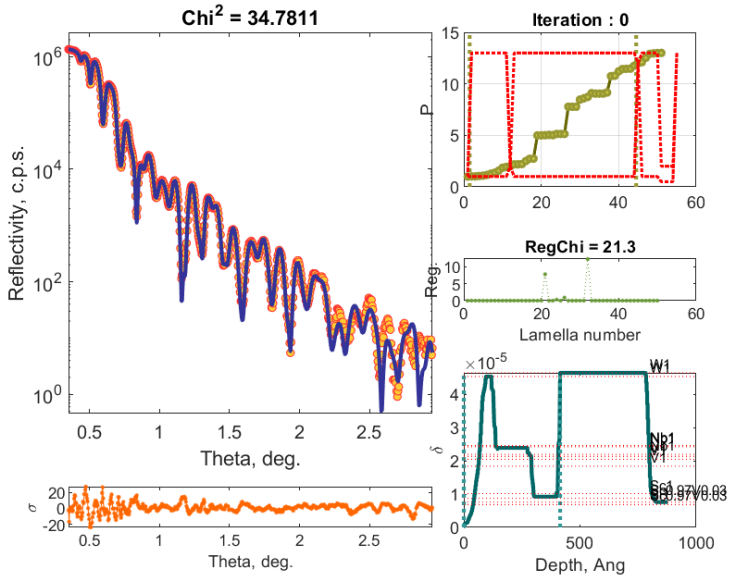


Figure 5.4: GIXR measurement fit of the the annealed V/Sc/Nb Sample.

The fitting quality of the XSW signal can be seen in Figure 5.5 including the resulting atomic profiles. The increase of the concentration of V on interfaces between the Scandium layer and Tungsten layer (depth - 400 Andstrom in Figure 5.5) and the interface between Sc and Nb (depth - 300 Angstroms in Figure 5.5) is the indication of preferential segregation of Vanadium to these interfaces.

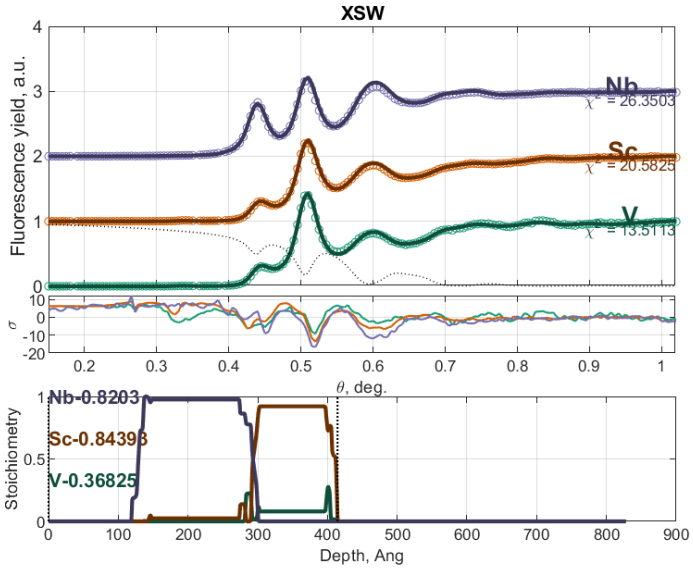


Figure 5.5: XSW measurement fit of the the annealed V/Sc/Nb Sample for the 3 different layer elements. On the bottom in the different color the respective atomic profiles are shown.

5.2 Sensitivity to the different interfaces

To demonstrate the sensitivity of XSW technique combined with the waveguide design to the 3 potential segregation sites (Top,Middle,Bottom), a set of samples with Chromium as dopant, Iron the matrix material and Co as ternary material was deposited and analyzed for each respective segregation site.

In Figure 5.6 on the the left side, the measured angular-dependent XSW signal for each of the samples is shown and compared. Each different deposition location expresses a unique signature in the XSW measurements. The signals are not interchangeable and cannot be linearly superimposed to obtain the other respective signals.

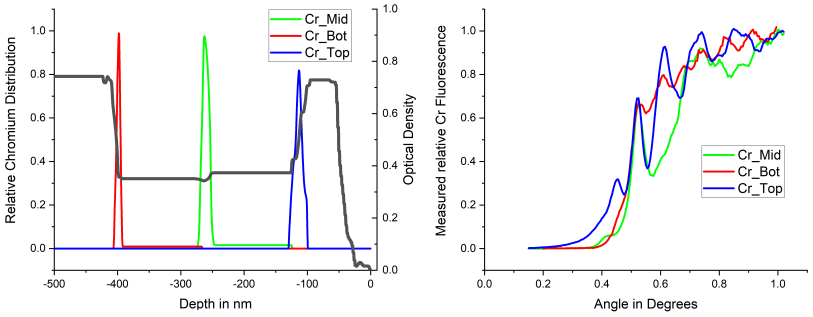


Figure 5.6: Comparison of the measured Cr fluorescence signal of the 3 samples where Cr has been pre-depositioned to the designated interfaces. On the left the resulting reconstructed Cr profiles are present and on the right side the respective measured angular fluorescence signals are shown.

Comparison of different signals After the reconstruction of the XSW & GIXR measurements the Chromium distribution are found in the respective interfaces to the designed positions. The reconstructed Chromium depth profiles resulting from the XSW measurements are shown on the right side in Figure 5.6. All profile reconstructions are almost identical to the as deposited structure indicating a strong sensitivity to the the different interfaces. The

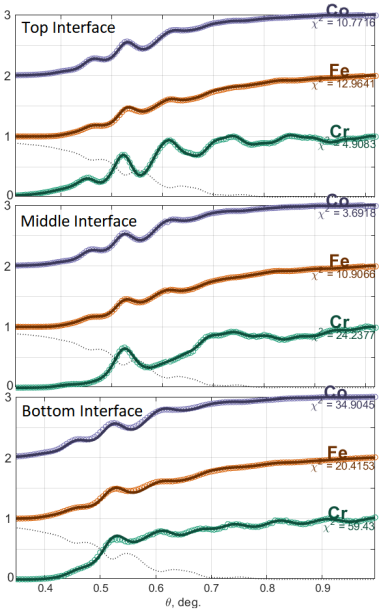


Figure 5.7: Measured and simulated signal of the 3 different samples and their respective reconstructions

agreement between the simulated signals and the measured signals is shown in Figure 5.7.

The measured GIXR curves are shown on the right side in Figure 5.8 with the respective reconstructions of the optical constant profile δ on the left side. The reconstructions of the measured samples are optically nearly identical, which is expected from the very similar measured GIXR curves. Implying that the difference in the fluorescence signal is clearly a consequence of the different Chromium distributions in the waveguide structure.

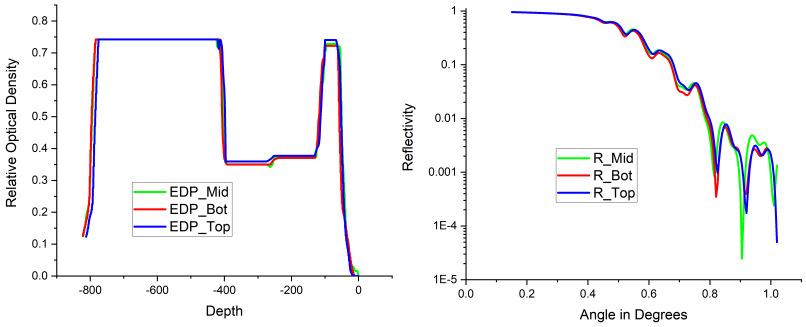


Figure 5.8: On the left the reconstructed δ profiles are shown for all samples and on the right side the respective measured GIXR curves for each test sample are shown.

5.3 HMC Implementation

Complementary to the X_{map} value obtained from the optimization procedure, the HMC sampler discussed in Section 4.4 is used to obtain a chain of samples with fitting parametrizations invariant to the distribution distribution by the likelihood of loss function discussed in Section 4.3. The chain of samples is used to obtain confidence bounds on the individual parametrization parameters.

Table 5.1: Technical details of the HMC implementation, for the details on each parameter refer to Section 4.4

Parameter	Average
h	0.0006
ΔX_{step}	0.0015
t_{step}	$\frac{1}{3}$ sec
SCS	90
ΔX_{Sample}	0.05
t_{sample}	30 sec

To test the performance of the HMC sampler, again the V/Sc/Nb structure is analysed. This is because of this structure also a Transmission Electron Microscopy measurement is done with which the results can be verified. This is discussed in the next section. The HMC chain quickly converges to a equilibrium state from a point drawn from the prior as can be seen in Figure 5.29. In equilibrium the HMC is it is sampling at a loss function value of in the range of the likelihood of the X_{map} value and likelihood indicating a efficient exploration of the typical space. A snapshot of the loss function value of the sampled chain over time in equilibrium is shown in Figure 5.9.

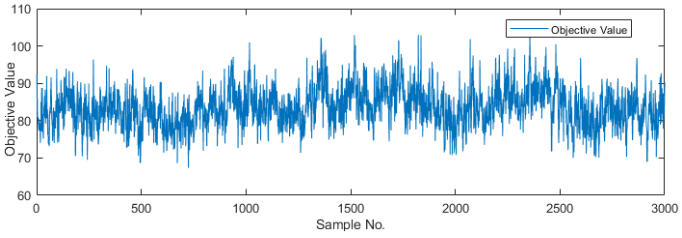


Figure 5.9: Objective value of the loss function for an interval of 3000+ samples in equilibrium sampling.

The truncation error in stable steps averages at a less than half of a percent of the potential energy (loss function) value indicating a stable integration of the equations of motion. For a temperature of $T = 2$ the optimal acceptance rate of

70% is achieved. In Table 5.1 the average value of the the technical parameters are shown. ΔX_{step} is the average distance travelled in in parameter space per step. t_{step} is the time is takes per step. SCS are the number of successive steps made in a correlated direction (steps per sample). ΔX_{Sample} is the average distance between each successive sample and t_{sample} is the time it takes to sample and accept such a sample.

In practice proving the convergence of a MCMC in finite time is not a tractable computation, hence the use of the Markov Chain. To obtain some measure of convergence, the HMC chain has been run from 3 different random initial conditions. In Figure 5.10 the average relative width of the confidence intervals in the stationary regime for different samples sizes averaged over different regions in the sampled markov chain for the different important parameters is shown. For the calculation the average relative confidence interval width for the different layer and interfaces. From an average of about 1500 samples all the chains seems to have obtained the same confidence intervals in comparison to the complete set of samples of all 3 chains that have sampled for 5000+ sample. This indicates good mixing is achieved and a good posterior estimation is made after 1500 samples.

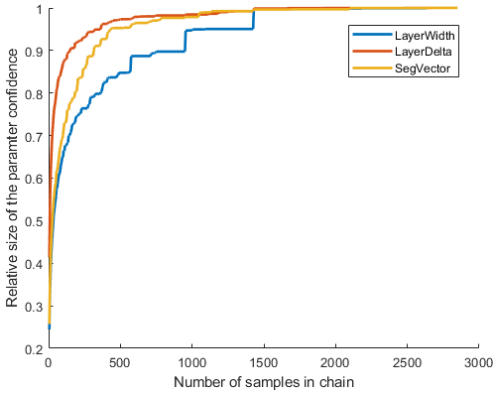


Figure 5.10: Average relative width of the 95% confidence interval in comparison to the complete chain confidence interval width for the different parameters.

As an example of the posterior, the resulting parameters from the analysis for the segregation are shown in a pyramid posterior plot in Figure 5.11. Using Equation 2.13 the confidence intervals of the individual parameters can be calculated and are shown in the diagonal entries. The off-diagonal entries show the inter-parameter correlations between pairs of parameters.

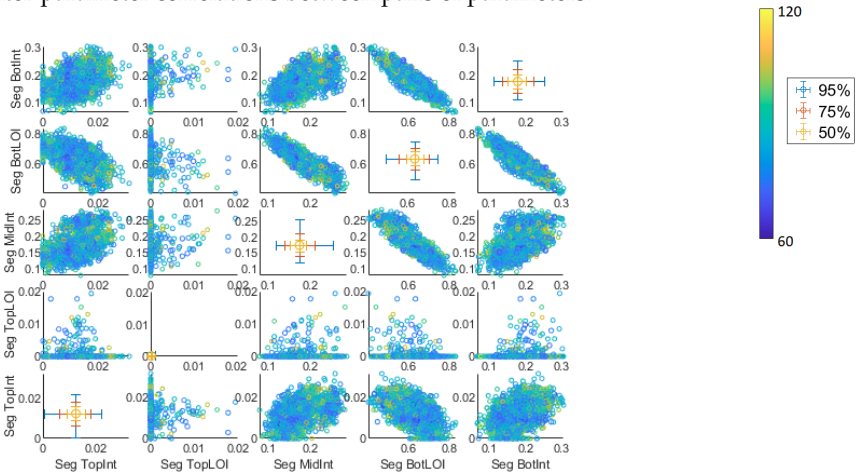


Figure 5.11: Pyramid plot showing the segregation vector for the individual samples. Legend in shown on the side with the color indicating the individual sample loss function values.

5.4 TEM comparison

One of the main challenges of the XSW-GIXR method is not only to reconstruct a structure that fits the measurement data but also to ensure that the reconstruction accurately represents the actual measured structure instead of a structure that yields a similar measurement. To address this, 2 structures which showed complex dopant distributions in its reconstruction have been characterized using TEM,EDX and EELS and thereby will be used as an experimental based study of the reliablity of GIXR-XSW for the designed waveguide samples.

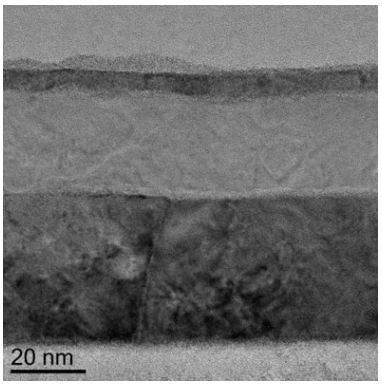


Figure 5.12: Close up TEM image of the Cr/Fe/Co Sample

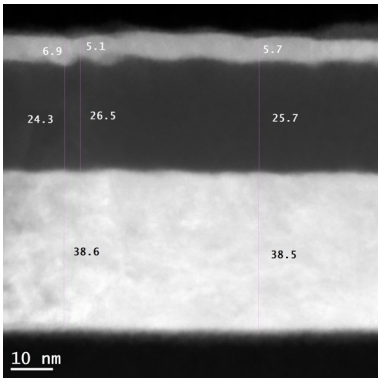


Figure 5.13: Spatial HAADF spectrum accumulated from the Cr/Fe/Co Sample

Sample Cr/Fe/Co

The Cr/Fe/Co sample was chosen since it showed very explicit segregation behavior after annealing. Figure 5.14 shows that the initial qualitative XSW signals for the dopant and the matrix layers before annealing are overlaping and differ greatly after annealing. From this was concluded that the co-deposition was indeed successful in making dopant and matrix atoms distributions overlap and having the dopant segregate after annealing.

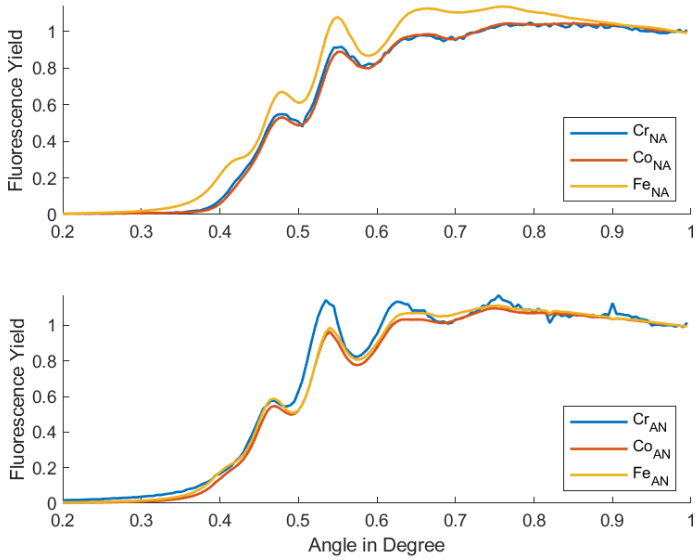


Figure 5.14: Above the measured angular fluorescence signals are shown from the Cr/Fe/Co sample before annealing, on the bottom the measurements after annealing are shown.

In Figures 5.15 , 5.16 and 5.17 a comparison of the reconstruction of the atomic distribution for Chromium, Iron and Cobalt respectively for the different measurement methods is shown. The reconstruction of the Chromium distribution with the GIXR & XSW analysis is in agreement with the EDX and EELS distributions. The small (1%) concentration of Chromium visible between -300 and -200 detected by the XSW method is below the sensitivity of the EDS and the EELS. Later we will discuss the significance of this low concentration "tail" for XSW data. The HMC sampling analysis in Table 5.2 shows that no solutions exist for which the bottom layer contains no significant amount of Chromium concluding that about 30% of the dopant is left in the bottom layer. Also the resulting relative portions of chromium in the interfaces obtained from the different measurement methods is consistent.

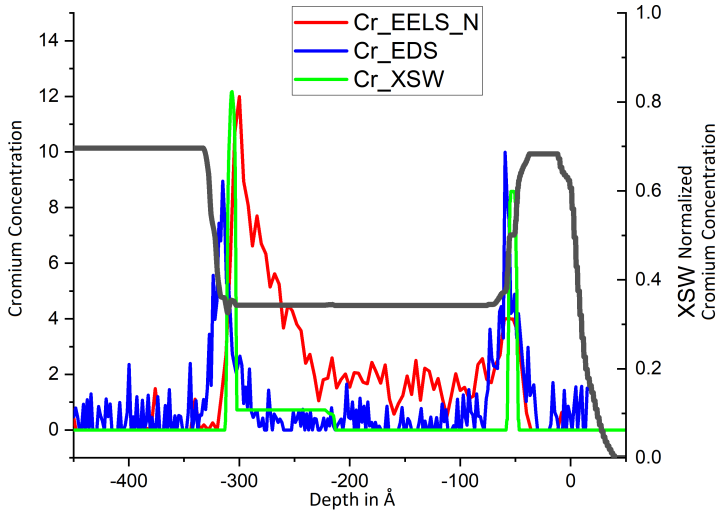


Figure 5.15: Comparison of the different Cr distributions found using XSW, EDS and EELS.

Table 5.2: Segregation vector from the HMC sampler on the Cr segregation.

Layer	Mean	Best	95% CI
TopInt	0.347	0.357	0.389 - 0.307
TopLOI	0	0	0 - 0
MidInt	0.002	0	0.021 - 0
BotLOI	0.297	0.305	0.395 - 0.177
BotInt	0.349	0.337	0.426 - 0.290

The intermixing of the Iron and Cobalt layers, though unexpected, confirmed by all measuring methods. Also a gradient in this mixture is perceived by all measurement reconstructions. There is small difference in the continuity of the distribution in the XSW analysis resulting from the limited forms the distribution can take due to the layer constraints and regularization. Table 5.3 contains the inferred confidence intervals from the HMC sampler of the layer purity of the as deposited materials for the 2 inner layers, also indicating that no probable solutions exist without the intermixing of LOI1 and LOI2.

Table 5.3: Layer purity after annealing of deposited material, results from the HMC sampler on the Cr Sample.

Layer	Mean	Best	95% CI
TopLOI	0.587	0.652	0.711 - 0.455
BotLOI	0.569	0.506	0.722 - 0.387

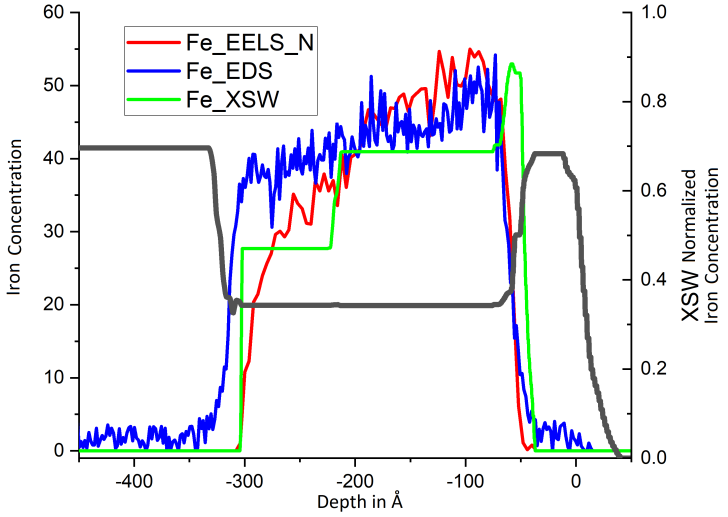


Figure 5.16: Comparison of the different Fe distributions found using XSW, EDS and EELS.

Also the gradient of the Cobalt distribution opposite to the Iron is perceived by the GIXR-XSW reconstruction.

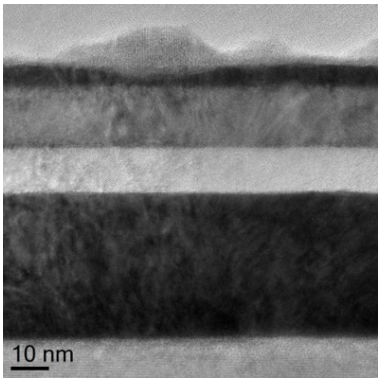


Figure 5.18: Close up TEM image of the V/Sc/Nb Sample

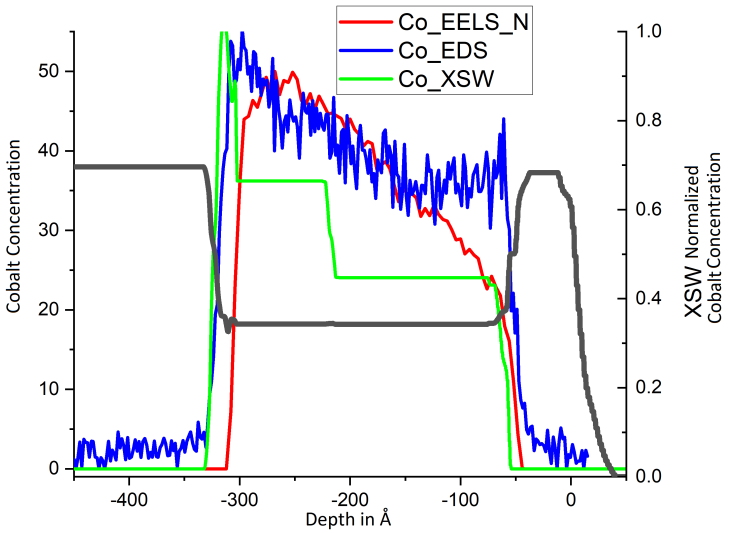


Figure 5.17: Comparison of the different Co distributions found using XSW, EDS and EELS.

Sample V/Sc/Nb

For the same reasons as with the Cr/Fe/Co sample the V/Sc/Nb was chosen to be analyzed to confirm the reconstruction of the XSW & GIXR measurements. The measured XSW signals of V/Sc/Nb are shown in Figure 5.19.

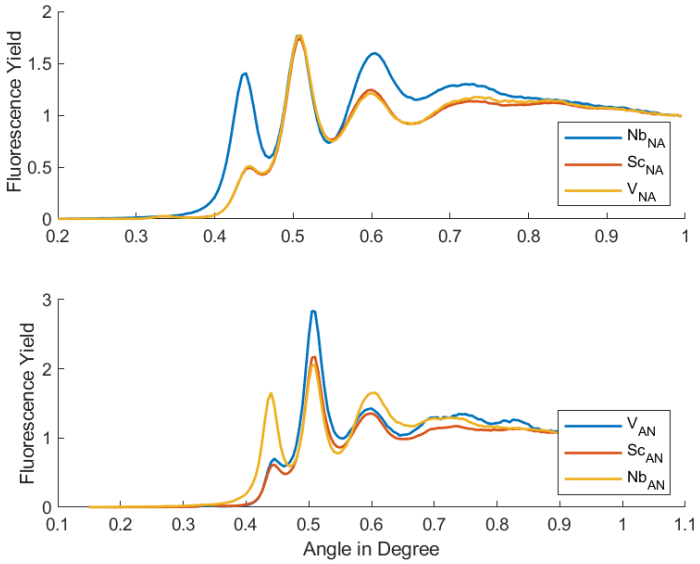


Figure 5.19: Measured XSW signals overlapped for different elements before and after of the V/Sc/Nb Sample

In Figures 5.20 , 5.21 & 5.22 the comparison of the reconstruction of the different atomic distributions for the different measurement methods are shown respectively for Vanadium, Scandium and Niobium. Again for the dopant material, Vanadium, the fractions and the location of the dopant present in the interfaces is in agreement for all measurement reconstructions. The dopant material left in the bottom layer is also confirmed by all methods. Also HMC sampling analysis for which the segregation results are presented in Table 5.4 shows that no solutions exist for different dopant locations. There is however

a slight uncertainty in the relative fractions of dopant material in the different segments in the reconstruction.

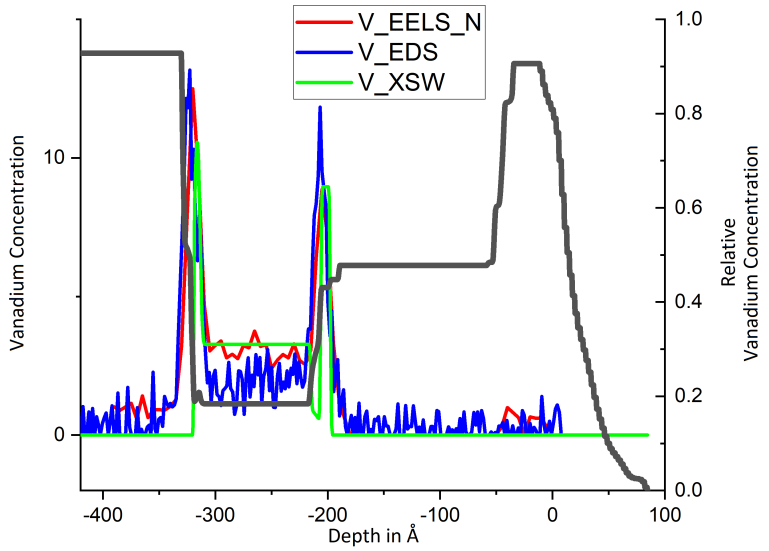


Figure 5.20: Comparison of the different V distributions found using XSW, EDS and EELS.

Table 5.4: Segregation results from the HMC sampler on the V segregation.

Layer	Mean	Best	95% CI
TopInt	0.013	0.014	0.022-0
TopLOI	0.000	0	0-0
MidInt	0.175	0.148	0.255-0.119
BotLOI	0.634	0.660	0.756-0.946
BotInt	0.178	0.177	0.253-0.113

The Scandium distribution is isolated to the location where it was originally deposited. The Scandium presence in the waveguide of the EELS is an artifact of the overlapping peaks in analysis, this is further explained in the Appendix .2. In Table 5.5 the resulting layer purity's with respect to the as deposited structure from the HMC sampler are shown. No solutions exists for intermixed systems and only a slight uncertainty of 2.4% is present in the top layer of interest.

Table 5.5: Layer purity after annealing of deposited material, results from the HMC sampler on the V Sample.

Layer	Mean	Best	95% CI
TopLOI	0.99	1.00	1 - 0.98
BotLOI	1.00	1.00	1 - 1

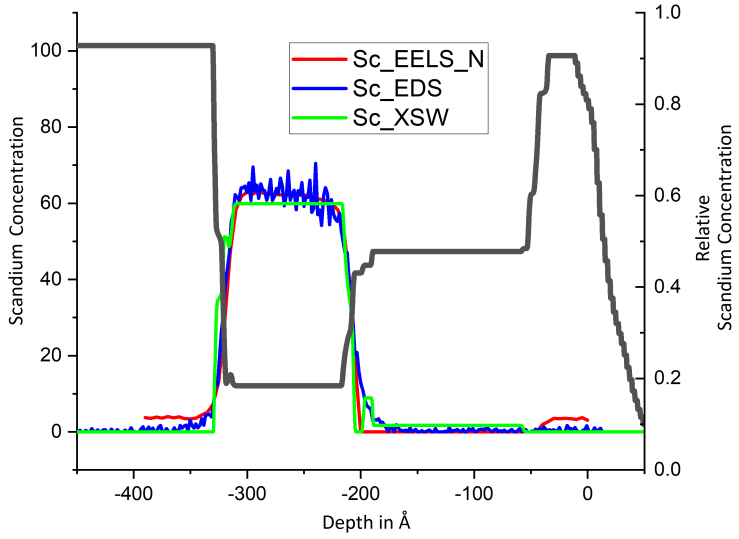


Figure 5.21: Comparison of the different Sc distributions found using XSW, EDS and EELS.

The Niobium layer is also still intact and the distribution is isolated to the top layer of interest. Again the presence of the Niobium in the waveguide indicated by the EELS profile is caused by the overlapping Fluorescence lines which is explained in detail in the appendix section .2. Again the no solutions exist for an intermixed top layer of interest as seen in Table 5.5.

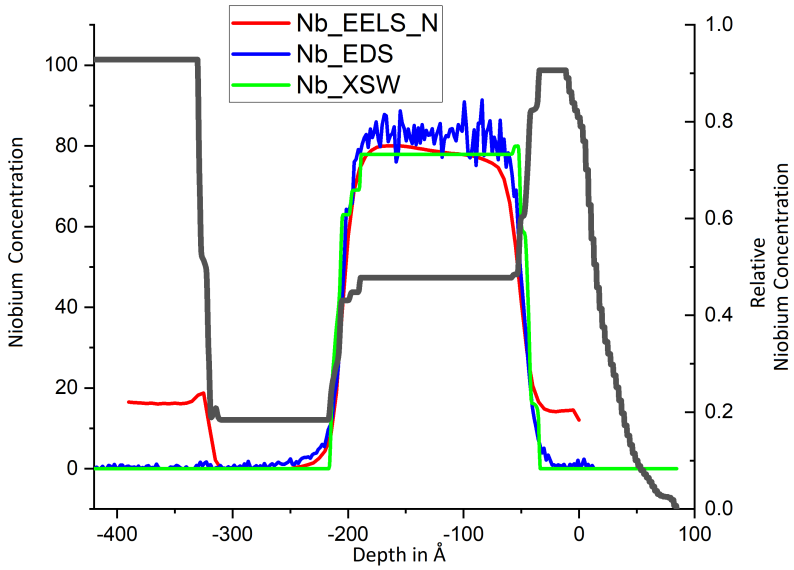


Figure 5.22: Comparison of the different Nb distributions found using XSW, EDS and EELS.

Gaussian Fits

The Gaussian fitting procedure (Described in Section 4.3) results in an almost identical distribution as the free-form parametrization distributions (Figures 5.23 for Cr/Fe/Co & 5.24 for V/Sc/Nb) , confirming the assumptions made on the distribution shape and location of potential segregation sites.

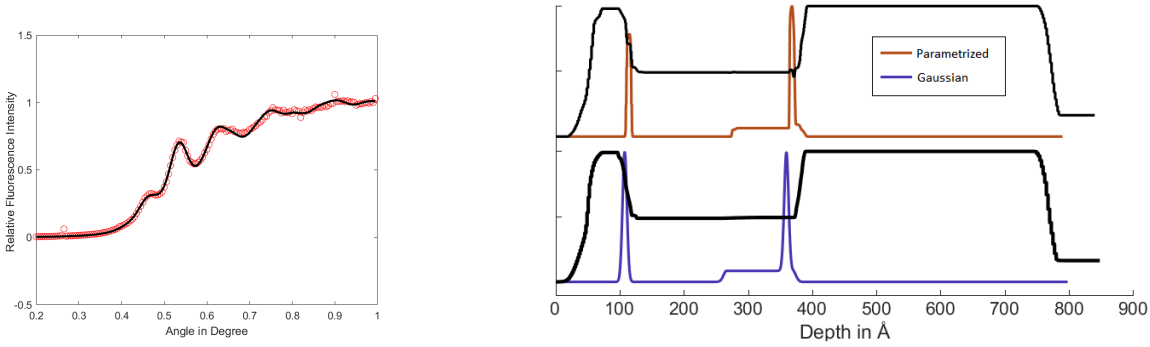


Figure 5.23: Resulting distribution Cr, obtained with the Gaussian fitting procedure (Cr/Fe/Co Sample). On the left side the measurement and resulting fit is shown of the XSW Measurement.

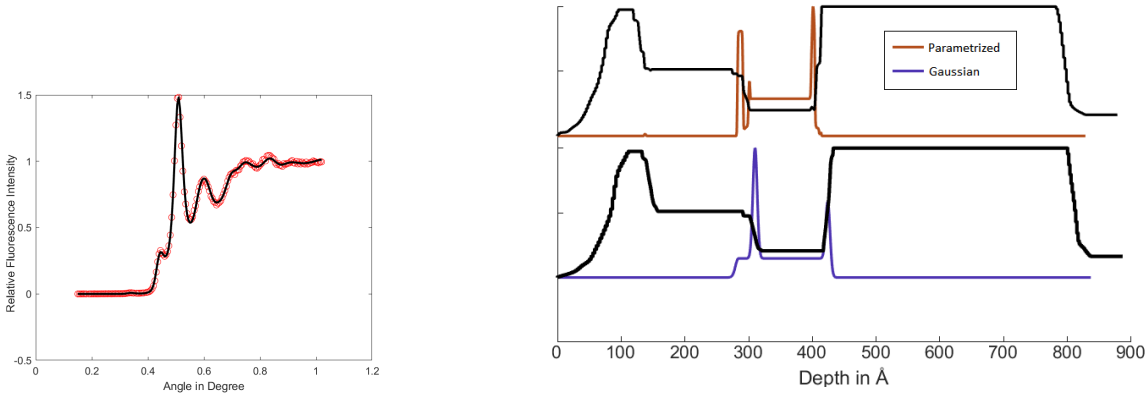


Figure 5.24: Resulting distribution of V, obtained with the Gaussian fitting procedure (V/Sc/Nb Sample). On the left side the measurement and resulting fit is shown of the XSW Measurement

5.5 Efficacy of the XSW measurements

To evaluate the effect of XSW data on the accuracy of the thin film analysis we have calculated the confidence intervals of thin film reconstructions obtained from XRR-only, XSW-only and combined XRR-XSW analysis sampled with the HMC sampler. The GIXR reconstructions are known to be prone to multi-modalities and corridors of acceptable reconstructions and the addition of extra data sets has known to reduce the number of acceptable solutions.

The XSW-only dataset is generally not used on its own since it is expected not to yield stringent results on but as additional information is can be used to narrow down the confidence intervals by restricting the number of acceptable solutions. In Figures 5.25 & 5.26 respectively the relative size of the 95% confidence intervals of the layer δ and layer widths for the V/Sc/Nb Sample are shown. The sampled parameters and the 95% confidence intervals for the combined data set are shown in Tables 5.6 and 5.7. For the GIXR data set this is shown in Tables 5.9 and 5.8. The figure 5.25 shows that the accuracy of the determination of the layers optical constant can be significantly increased by adding the XSW measurements to the GIXR measurements. Also the layer thickness reconstruction benefits from the addition of the XSW data-set, although not as much.

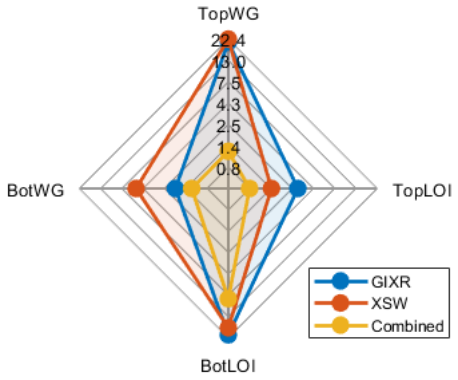


Figure 5.25: Relative width of the 95% confidence intervals in percentages of the layer δ from different datasets resulting from the HMC sampling of the V/Sc/Nb Sample.

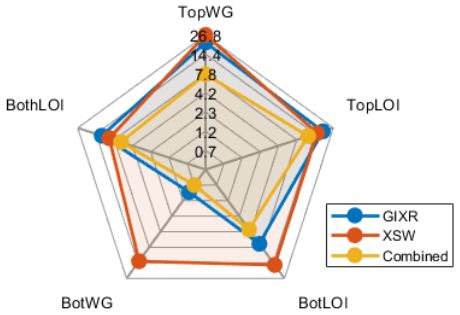


Figure 5.26: Relative width of the 95% confidence intervals in percentages of the layer thickness from different datasets resulting from the HMC sampling of the V/Sc/Nb Sample.

Both of the pyramid plots of the posterior of all structural parameters are shown below in the Figures 5.27 and 5.28. While the combined data set posterior shows inferences of convex shape and no or linear correlations between parameter pairs, the GIXR only data set shows strong non-convex inferences, some multi-modalities and also complex correlations are visible. Indicating

Table 5.6: Thickness of the combined dataset.

Layer	Mean	Best	95%
TopWG	65.72	66.74	68.45-63.64
TopLOI	175.5	176.2	191.8-170.0
BotLOI	117.6	115.6	120.3-115.3
BotWG	382.4	383.9	383.6-381.2
BothLOI	293.2	291.7	307.2-287.8

Table 5.7: δ of the combined dataset ($*10^5$)

Layer	Mean	Best	95%
TopWG	4.516	4.502	4.549-4.493
TopLOI	2.393	2.394	2.404-2.385
BotLOI	9.236	0.932	0.962-0.877
BotWG	4.556	4.559	4.583-4.527

Table 5.8: Thickness of the GIXR dataset

Layer	Mean	Best	95%
TopWG	64.33	65.41	73.19-59.42
TopLOI	179.4	172.3	200.3-167.1
BotLOI	119.5	116.9	124.3-116.6
BotWG	383.57	382.6	384.2-380.9
BothLOI	299.4	289.2	321.6-286.6

Table 5.9: δ of the GIXR dataset ($*10^5$)

Layer	Mean	Best	95%
TopWG	4.229	4.449	4.483-4.537
TopLOI	2.391	2.396	2.429-2.360
BotLOI	0.937	0.942	1.012-0.818
BotWG	4.565	4.577	4.606-4.520

the XSW data set is effective at eliminating the modes that are not of interest and severely restricting the solution space.

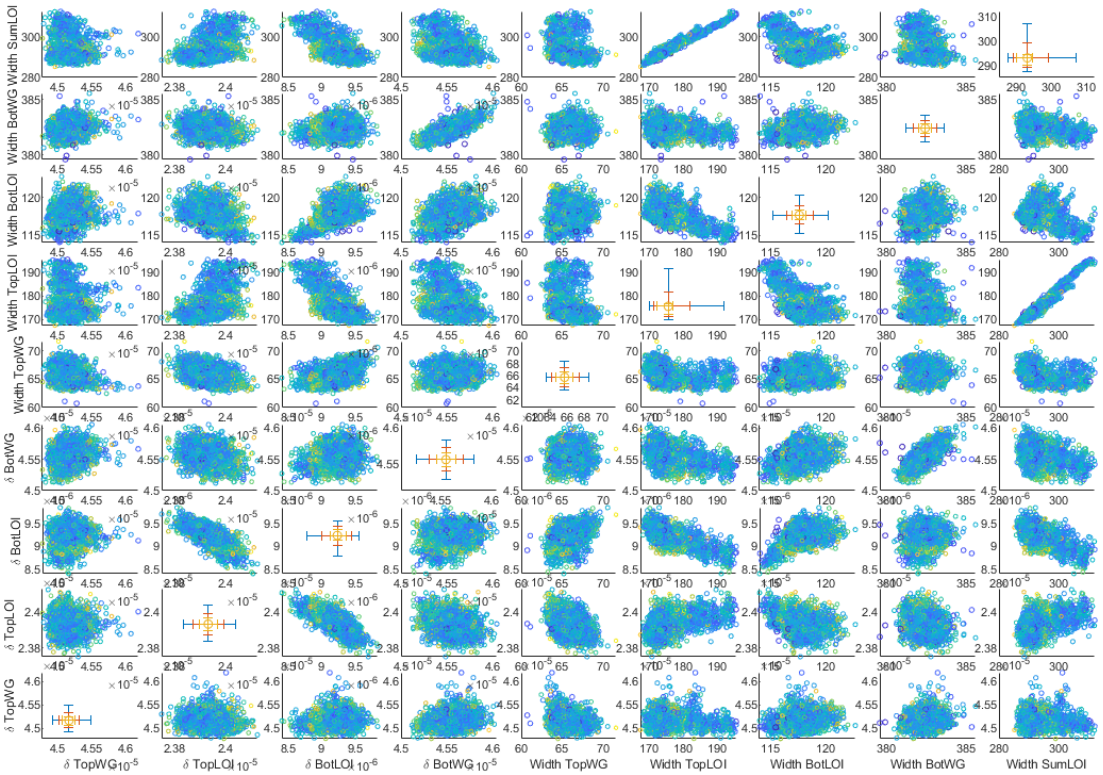


Figure 5.27: Pyramid plot showing the structural parameters for the individual samples for the GIXR-XSW data.

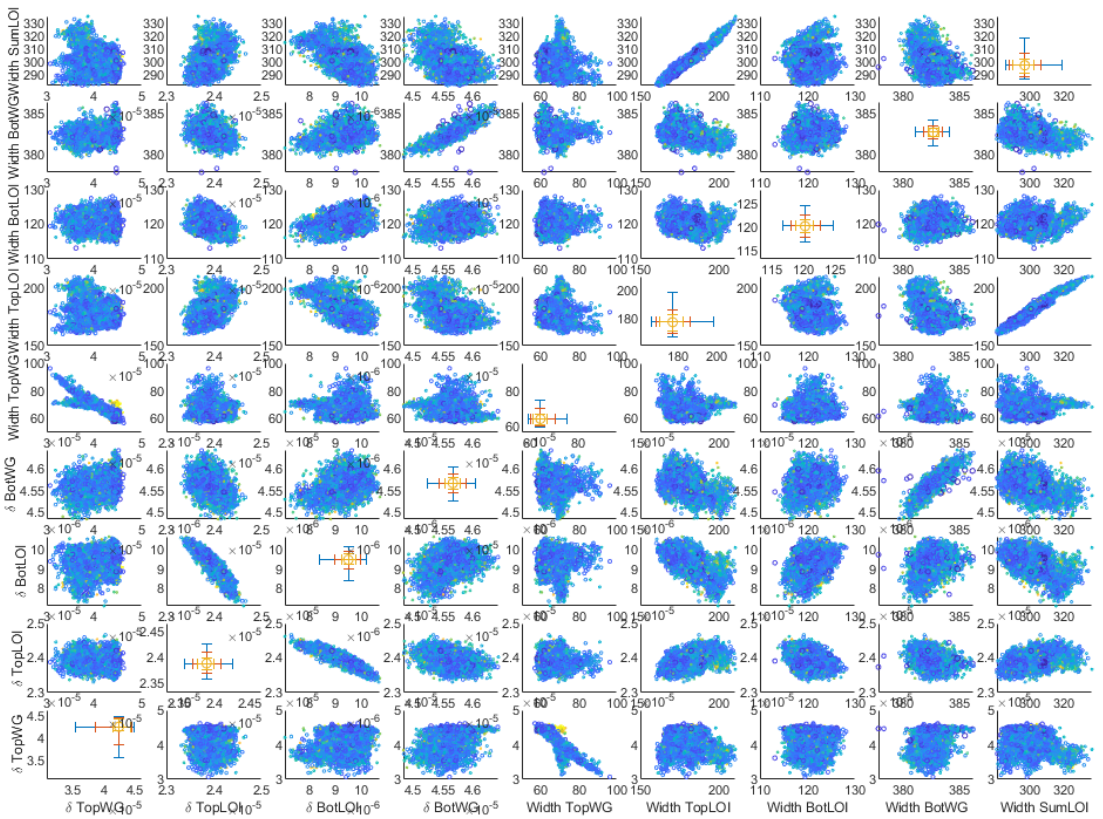


Figure 5.28: Pyramid plot showing the structural parameters for the individual samples for the GIXR data.

5.6 MCMC performance comparison

For both the HMC and Metropolis-Hastings, a long round of sampling has been done to sample 10000+ samples of the V/Sc/Nb annealed sample measurement data set to make a performance comparison. The Figure 5.29 shows the comparison of the burn in period between the Metropolis-Hastings sampling and the HMC sampling from an identical starting point. According to Figure 5.29 it is clear that the classic style Metropolis-Hastings burn in period in practice is much longer, even to the point that after 5 hours the loss function equilibrium sampling value was still more then 5 times higher than at which the HMC was sampling. The HMC was sampling at the objective value that is comparable to the X_{MAP} that is reached when using the deterministic optimization methodology, meaning the samples where of comparable likelihood as the most likely found parametrization.

After observing that the burn in phase had been stabilized at a level that was considered outside the acceptable range of the loss functions value, the σ of the proposal distribution of the Metropolis-Hastings was divided by 4 to improve the acceptance rate by making smaller steps and therefore increasing the probability to step towards an acceptable probability. The result, starting from the black arrow in Figure 5.29 was a very slow down sloping trend which would not within considerable time reach the preferred sampling loss function value. Reducing the temperature further only slowed down the acceptance rate without going down in objective value faster.

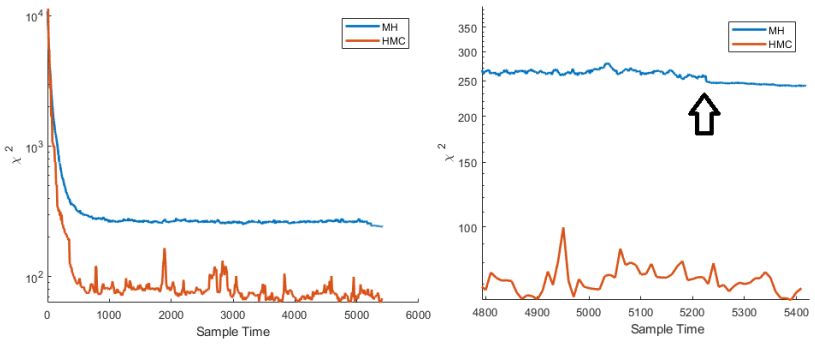


Figure 5.29: On the left is the behavior of the burn in period is shown visible, on the right side a zoom is shown when the σ of the HM proposal distribution is cut 4.

At this point it was decided to restart the sampling in the highest probability region already found by the HMC sampler. After restarting in a high-probability region, the Metropolis-Hastings chain manages to stay sampling at the preferred value as seen in Figure 5.30. Any increase in temperature or σ did lead again to divergence from this region.

Once in equilibrium, The behavior of the different sampling methods is very different. The number of Metropolis-Hastings samples per unit of time is much higher than the HMC since the HMC samples are more expensive. In turn the correlation between successive samples for the HMC chain is much lower. In Table 5.10 different behavioral specifications are shown of the 2 different sampling methods. ΔX_{step} is the average distance travelled in in parameter space per step. t_{step} is the time is takes per step. SCS are the number of successive steps made in a correlated direction (steps per sample). ΔX_{Sample} is the distance between each successive samples and t_{sample} is the average time it takes to sample and accept such a sample.

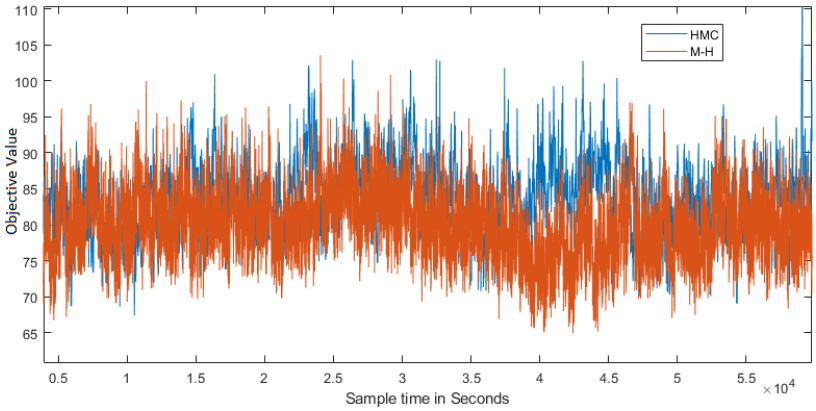


Figure 5.30: Comparison of the Objective value of the loss function of the samples versus the sample time for both MCMC methods.

Table 5.10: Comparison Metropolis-Hastings to HMC

-	M-H	HMC
ΔX_{step}	0.002	0.0015
t_{step}	1 sec	$\frac{1}{3}$ sec
SCS	1	90
ΔX_{Sample}	0.002	0.05
t_{sample}	1 sec	30 sec

Although the ΔX_{step} for the HMC method and Metropolis-Hastings is very comparable, the acceptance rate for the M-H is not very high ($\frac{1}{20}$), which for the HMC is around 70%. Leading to an average 3 times more time needed for a single step travelled using the M-H. The most important difference however is that for the HMC per sample on average 50 steps are made in a similar direction, thereby overcoming the random walk behavior, leading to much larger distances in parameter space in successive samples, in turn leading to a much smaller auto-correlation in the sample chain.

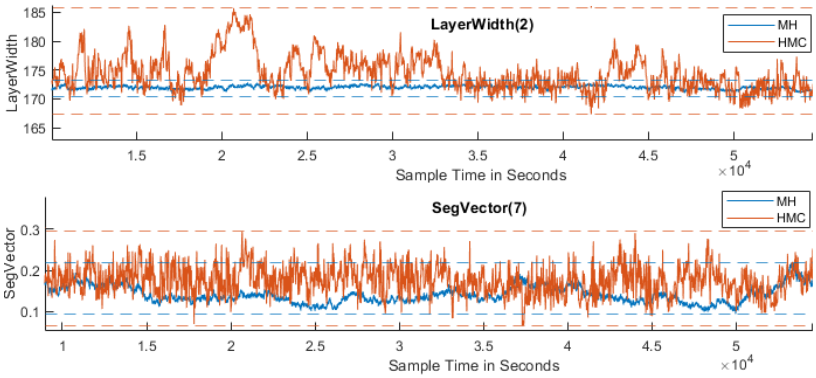


Figure 5.31: Comparison of parameter range that is sampled during identical time periods of the width of the top layer of the waveguide on top and the segregation vector value of the bottom interface on bottom, for V/Sc/Nb Sample, for both MCMC methods.

Equation 5.1

Expected time to travel an abitrary distance d by random walk:

$$\sim (\frac{d}{\Delta X_{Sample}})^2 t_{sample} \tag{5.1}$$

In Figure 5.31 the thickness of the top layer of interest and the segregation vector value of the bottom interface as an example are plotted during the sampling for the 2 different sampling methods. The dotted lines are on the level of the maximum and minimum of the value of the designated parameter that has been sampled for the respective method. The neighbouring values of the HMC samples are clearly less correlated and within comparable time a much greater part of the parameter space is explored as can be seen from the extreme values that have been explored. Using a random walk analogy we can compare the expected time to travel an arbitrary distance d from a starting point expressed in Equation 5.1. Using the values from Table 5.10 the ratio of expected time to travel a distance d is 31.125 in favour of the HMC method, explaining the lack of explored values by the MCMC chain in the same time.

Table 5.11: Comparison Metropolis-Hastings to HMC

-	M-H	HMC
$2\sum_{i>1}(X_1, X_i)$	1.5	$7.8 \cdot 10^{-3}$
N	$6.1 \cdot 10^4$	$3.0 \cdot 10^3$
$\sqrt{\frac{\sigma(F)^2}{N}}$	$5.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$

Where N is the number of samples sampled after 24 hours on a modern desktop PC (Intel i7-2600K) .

To use Equation 2.18, first the chain should enter the stationary regime where $\{X_i\}$ becomes independent of i to get an accurate measure of the averages

and variances of the distribution. For the HMC sampler this condition is reached after 5000 seconds (about 150 samples), seeing that that no correlation is present between previous points at this distance. From Figure 5.31 one can see that the normal Metropolis-Hastings sampler is not able to explore some regions that the HMC is able to explore, indicating that even after 60000 seconds stationary is likely not achieved.

From the values shown in Figure 5.31 that are explored by both samplers, it is clear that the normal Metropolis-Hastings chain is not able to explore the parameter space of acceptable values as effectively as the HMC sampler is. The term $\sum_{i>1}(X_1, X_i)$ will be significantly larger for the Metropolis-Hastings sampler than for the HMC since also the correlation between successive samples is visibly much larger.

To use Equation 2.18, a stationary state of the Markov chain should be used. Using the Metropolis-Hastings chain as an input will only yield a upper bound on the convergence since this chain after 60k samples is not yet stationary. Assuming that stationary is archived for the metropolis-Hastings sampler, the convergence comparison can be calculated. Using as measure the segregation vector to the bottom interface (shown in the bottom of Figure 5.31), the convergences are calculated and shown in Table 5.11. The term $\sigma_0(F) = 8.7 \cdot 10^{-5}$, calculated using the HMC values for both chains show negligible contribution to the $\sigma(F)$ term. The variance of the the difference between the 2 methods for a sample time of 24 hours on a regular modern consumer PC shows a factor of 10 difference in the variance of the converging distribution, indicating a much faster convergence for the HMC chain.

5.7 Material dependent segregation behavior analysis

All material combinations that have been succesfully deposited and annealed are shown in Table 5.12. The materials are listed as Pri(dopant), Sec(matrix) and Ter(neighbouring top layer). In the standard notation this is noted as Pri/Sec/Ter. In the column SegVec the observed interface segregation behavior after the annealing is shown. A label B(Bottom), M(Middle), T(Top) is given to the samples showing significant segregation to this respective interface. Significant segregation is defined as at least a segregated fraction of 0.05 of the total dopant material present in the structure.

In the Table 5.12 the material parameters of interest are listed and labeled by R(atomic radius), S(Surface energy) and C(crystal structure). Displayed in the column ΔPar, a plus or minus is added to indicate the difference between the dopant value of the parameter and the matrix value. The individual parameter isolation did not yield any predictive power for the occurance of interface segregation.

Miedema's value for the enthalpy prediction (Equation 2.1) of the interface segregation is shown in column $H^{segr}_{Interface}$ of Table 5.12 for the different interfaces. For the middle interface the material labels are defined as Pri/Sec/Ter = A/B/C. For the bottom interface C is changed with the waveguide material Tungsten. The model however is not formulated for the top interface since this involves the segregation to a non-neighbouring interface. To correct for this the H^{segr}_1 is replaced with H^{sol}_{AinB} to correct for the movement of atom A to the non-neighbouring interface and H^{segr}_2 and H^{segr}_3 are taken for the interface between the Ternary material and the waveguide material. (Respectively B and C in Equation 2.1)

Table 5.12: Table showing the individual segregation behavior of the samples after annealing.

Lab	Pri	Sec	Ter	SegVec	ΔPar	H^{segr}_{Top}	H^{segr}_{Mid}	H^{segr}_{Bot}
A	Cr	Ag	Cu	T,B	-R,C,+S	-95.2	-18.1	-87.7
B	Ti	Zr	Nb	M,B	-R,-S	-10.6	-8.13	-8.01
C	Ti	Sc	Nb	M	-R,+S	-38.7	-34.5	-40.9
D	Cr	Fe	Co	T,B		-9.07	-4.51	-3.58
E	Fe	Cr	Co	-		0.366	7.21	-4.75
F	Co	Cr	Fe	-	C	7.51	7.09	-2.11
G	Cr	Co	Fe	T,B	C	9.79	3.68	2.17
H	Co	Fe	Cr	-	C	-19.4	-2.14	-6.2
I	Fe	Co	Cr	-	C	-8.42	5.96	-3.07
J	Sc	V	Nb	-	+R,C,-S	34.5	0.313	-20
K	V	Sc	Nb	M,B	-R,C,+S	-25.3	-29.6	-25.8
L	V	Nb	Sc	-	-S	1.48	2.95	4.35
M	Sc	Nb	V	M	+R,C,-S	-50.5	-27.2	-21.1
N	Pd	Co	Fe	T,B	C	-34	-7.1	-14.8
O	Pd	Fe	Co	T,B	C	-11.5	0.584	-13
P	Fe	Pd	Co	-	C	10.1	6.82	7.48
Q	Co	Fe	Pd	-	C	-2.83	7.22	-6.2
R	Fe	Co	Pd	-	C	-5.94	0.666	-3.07
S	Co	Pd	Ti	B	C	-61.1	40	2.05
T	Ti	Co	Pd	M	+R	-64.8	-50.8	20.5
U	Ti	Hf	Ru	B	-S	-136	-2.63	-11.7
V	Ti	Ru	Hf	-	-S	161	126	52.2
W	Ru	Ti	Hf	-	+S	30.4	5.5	50.3
X	Ru	Hf	Ti	-	-R	57.6	25.1	63

In Figure 5.32 the total dopant fraction segregated to the bottom interface

is shown versus Miedema's interface segregation enthalpy to this respective interface. A clear threshold behavior is visible in which the segregation tends to occur at a negative enthalpy. The enthalpy prediction however does not seem to hold any predictive power over the total fraction of segregated material and even samples are present with negative enthalpy for which no segregation occurs. The samples S and G both suffer from intermixing of the top and bottom layer of interest that happened during the annealing process, possibly changing the segregation enthalpy explaining the segregation while having a slightly positive enthalpy .

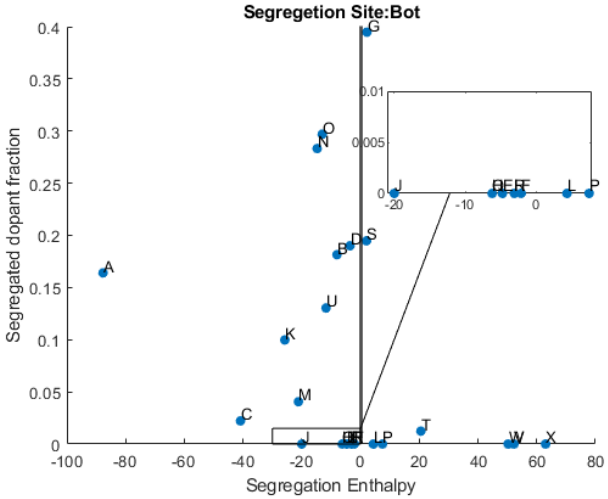


Figure 5.32: Segregation intensity versus the segregation enthalpy of the top interface for the dopant as segregant.

In Figure 5.33 the total segregated dopant fraction to the middle interface for the different materials is shown versus Miedema's interface segregation enthalpy to this respective interface. Again a clear threshold behavior is visible. The absence of segregation in samples N,D and H can be explained by the absence of an Co/Fe interface due to the intermixing of the top and bottom layer of interest that happened during the annealing process. The segregation for A is probably suppressed by the segregation to the Top and bottom interface since these sites are much more energetically favourable.

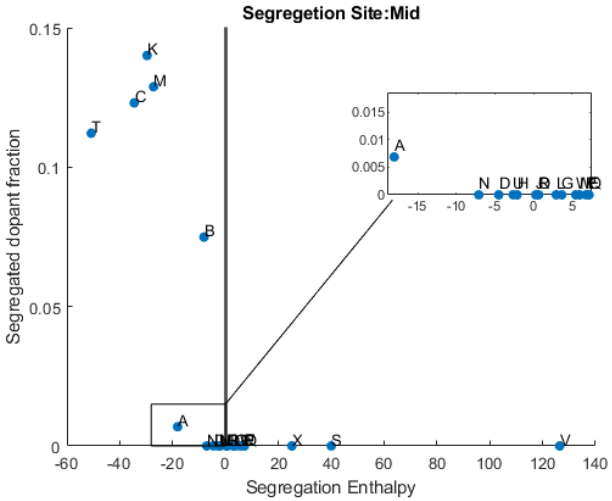


Figure 5.33: Segregation intensity versus the segregation enthalpy of the top interface for the dopant as segregant.

The segregation fraction of the dopant for the different materials to the top interface shown in Figure 5.34. The predicted value for this segregation site are not as well behaved as for the middle and bottom interface, very likely due to the rough approximation to the to different scenario that this model is not applicable to.

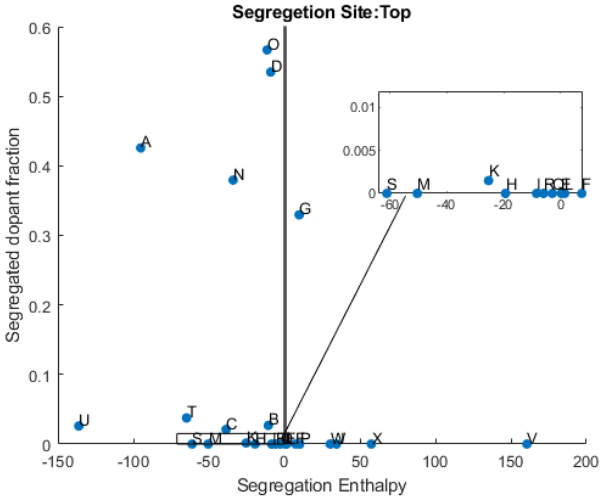


Figure 5.34: Segregation intensity versus the segregation enthalpy of the top interface for the dopant as segregant.

Discussion & Conclusion 6

6.1 Wavegudes and X-ray standing waves for the analysis of interfacial segregation

The application of the waveguide for the generation of the standing wave has proven to be a powerful tool in the analysis of the atomic depth profiles in thin films. The X-ray standing wave, generated inside the waveguide effectively modulates the fluorescence yield from even distributions of single atomic percentages. Alternatively, these kind of reconstructions were mostly limited to more expensive, destructive and time consuming methods like electron microscopy or sputter-depth profile photo-electron spectroscopy.

The often noted disadvantage of the results obtained from the GIXR and XSW measurements is the absence of any stringent measure of uniqueness and tolerance on the individual parameters. This unique combination of the free-form parametrization combined with a Bayesian inference allows for a very stringent analysis into these previously ambiguous uncertainties of the reconstruction parameters.

From the Bayesian inference analysis, the addition of the XSW measurements has shown not only to be beneficial in the reconstruction of (low-concentration) atomic distributions but also the decreases the uncertainty in the in the reconstruction parameters of the layer thicknesses and densities.

The Hamiltonian Monte Carlo sampler converges notably faster than the classical-style guess-and-check Metropolis-Hastings sampler. An accurate parameter inference can be calculated within a couple of hours on a modern consumer style desktop PC. Such speed increase can facilitate a wider adoption in GIXR and XSW metrology in thin film RD departments.

No direct correlations or connections could be found in the single parameter isolation. A found threshold-dependence of the segregation process on the thermodynamic parameters of atoms in thin films was found on the predictions made in [7] (see Section 5.7). Because the annealing temperature of 400 °C was determined by the limitations of the equipment in the laboratory, it is not ensured complete mobility is archived for all samples, possibly explaining some of the anomalies in Section 5.7.

6.2 Limitations

Although the TEM measurements confirmed the resulting reconstructions and agree with the inferred probabilities one can never be sure to have explored the complete space of possible solutions present in the posterior since this objective is intractable in practice and not explored multi-modalities can always be present. Although theoretically the inferred confidence intervals on the reconstruction only serve as a lower bound, we have shown both with the sensitivity analysis in Section 5.2, the interval comparisons of multiple sampled chains and the TEM measurements in Section 5.4 that in practice the results agree completely with the as deposited sample structures and standardized measuring techniques.

One more limitation of the current implementation of the XSW method is the lack of quantitative power in characterization of atomic concentrations. Only the relative distribution profile of the different atoms present in the thin film can is obtained. Potentially as the XRF measuremens that are performed during the XSW experiment have gain quantitative sensitivity, the absolute atomic concentrations can be obtained by the calibration the XRF measurements against thin-film samples.

Even though the results have shown the location of segregated layers can be obtained with the sub-nanometer accuracy, there is a significant uncertainty from the inference in the quantitative fractions of dopant in the interfaces and layer of about 10% of the total fraction. This is an inherent limitation of the measurement techniques and could possibly be addressed by adding more data sets measured at different wavelengths.

6.3 Recommendations

The presence of multi-modalities in the results of X-ray-interference-based techniques poses an issue for the widespread adaptation of these methods. Although for the reconstructions and the inference of the parameters tolerances of the specific samples in this thesis show no issues identifying non-physical solutions. An interesting and useful extension of this research could be into the independent applicability of GIXR and XSW measurement reconstructions to a more generalized set of samples.

At the moment there was no requirement for a thorough quantitative analysis. In the future, to broaden the application and increase the usage of the developed reconstruction methods, a quantitative analysis can be included on the concentrations of atoms present at every depth position. This should be done by making a model that effectively calculates the shape of the excitation beam in the macro level of the sample and takes into account the cross sections of the different elements and the potential secondary excitation effects.

The time duration of the sampling of the posterior at the moment can still be hindrance in the adoption for routine reconstruction calculations that require instant results. Therefore a logical next step would be to improve the Hamiltonian Monte Carlo convergence by further optimizing computation cost of the derivatives. One approach could be to parallelize the calculations and use a multi-core CPU.

Currently the XSW data acquisition is focused on the accurate reconstruction thin films made of transition metals and operates best for chemical elements with periodic numbers between 16 and 65. This is due to the ambient atmosphere in which is measured and the Cu-kα radiation that is used for the excitation. To broaden the applications to, for example biological films, also a sensitivity in the lighter elements should be enabled. For this a step has to be made to an in-vacuum implementation.

At the moment the segregation behavior in thin films is still a relatively unexplored area of research. The clear threshold behavior that has been found following the predictions made in [7] could potentially be used as a rough predictor when to expect segregation in nano-meter thin films but the enthalpy predictions do not show to have good predictive power in the magnitude of segregation. Follow-up research should be guided towards the explanation of the magnitude of the segragation behavior.

6.4 Conclusion

The heat-induced segregation inside nano-meter thin films was characterized for 24 material combinations. The trends of segregation of various elements in various bi-layer system was characterized from the analysis of the atomic profiles of low concentration dopant atoms before and after annealing of the bi-layer system. The atomic profiles of the doping-atoms were obtained using the combined analysis of angular dependent X-ray fluorescence (the X-ray standing wave technique) and Grazing incidence X-ray reflectively measurements. The segregation that was observed in the different material combinations shows a clear threshold behavior following the predictions on formation enthalpy of segregation in layered systems done in [7].

An optimized waveguide structure has been designed to create an X-ray standing wave inside the thin films structure for this analysis. The special requirements such as optimal sensitivity to segregation of doping atoms to interfaces and the heat resistance of waveguide layers and bi-layer up to 400C were taken into account during the designing process. The free-form thin film parametrization was used for the analysis of the GIXR-XSW measurements. The accuracy of the GIXR-XSW analysis was tested experimentally by the comparison of the reconstructions from the X-ray measurements atomic profiles to the results of Transmission electron microscopy energy dispersive X-ray and electron energy loss spectroscopy images. This confirmed the correctness of the reconstructions to a sub nano-meter accuracy in the determination of the segregation depth for all segregation positions.

A Bayesian inference was applied to the free-form parametrization reconstructions sampled using Hamiltonian Monte Carlo methods. From the inferred probabilities the confidence intervals of the individual reconstruction parameters could effectively calculated thereby tackling ambiguity of the reconstruction uncertainty that has long been the main hindrance in the indepentent usage for complete thin film reconstuctions using only the GIXR-XSW measurements. Hamiltonian Monte Carlo methods have been successfully implemented to the free form reconstruction methodology and can effectively sample the posterior of possible reconstructions from the given hybrid GIXR and XSW data set. Due to the higher acceptance rate and lower correlation between successive samples, the convergence has shown to be notably faster in sampling of the posterior than classical guess and check methods.

Using the sampled posterior obtained with the Bayesian inference framework with and without the addition of the XSW dataset, it is concluded that the XSW dataset is not only beneficial for the reconstruction of dopant distributions and other atomic distributions but also reduces the uncertainty in the reconstruction of the optical constants and thicknesses of the individual layers.

APPENDIX

.1 Appendix: Pre-Segregated depositions test samples

Bottom Position

In the figure below the measurements and the fits of the Fluorescence can be seen for the different elements for the bottom position sample. The proposed fit with the Chrome being located on the bottom position clearly exhibits all features of the measured fluorescence signal.

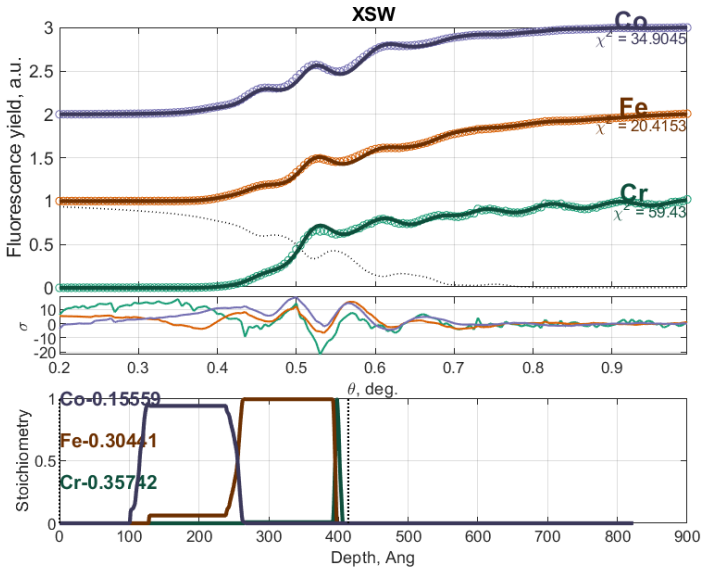


Figure .1: The Fluorescence Fit of the optimized structure of the to the bottom position depositioned Cr/Fe/Co sample.

Also the GIXR measurement can be explained by the fit that has been proposed.

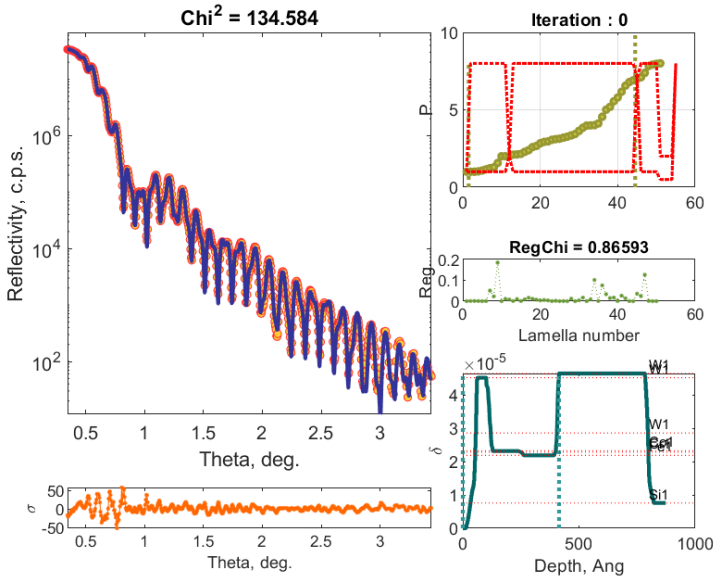


Figure .2: The GIXR Fit of the optimized structure of the to the bottom position depositioned Cr/Fe/Co sample.

Middle Position

Also for the middle position both the fluoresence and the GIXR measurements can be clearly described by the as deposited structure with the Chrome being isolated in the middle position.

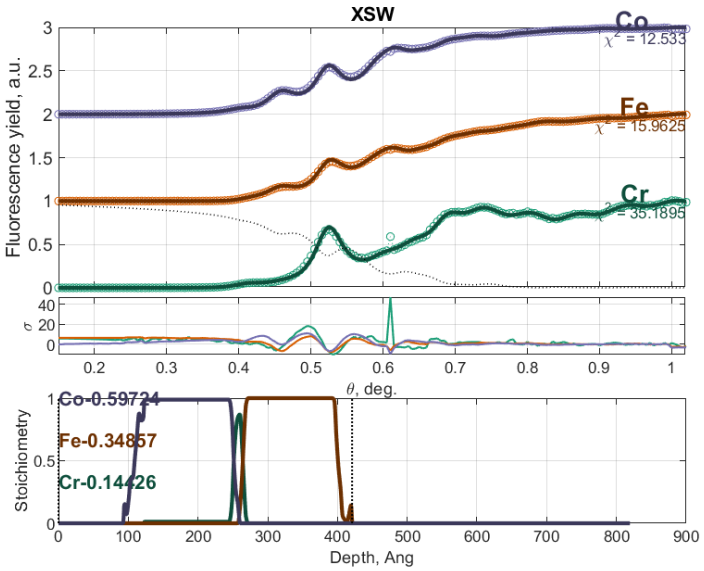


Figure .3: The Fluorescence Fit of the optimized structure of the to the middle position depositioned Cr/Fe/Co sample.

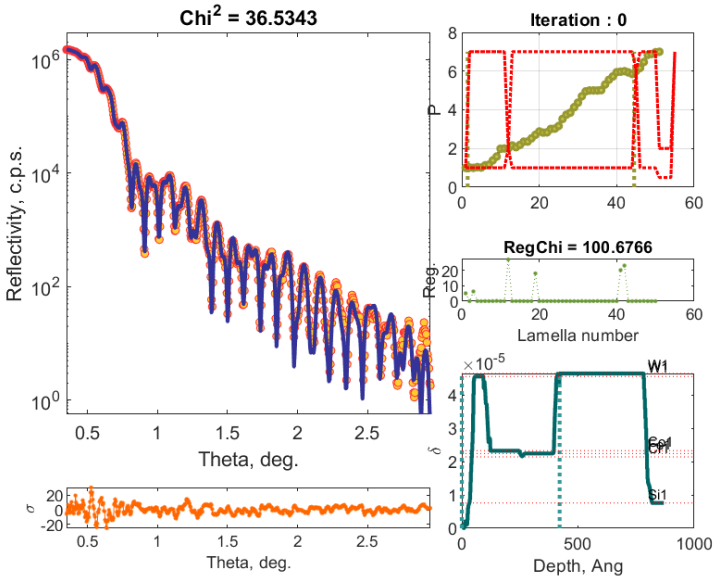


Figure .4: The GIXR Fit of the optimized structure of the to the middle position depositioned Cr/Fe/Co sample.

Top Position

The same holds true for the top position.

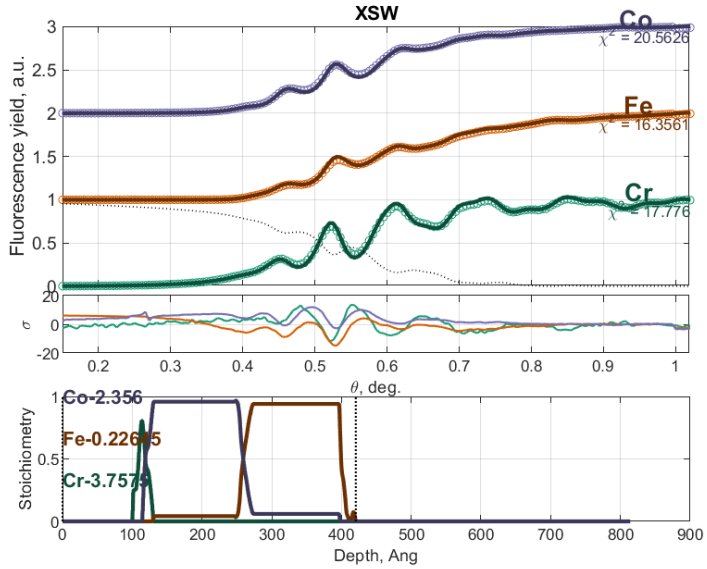


Figure .5: The Fluorescence Fit of the optimized structure of the to the top position depositioned Cr/Fe/Co sample.

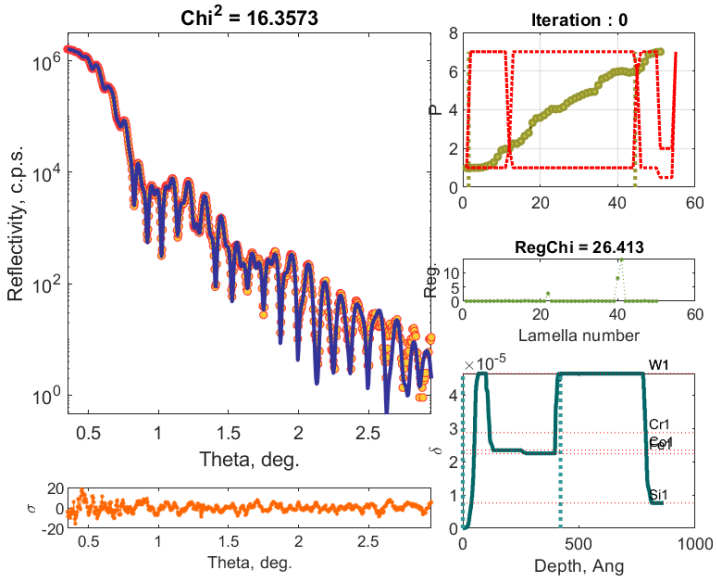


Figure .6: The GIXR Fit of the optimized structure of the to the Top position depositioned Cr/Fe/Co sample.

.2 Appendix: TEM data analysis of the Cr/Fe/Co & V/Sc/Nb Samples

GIXR & XSW

The sample analysis of XSW GIXR before annealing shows the measurements can easily be explained by the as deposited structure. The GIXR fit accurately described all thicknesses and densities of the layers as expected.

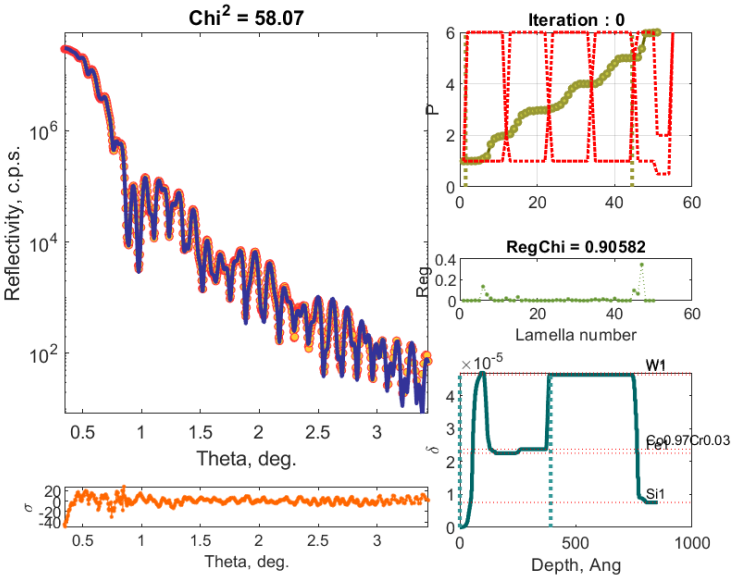


Figure .7: GIXR fit before Annealing of Cr/Fe/Co Sample.

Also the XSW signals are accurately described by the as deposited fitted structure.

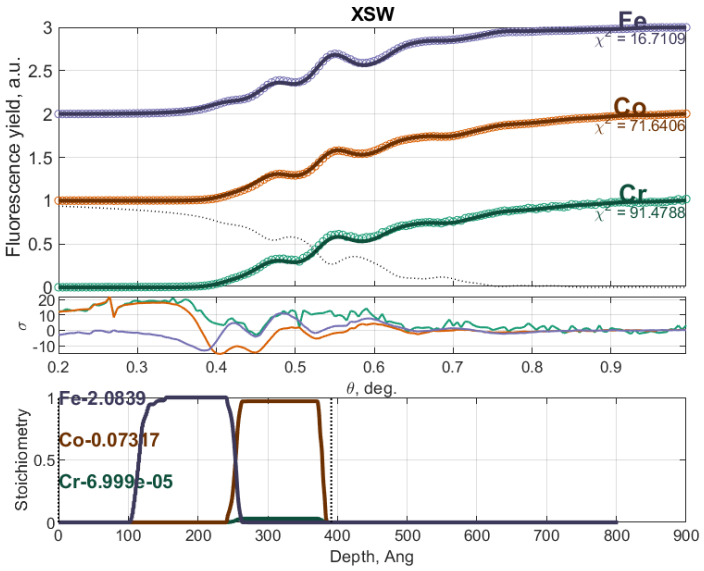


Figure .8: XSW fit before Annealing of Cr/Fe/Co Sample.

After annealing it seems that the Fe and Co layers have become intermixed. The GIXR percieves no more density differences between the layers and the XSW signal seems to indicate intermixing aswell due to the more overlapping signals for Fe and Co. The Cr XSW signal seems to indicate a complex segregation behavior which can be explained by segregation to both interfaces with with Tungsten.

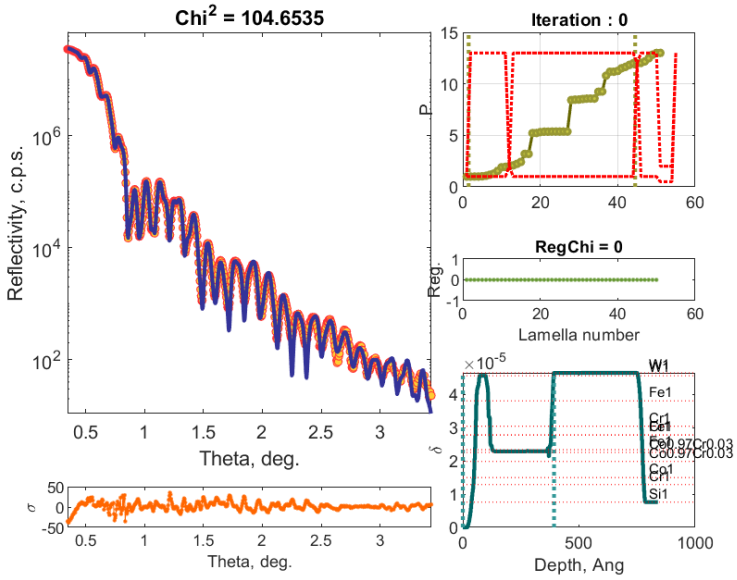


Figure .9: GIXR fit after Annealing of Cr/Fe/Co Sample.

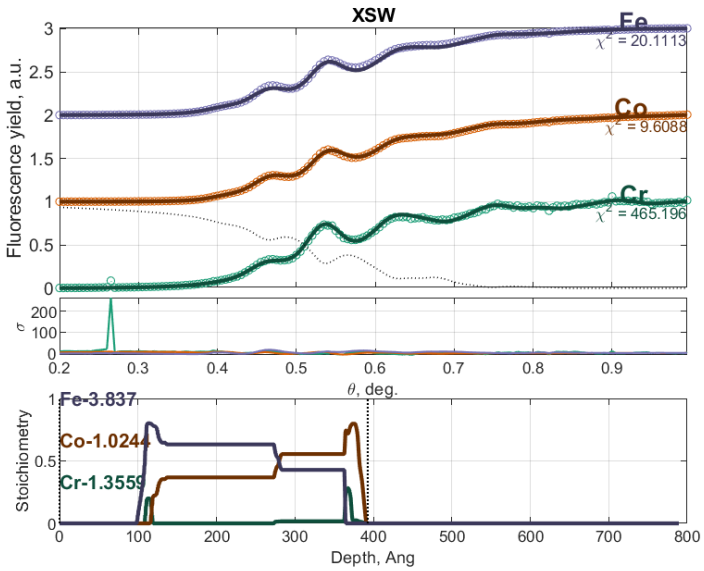


Figure .10: Fluorescence fit after Annealing of Cr/Fe/Co Sample.

Both before and after annealing a structure can be found which accurately describe the measurements of these samples before and after annealing.

TEM,EDX EELS

Since the TEM analysis resources are very limited only the sample after annealing was chosen to be analysed. The sample before annlealing did not show any unexpected results varying from the as deposited structure and is therefore assumed to be correctly described.

From the TEM image it is visible that after annealing the Co and Fe layers have merged. The waveguide layers seems to be intact with some oxidation

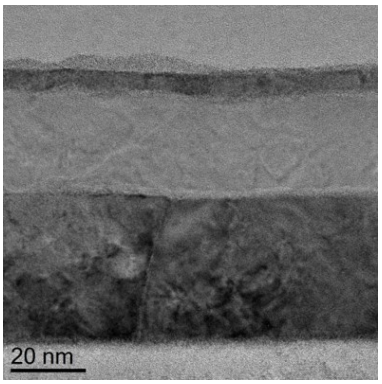


Figure .11: Close up TEM image of the Cr/Fe/Co Sample

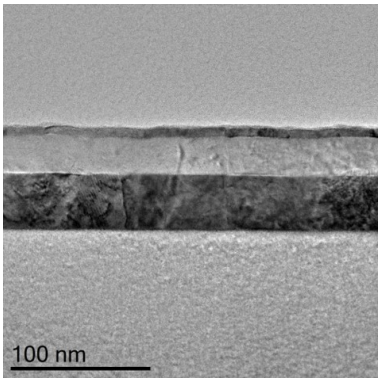


Figure .12: At a distance TEM image of the Cr/Fe/Co Sample

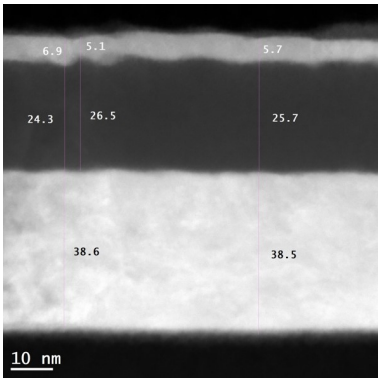


Figure .13: spatial HAADF spectrum accumulated from the Cr/Fe/Co Sample

on the top which is only scarcely observed on the sample. There is some low frequency roughness which is known to occur when making these samples. When looking at the details, a small interface can be observed between the W and the inner layers which could be the proposed Cr segregation.

From the TEM and HAADF it can be seen that the structural integrity of the waveguide is maintained and all the layers have the widths that also the XRR analysis reconstructed.

The HAADF analysis shows that the waveguide structure only is compromised of Tungsten and in the middle layers no tungsten is present, confirming the waveguides structural integrity.

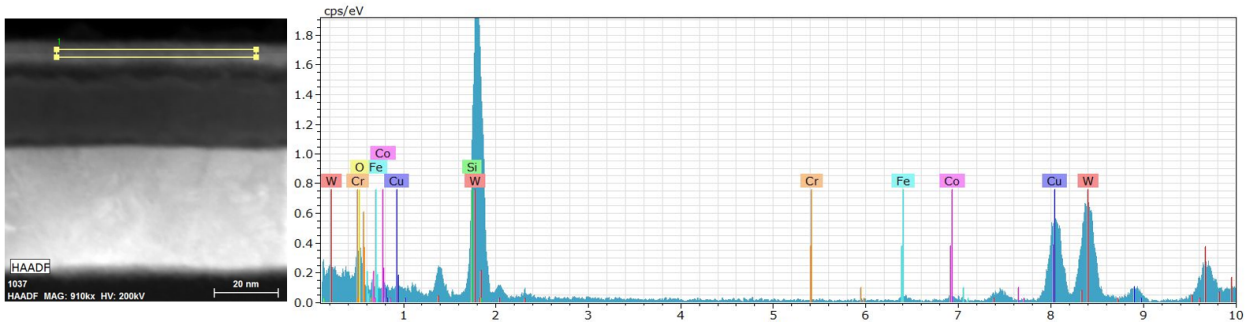


Figure .14: Elemental analysis of the top waveguide layer of the structure

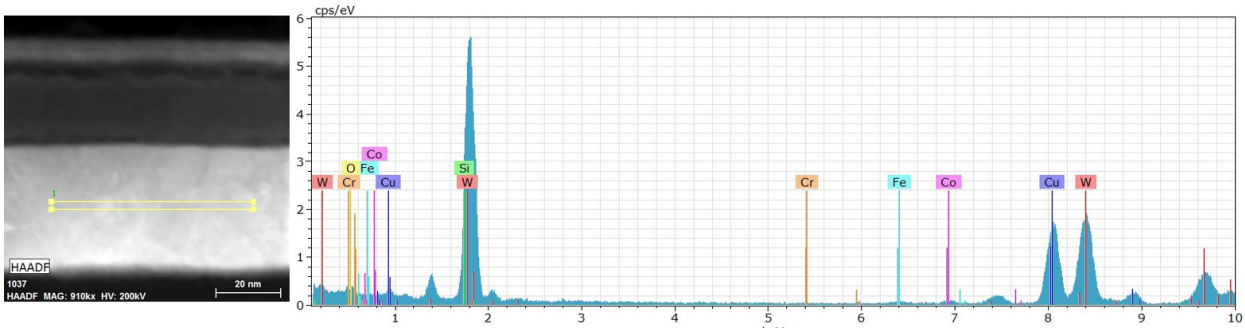


Figure .15: Elemental analysis of the bottom waveguide layer of the structure

In both interfaces neighbouring the Tungsten waveguide Chrome is present in higher concentrations than any other place in the structure indicating segregation to these interfaces of the Chrome dopant. The concentration on the bottom interface seems to be a bit higher than on the bottom, but is also seems to be a bit broader.

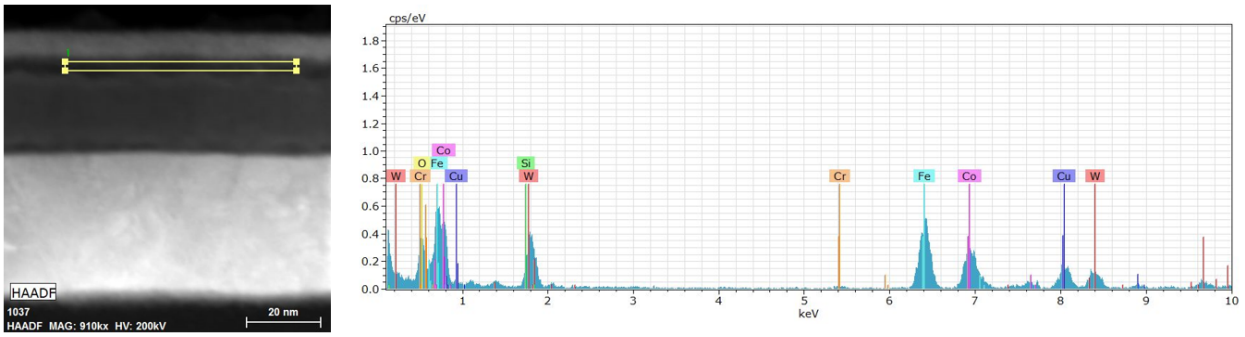


Figure .16: Elemental analysis of the top interface of the structure

Whether Chrome is still present in the bottom layer of interest is inconclusive from this analysis since the expected concentration is is equal to the sensitivity of the HAADF machine. It also seems some oxidation has occured to the Fe/Co compound layer that has formed due to the annealing.

The EELS analysis confirms the results found with the HAADF methods. The Tungsten waveguide structure is uncompromised with minor oxidation on the top. Using IELS also the oxidation of the middle layer is confirmed. Also the intermixing of the Fe and Co is again visible using EELS.

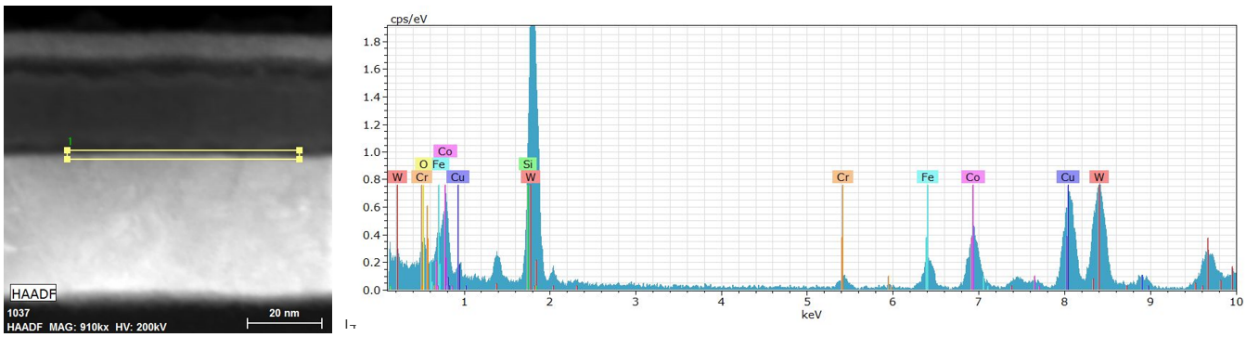


Figure .17: Elemental analysis of the bottom interface of the structure

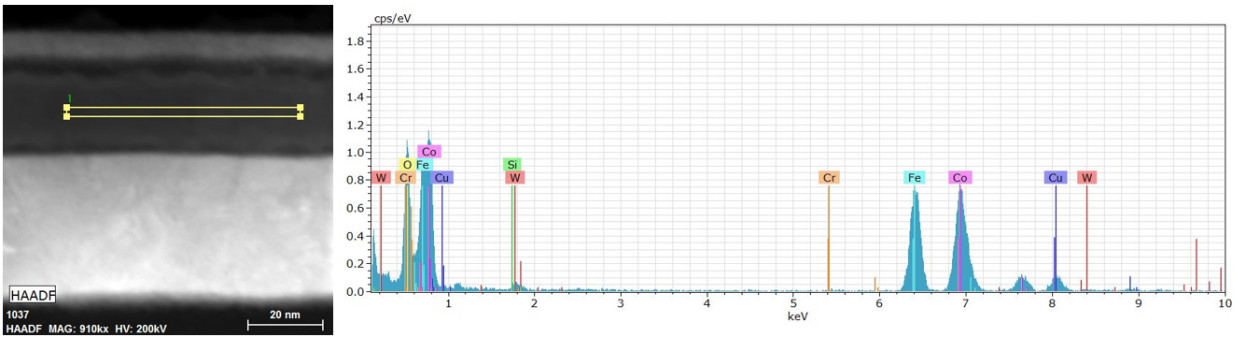


Figure .18: Elemental analysis of the layer of interest of the structure

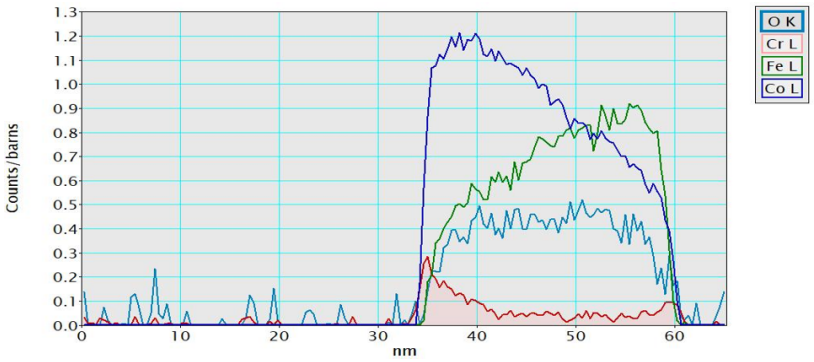


Figure .19: EELS analysis of the annealed Cr/Fe/Co sample.

The EDS analysis shows similar results. The Si and C and part of the Cr signal are both caused by the overlap of their fluorescence peaks.

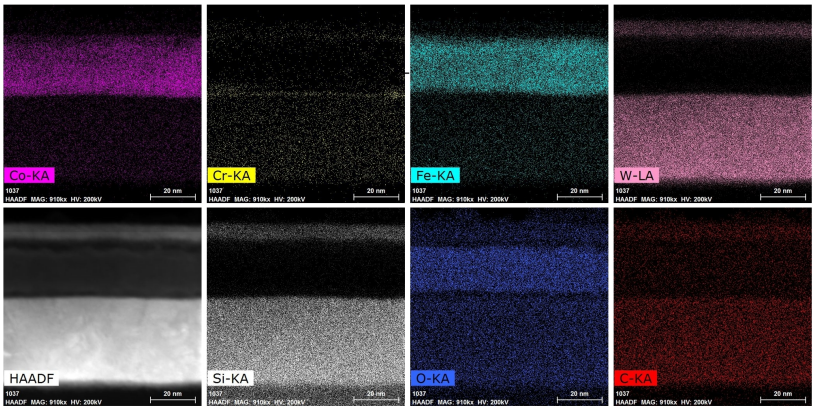


Figure .20: EDS analysis of the annealed Cr/Fe/Co sample.

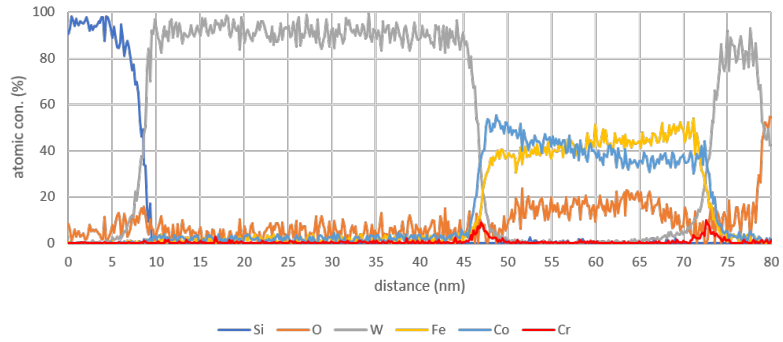


Figure .21: Resulting depth profiles of the different elements resulting from the EDS analysis of the annealed Cr/Fe/Co sample.

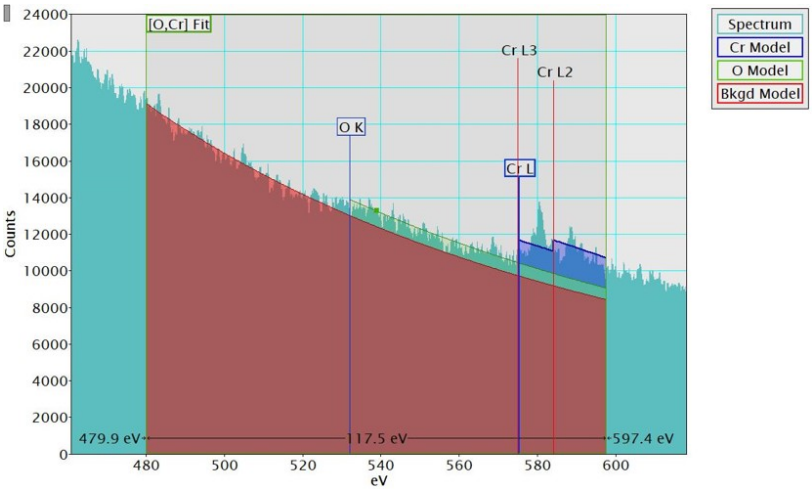


Figure .22: EELS peak extraction deconvolution for the Cr/Fe/Co Sample.

Sample V/Sc/Nb

Also a good fit of these structures was found leading it to be an ideal candidate to test the robustness of this method of analysis.

H

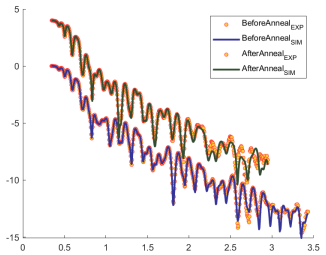


Figure .23: Fitquality of GIXR measurements.

GIXR & XSW

The sample analysis of XSW GIXR before annealing shows the measurements can easily be explained by the as deposited structure. The GIXR fit accurately described all thicknesses and densities of the layers as expected.

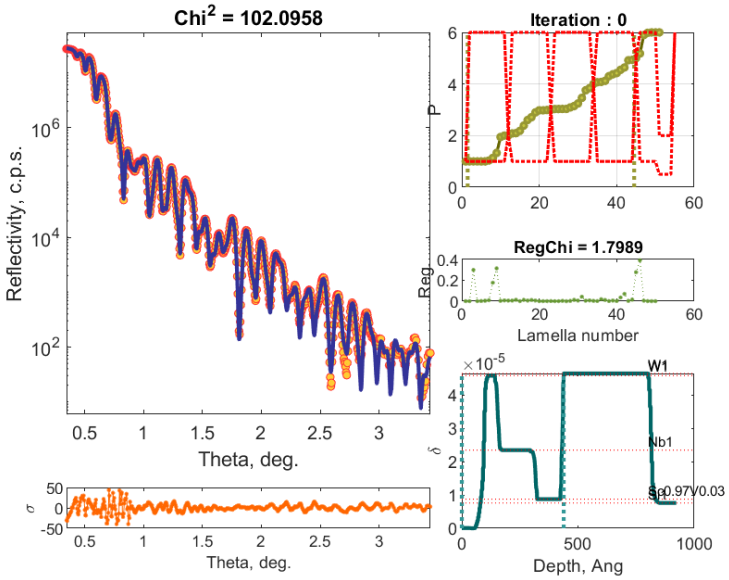


Figure .24: GIXR fit before Annealing of V/Sc/Nb Sample.

Also the XSW signals are accurately described by the as deposited fitted structure.

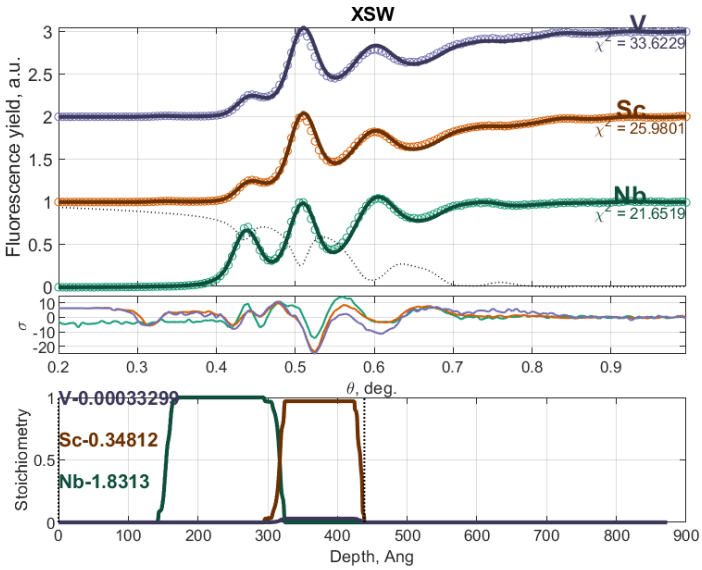


Figure .25: XSW fit before Annealing of V/Sc/Nb Sample.

After annealing it seems that the Fe and Co layers have become intermixed. The GIXR percieves no more density differences between the layers and the XSW signal seems to indicate intermixing aswell due to the more overlapping

signals for Fe and Co. The Cr XSW signal seems to indicate a complex segregation behavior which can be explained by segregation to both interfaces with with Tungsten.

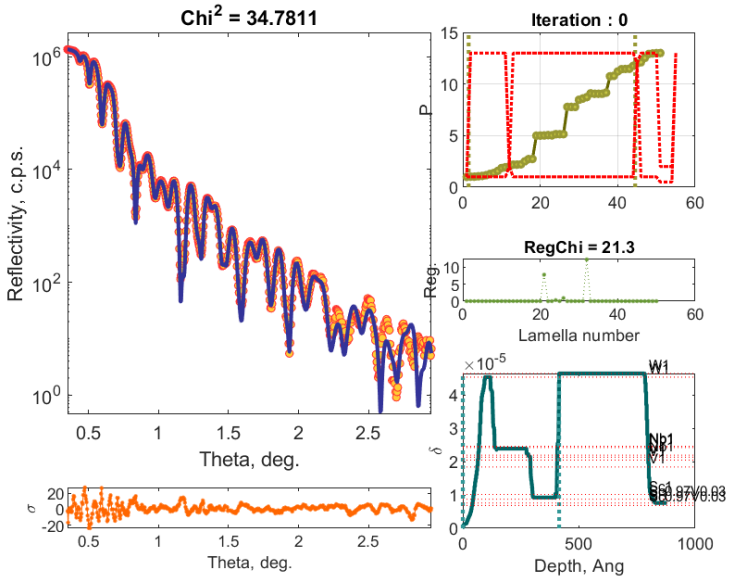


Figure .26: GIXR fit after Annealing of V/Sc/Nb Sample.

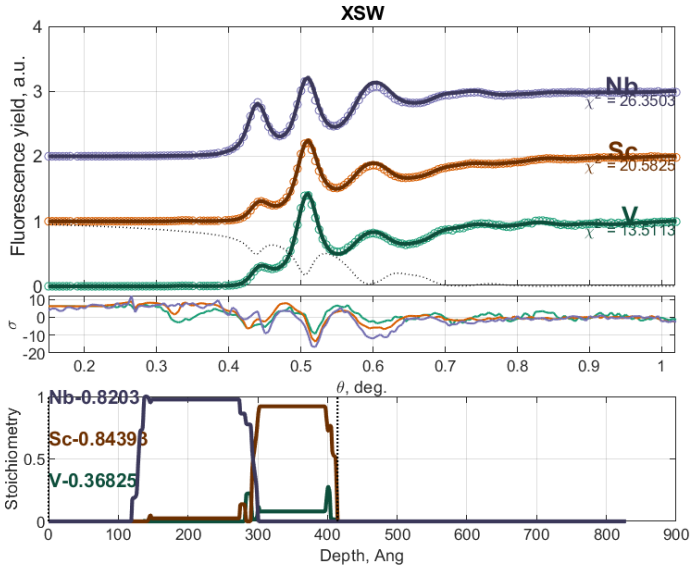


Figure .27: Fluorescence fit after Annealing of V/Sc/Nb Sample.

Both before and after annealing a structure can be found which accurately describe the measurements of these samples before and after annealing.

TEM,EDX EELS

From the TEM image it is visible the Sc and Nb layers still seem to be separate systems. The wave guide layers seems to be almost intact. In some places oxidation took place and a small hill emerged on the surface with a prominence

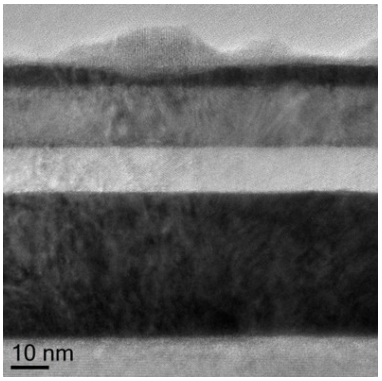


Figure .28: Close up TEM image of the V/Sc/Nb Sample

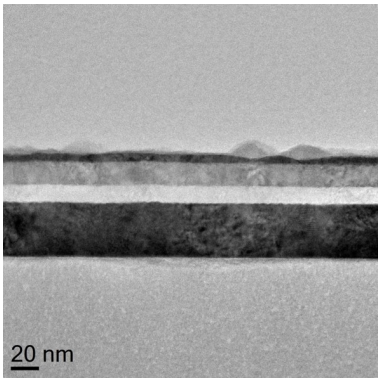


Figure .29: At a distance TEM image of the V/Sc/Nb Sample

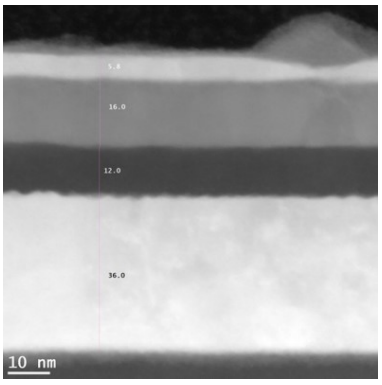


Figure .30: spatial HAADF spectrum accumulated from the V/Sc/Nb Sample

of about 6nm. Since the samples have been annealed and measured 4 months before the TEM analysis took place, this oxidation could have happened in the meantime . There is some low frequency roughness which is known to occur when making these samples although it is less than is present on the Cr/Fe/Co sample. When looking at the details, a small interface can be observed between the bottom and middle interface the by the XSW proposed V segregation.

From the TEM and HAADF it can be seen that the structural integrity of the waveguide is maintained and all the layers have the widths that also the XRR analysis reconstructed.

The HAADF analysis shows that the waveguide structure only is compromised of Tungsten and in the middle layers no tungsten is present. The EDS confirms that the small hills are indeed Tungsten Oxide. Since the waveguide did not show signs of disintegration right after annealing on the GIXR it is assumed this formation happend after annealing. To confirm wether these have formed during annealing or after annealing a remeasurement can be done to test this hypothesis.

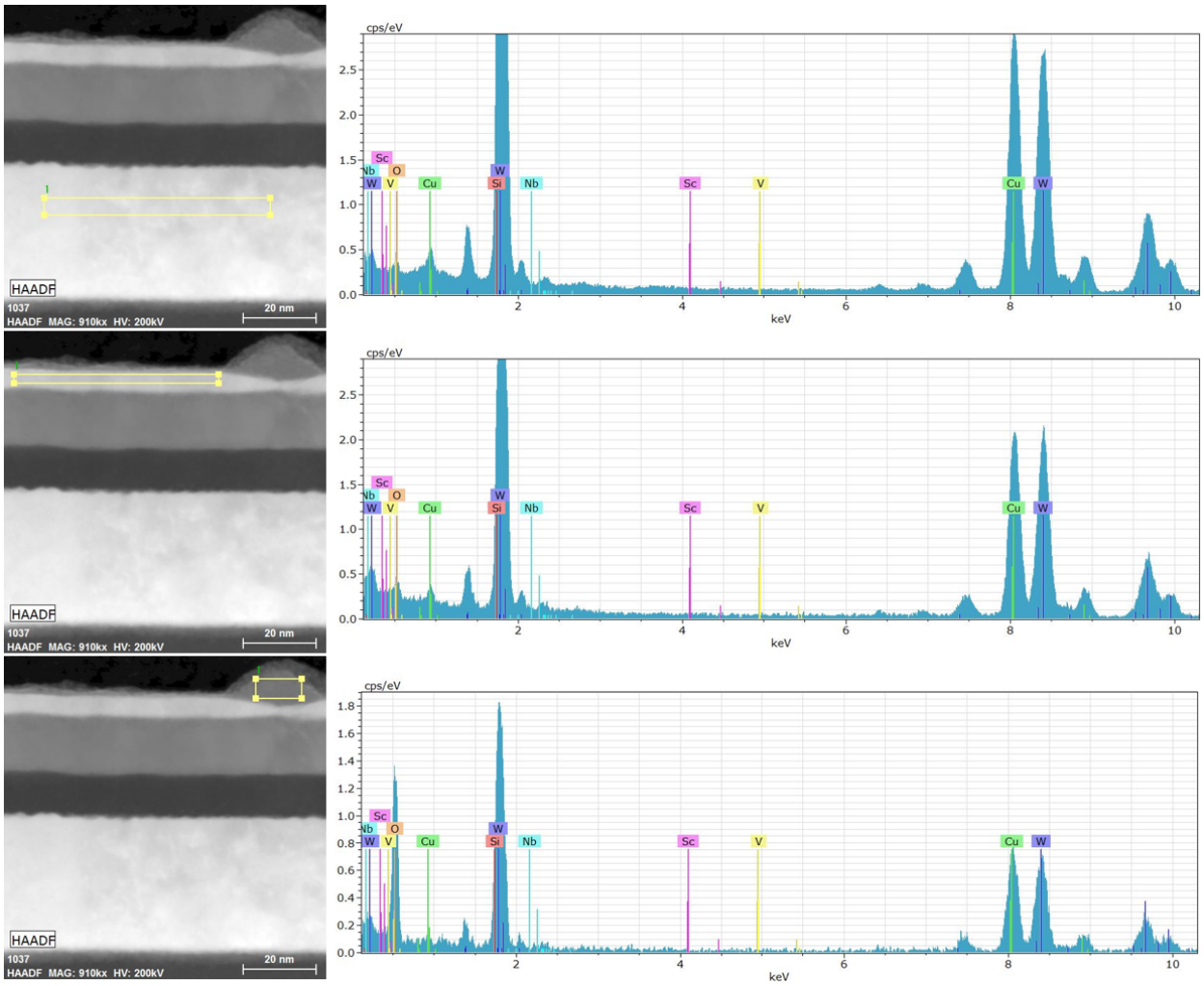


Figure .31: Elemental analysis of the top waveguide layer of the structure of the V/Sc/Nb sample.

The EDS also indicates no intermixing has occured between the layers of interest. Both the Nb and Sc layers do not show any sign of intermixing. The EDS indicates that still in the Sc layer the co-deposited Vanadium is still present. No Vanadium is present in the Nb layer of the system.

As seem in the XSW fit, the segregation seems to have occurred to the middle and bottom interface with a small preference for the bottom interface. This is concluded since the intensity of the Vanadium signal is a multitude larger than it is in the Sc layer.

The EELS analysis confirms the results found with the HAADF-EDS methods. The Tungsten waveguide structure is partly compromised with locally significant oxidation on the top. Using EELS also the oxidation the Scandium scandium layer is seen. The EELS also confirms no intermixing of the layers.

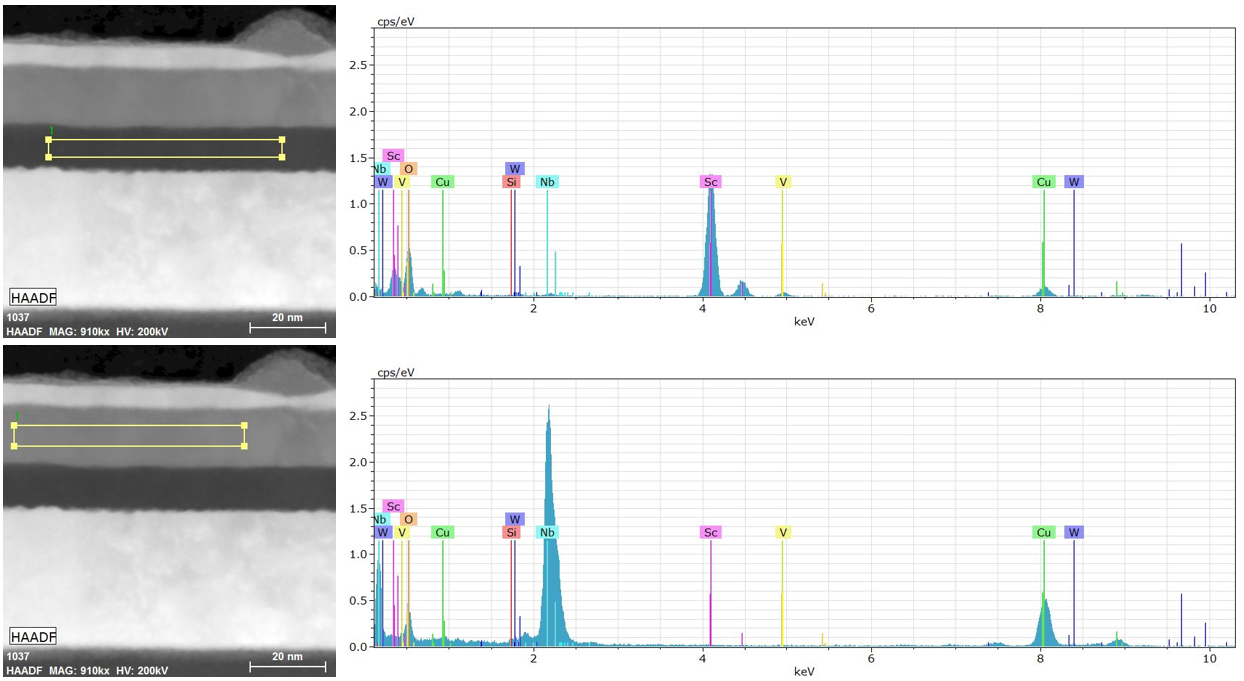


Figure .32: Elemental analysis of the top interface of the structure of the V/Sc/Nb sample.

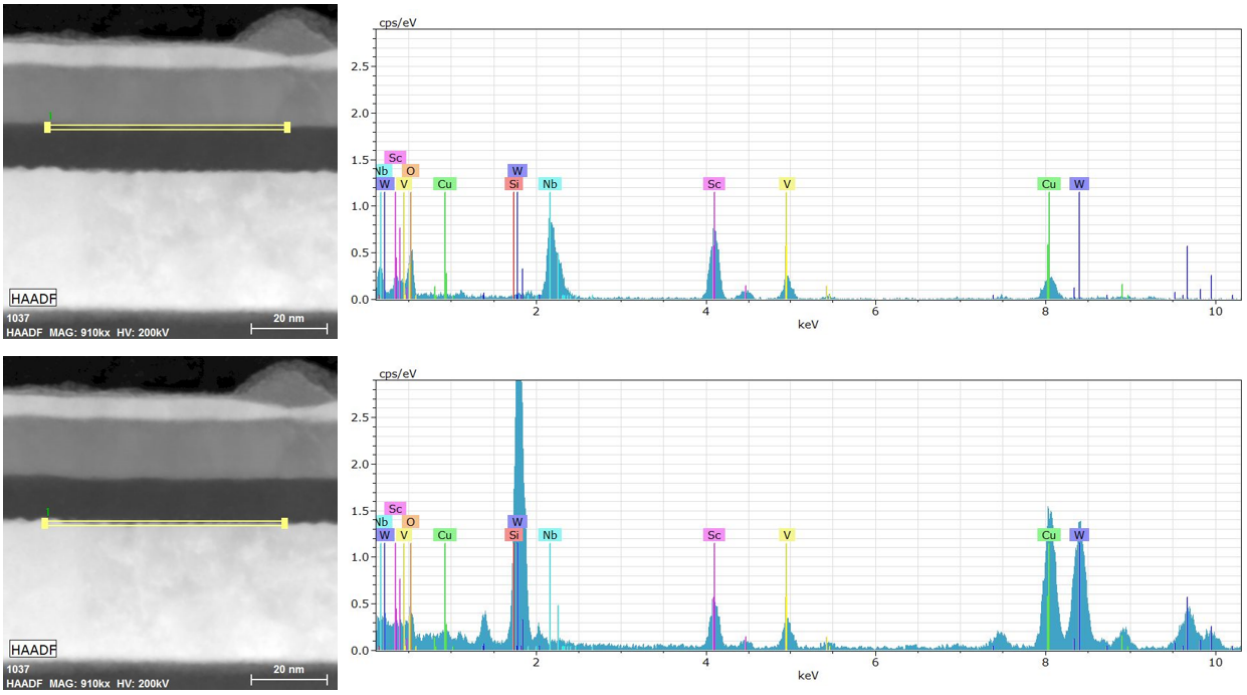


Figure .33: Elemental analysis of the bottom waveguide layer of the structure of the V/Sc/Nb sample.

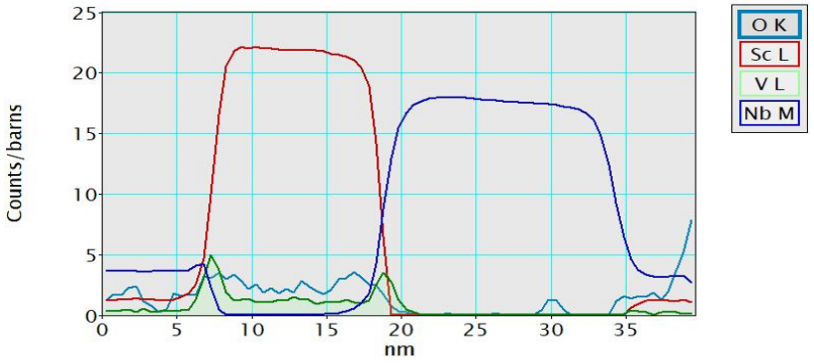


Figure .34: Resulting profiles of the EELS analysis of the annealed V/Sc/Nb sample.

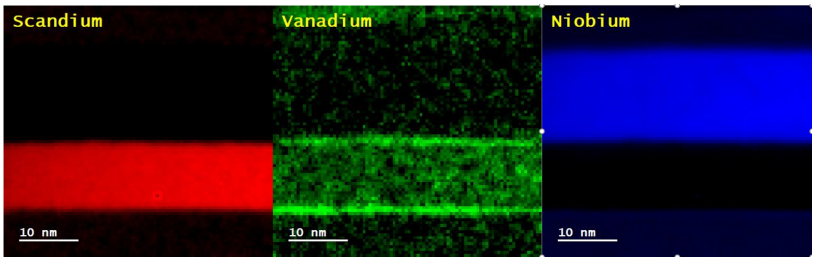


Figure .35: RAW images of EELS analysis of the annealed V/Sc/Nb sample.

The EDS analysis shows similar results. The same patterns of segregation are seen and again no intermixing of Sc and Nb are observed. The oxidation of Sc is visible but for the other layers the background is too high to perceive any grade of oxidation. The Si and C and part of the Cr signal are both caused by the overlap of their fluorescence peaks and background.

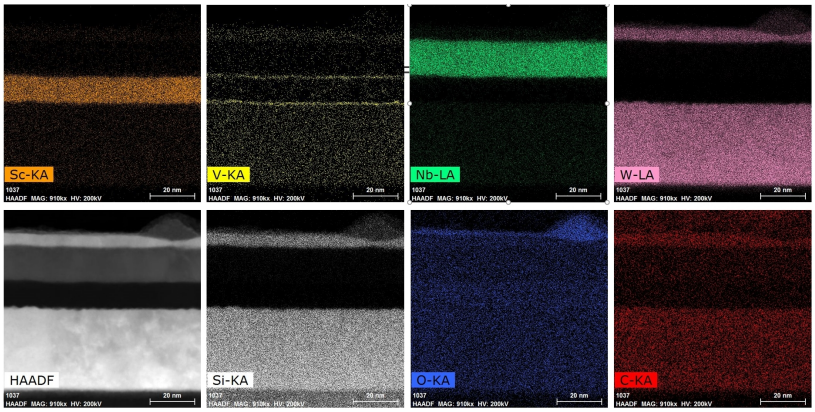


Figure .36: EDS analysis of the annealed V/Sc/Nb sample.

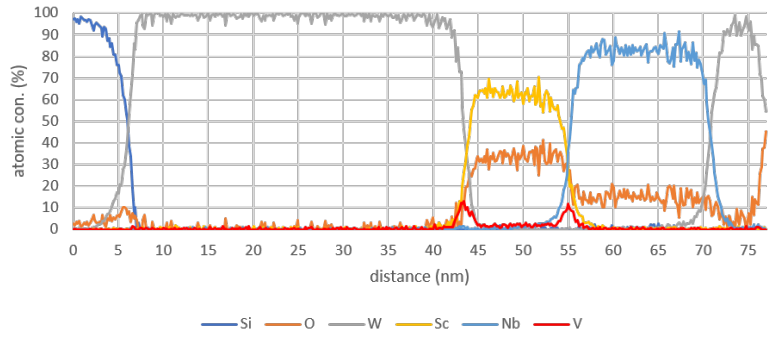


Figure .37: Resulting depth profiles of the different elements resulting from the EDS analysis of the annealed Cr/Fe/Co sample.

.3 Greater Table

PriMat	SecMat	TerMat	Seg	CrystalStruc	AtomicR	SurfaceE	$H_{segTable}$
Cr	Ag	Cu	T,B	BCC FCC FCC	1.28 1.44 1.28	1.2 0.828 0.944	-2.75 -18.9 -88.3
Ti	Zr	Nb	B	HCP HCP BCC	1.47 1.6 1.46	1.42 1.63 1.85	-1.13 -2.81 6.12
Ti	Sc	Nb		HCP HCP BCC	1.47 1.62 1.46	1.42 1.09 1.85	-3.14 -32 -29.7
Cr	Fe	Co	T,B	BCC BCC HCP	1.28 1.26 1.25	1.2 1.28 1.27	0.192 -3.34 3.12
Fe	Cr	Co	-	BCC BCC HCP	1.26 1.28 1.25	1.28 1.2 1.27	1.17 7.23 0.764
Co	Cr	Fe	-	HCP BCC BCC	1.25 1.28 1.26	1.27 1.2 1.28	1.9 7.1 3.07
Cr	Co	Fe	T,B	BCC HCP BCC	1.28 1.25 1.26	1.2 1.27 1.28	2.07 4.85 8.88
Co	Fe	Cr	-	HCP BCC BCC	1.25 1.26 1.28	1.27 1.28 1.2	0.24 -3.21 -0.908
Fe	Co	Cr	-	BCC HCP BCC	1.26 1.25 1.28	1.28 1.27 1.2	2.5 4.72 2.42
Sc	V	Nb	-	HCP BCC BCC	1.62 1.34 1.46	1.09 1.52 1.85	3.07 15.6 2.61
V	Sc	Nb	M,B	BCC HCP BCC	1.34 1.62 1.46	1.52 1.09 1.85	-3.67 -35.1 -22
V	Nb	Sc	-	BCC BCC HCP	1.34 1.46 1.62	1.52 1.85 1.09	3.57 -12.4 11.6
Sc	Nb	V	M	HCP BCC BCC	1.62 1.46 1.34	1.09 1.85 1.52	0.414 -12.8 -1.41
Pd	Co	Fe	T,B	FCC HCP BCC	1.37 1.25 1.26	1.23 1.27 1.28	-2.19 -3.3 -4.83
Pd	Fe	Co	T,B	FCC BCC HCP	1.37 1.26 1.25	1.23 1.28 1.27	-0.641 4.8 -2.44
Fe	Pd	Co	-	BCC FCC HCP	1.26 1.37 1.25	1.28 1.23 1.27	2.35 6.63 13.9
Co	Fe	Pd	-	HCP BCC FCC	1.25 1.26 1.37	1.27 1.28 1.23	1.26 4.16 -0.908
Fe	Co	Pd	-	BCC HCP FCC	1.26 1.25 1.37	1.28 1.27 1.23	1.32 -2.65 2.42
Co	Pd	Ti	B	HCP FCC HCP	1.25 1.37 1.47	1.27 1.23 1.42	-4.39 53.2 8.81
Ti	Co	Pd	-	HCP HCP FCC	1.47 1.25 1.37	1.42 1.27 1.23	-9.7 -42.3 36.5
Ti	Co	Pd	-	HCP HCP FCC	1.47 1.25 1.37	1.42 1.27 1.23	-9.7 -42.3 36.5
Ti	Hf	Ru	B	HCP HCP HCP	1.47 1.59 1.34	1.42 1.71 1.75	-10.5 17.8 1.68
Ti	Ru	Hf	-	HCP HCP HCP	1.47 1.34 1.59	1.42 1.75 1.71	7.28 122 64.6
Ru	Ti	Hf	-	HCP HCP HCP	1.34 1.47 1.59	1.75 1.42 1.71	-2.36 -6.85 51
Ru	Hf	Ti	-	HCP HCP HCP	1.34 1.59 1.47	1.75 1.71 1.42	0.353 6.45 59.5

.4 Derivatives Matlab Implementation

GIXR Derivative Function

```
for n=Nz:-1:1
    A1(n,:) = c11(n,:).*A1(n+1,:) + c12(n,:).*A2(n+1,:);
    A2(n,:) = c21(n,:).*A1(n+1,:) + c22(n,:).*A2(n+1,:);
end

for n=1:1:Nz
    RM11(n+1,:) = RM11(n,:).*c11(n,:)+RM12(n,:).*c21(n,:);
    RM21(n+1,:) = RM21(n,:).*c11(n,:)+RM22(n,:).*c21(n,:);
    RM12(n+1,:) = RM11(n,:).*c12(n,:)+RM12(n,:).*c22(n,:);
    RM22(n+1,:) = RM21(n,:).*c12(n,:)+RM22(n,:).*c22(n,:);
end

for n=2:1:Nz-1
    A1temp = cd11(n,:).*A1(n+1,:) + cd12(n,:).*A2(n+1,:);
    A2temp = cd21(n,:).*A1(n+1,:) + cd22(n,:).*A2(n+1,:);

    A1D(n+1,:) = RM11(n,:).*A1temp + RM12(n,:).*A2temp;
    A2D(n+1,:) = RM21(n,:).*A1temp + RM22(n,:).*A2temp;
end

for n=2:1:Nz-1
    A1temp = cdd11(n,:).*A1(n+1,:) + cdd12(n,:).*A2(n+1,:);
    A2temp = cdd21(n,:).*A1(n+1,:) + cdd22(n,:).*A2(n+1,:);

    A1Dd(n+1,:) = RM11(n,:).*A1temp + RM12(n,:).*A2temp;
    A2Dd(n+1,:) = RM21(n,:).*A1temp + RM22(n,:).*A2temp;
end

ro = (sqrt(epsilon(1))./(sin(theta'))).*A2(1,:)./A1(1,:);
R = ((abs((1-ro)./(1+ro))).^2);

dA1_dq = A1D(3:end,:);
dA2_dq = A2D(3:end,:);
dA1_dd = A1Dd(3:end,:);
dA2_dd = A2Dd(3:end,:);

qDn = sqrt(epsilonDn-epsilon(1)*cos(theta').^2);

dq_dp = (qDn-q)./dp;

dr_dA1 = -(sqrt(epsilon(1))./(sin(theta'))).*(A2(1,:)./ (A1(1,:).^2) );
dr_dA2 = (sqrt(epsilon(1))./(sin(theta'))).*( 1./A1(1,:) );
dr_dq = (dr_dA1.*dA1_dq + dr_dA2.*dA2_dq);
dr_dp = dr_dq.* dq_dp(2:end-1,:);%.*dq_dp
dr_dd = (dr_dA1.*dA1_dd(:,:) + dr_dA2.*dA2_dd(:,:));%.*dq_dp

j=0;
for e = 1:length(BasicMat)
    if BasicMat(e).FitDensity == 1
        j = j+1;
        epsilonPn=(  real(dNPkArray(:,j)) +1i*imag(dNPkArray(:,j))  ).^2;    % Permittivit
        qPn = sqrt(epsilonPn-epsilon(1)*cos(theta').^2);
        dq_dP = (qPn-q)./dp;
        dr_dP(j,:) = sum(dr_dq.*dq_dP(2:end-1,:),1);%*BasicMat(e).Density;
    end
end

nn = 1;
for dd = 1:size(FitMatrix,1)
    next = length(FitMatrix{dd,2});
    dd_dL = FitMatrix{dd,1}/(next);
    dr_dD(dd,:) = sum( dr_dd(nn:nn+next-1,:),1 );%*dd_dL;
    nn = nn + next;
end
```

XSW Derivative Function

```
for n=Nz:-1:1
    A1(n,:) = c11(n,:).*A1(n+1,:) + c12(n,:).*A2(n+1,:);
    A2(n,:) = c21(n,:).*A1(n+1,:) + c22(n,:).*A2(n+1,:);
end

R = [];
absAp2 = (abs( A1(1:end-1,:) ) ).^2;
f = abs(  (  A1(1,:)  + sqrt(epsilon(1)).*A2(1,:)./sin(theta')) /  2*epsilon(1)   ) ;

WF(:,:) = absAp2./(f.^2);

j=0;
for e = 1:length(BasicMat)
    if BasicMat(e).FitDensity == 1
        j = j+1;
        AOI = find((dNPkArray(:,j)-Structure(:,2))~=0, 1, 'last' );
        AOI2 = find((dNPkArray(:,j)-Structure(:,2))~=0, 1, 'first' );
        if isempty(AOI)==0
            epsilonPn=(  real(dNPkArray(AOI2:AOI,j)) + 1i*imag(dNPkArray(AOI2:AOI,j))  ).^2;
            qPn = sqrt(epsilonPn-epsilon(1)*cos(theta').^2);
            A1P(AOI+1,:) = A1(AOI+1,:);
            A2P(AOI+1,:) = A2(AOI+1,:);
            PhidPn=K0*qPn(:,:).*Ld(AOI2:AOI);
            CosPhidPn=cos(PhidPn(:,:));
            SinPhidPn=sin(PhidPn(:,:));
            p11=CosPhidPn(:,:);
            p22=CosPhidPn(:,:);
            p12=(1i./(qPn)).*SinPhidPn(:,:);
            p21=1i.*(qPn).*SinPhidPn(:,:);
            for n = AOI:-1:1
                if n >= AOI2
                    m = n-AOI2+1;
                    A1P(n,:) = p11(m,:).*A1P(n+1,:) + p12(m,:).*A2P(n+1,:);
                    A2P(n,:) = p21(m,:).*A1P(n+1,:) + p22(m,:).*A2P(n+1,:);
                else
                    A1P(n,:) = c11(n,:).*A1P(n+1,:) + c12(n,:).*A2P(n+1,:);
                    A2P(n,:) = c21(n,:).*A1P(n+1,:) + c22(n,:).*A2P(n+1,:);
                end
            end
            WFPn(1:end-1,:,j) = [(abs( A1P(1:AOI,:) ) ).^2;absAp2(AOI+1:end-1,:)]./( ( abs( (A1P(1,:) + sqrt(
     epsilon(1)).*A2P(1,:)./sin(theta')) /  2*epsilon(1)   ) ).^2 );
        else
            WFPn(:,:,j) = WF(:,:);
        end

    end
end

for n = Nn-1:-1:2
    A1N(MazeList(n)+1,:) = A1(MazeList(n)+1,:);
    A2N(MazeList(n)+1,:) = A2(MazeList(n)+1,:);
    for m = MazeList(n):-1:1
        if m >= MazeList(n-1)+1
            A1N(m,:) =  d11(m,:).*A1N(m+1,:) +  d12(m,:).*A2N(m+1,:);
            A2N(m,:) =  d21(m,:).*A1N(m+1,:) +  d22(m,:).*A2N(m+1,:);
        else
            A1N(m,:) = c11(m,:).*A1N(m+1,:) + c12(m,:).*A2N(m+1,:);
            A2N(m,:) = c21(m,:).*A1N(m+1,:) + c22(m,:).*A2N(m+1,:);
        end
    end
    WFDn (1:end-1,:,n-1) = [(abs( A1N(1:MazeList(n),:) ) ).^2;absAp2(MazeList(n)+1:end-1,:)]./( ( abs( (A1N
     (1,:) + sqrt(epsilon(1)).*A2N(1,:)./sin(theta')) /  2*epsilon(1)   ) ).^2 );
end

DL=0;
for n = size(FitMatrix,1):-1:1
    D_Bot = Nn-DL-1;
```

```
66      DL = DL + size(FitMatrix{n,2},2);
67      D_Top = Nn-1-DL;
68      A1D(MazeList(D_Bot)+1,:) = A1(MazeList(D_Bot)+1,:);
69      A2D(MazeList(D_Bot)+1,:) = A2(MazeList(D_Bot)+1,:);
70      for m = MazeList(D_Bot):-1:1
71          if m >= MazeList(D_Top)+1
72              A1D(m,:) = dd11(m,:).*A1D(m+1,:) + dd12(m,:).*A2D(m+1,:);
73              A2D(m,:) = dd21(m,:).*A1D(m+1,:) + dd22(m,:).*A2D(m+1,:);
74          else
75              A1D(m,:) = c11(m,:).*A1D(m+1,:) + c12(m,:).*A2D(m+1,:);
76              A2D(m,:) = c21(m,:).*A1D(m+1,:) + c22(m,:).*A2D(m+1,:);
77          end
78      end
79      WFdDn(1:end-1,:,n) = [(abs( A1D(1:MazeList(D_Bot),:) ) ).^2;absAp2(MazeList(D_Bot)+1:end-1,:)]./( ( abs( (
         A1D(1,:) + sqrt(epsilon(1)).*A2D(1,:)./sin(theta')) /  2*epsilon(1)   ) ).^2 );
80  end
```

Bibliography

Here are the references in citation order.

[1] A.R Miedema. 'A Simple Model for Alloys'. In: *Philips Technical Review* 33.6 (1973), pp. 149–161 (cited on pages 1, 5).

[2] P. M. Ossi. 'Surface segregation in transition metal alloys: Experiments and theories'. In: *Surface Science* 201.3 (1988), pp. 519–531. DOI: 10.1016/0039-6028(88)90491-8 (cited on page 1).

[3] Heather A. Murdoch and Christopher A. Schuh. 'Estimation of grain boundary segregation enthalpy and its role in stable nanocrystalline alloy design'. In: *Journal of Materials Research* 28.16 (2013), pp. 2154–2163. DOI: 10.1557/jmr.2013.211 (cited on page 1).

[4] Yang Hu, Jennifer D. Schuler, and Timothy J. Rupert. 'Identifying interatomic potentials for the accurate modeling of interfacial segregation and structural transitions'. In: *Computational Materials Science* (2018). DOI: 10.1016/j.commatsci.2018.02.020 (cited on page 2).

[5] D. Raabe et al. 'Segregation engineering enables nanoscale martensite to austenite phase transformation at grain boundaries: A pathway to ductile martensite'. In: *Acta Materialia* 61.16 (Sept. 2013), pp. 6132–6152. DOI: 10.1016/j.actamat.2013.06.055 (cited on page 2).

[6] G. Rao, D. B. Zhang, and P. Wynblatt. 'A determination of interfacial energy and interfacial composition in CuPb and CuPbX alloys by solid state wetting measurements'. In: *Acta Metallurgica Et Materialia* (1993). DOI: 10.1016/0956-7151(93)90062-W (cited on page 2).

[7] J. Gerkema and A. R. Miedema. 'Adhesion between solid metals: observations of interfacial segregation effects in metal film lubrication experiments'. In: *Surface Science* 124.2-3 (Jan. 1983), pp. 351–371. DOI: 10.1016/0039-6028(83)90796-3 (cited on pages 2, 4, 5, 53–55).

[8] Z. Erdélyi et al. 'Nanoresolution interface studies in thin films by synchrotron x-ray diffraction and by using x-ray waveguide structure'. In: *X-Ray Spectrometry* 38.4 (July 2009), pp. 338–342. DOI: 10.1002/xrs.1170 (cited on page 2).

[9] I. A. Makhotkin et al. 'Determination of the density of ultrathin La films in La/B 4 C layered structures using X-ray standing waves'. In: *physica status solidi (a)* 208.11 (Nov. 2011), pp. 2597–2600. DOI: 10.1002/pssa.201184256 (cited on page 2).

[10] Andrey Zameshin et al. 'Reconstruction of interfaces of periodic multilayers from X-ray reflectivity using a free-form approach'. In: *Journal of Applied Crystallography* 49.4 (2016), pp. 1300–1307. DOI: 10.1107/S160057671601044X (cited on pages 2–4, 6, 9, 24, 25).

[11] Anton Haase et al. 'Multiparameter characterization of subnanometre Cr/Sc multilayers based on complementary measurements'. In: *Journal of Applied Crystallography* 49.6 (2016), pp. 2161–2171. DOI: 10.1107/S1600576716015776 (cited on page 2).

[12] Radford M. Neal. 'MCMC using hamiltonian dynamics'. In: *Handbook of Markov Chain Monte Carlo* (2011), pp. 113–162. DOI: 10.1201/b10905-6 (cited on pages 2, 14).

[13] Lars Gebraad, Christian Boehm, and Andreas Fichtner. 'Bayesian Elastic Full-Waveform Inversion Using Hamiltonian Monte Carlo'. In: *Journal of Geophysical Research: Solid Earth* 125.3 (2020), pp. 1–18. DOI: 10.1029/2019JB018428 (cited on page 2).

[14] Michiko Yoshitake, Yarrama-Reddy Aparna, and Kazuhiro Yoshihara. 'General rule for predicting surface segregation of substrate metal on film surface'. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 19.4 (July 2002), pp. 1432–1437. DOI: 10.1116/1.1376699 (cited on page 5).

[15] Albert Tarantola. *Inverse problem theory*. 2004, pp. 1–358 (cited on page 8).

[16] M Hadamard and M Hadamard Les. 'Les problemes aux limites dans la theorie des equations aux d'erivees partielles'. In: (1907) (cited on page 9).

[17] Kambe J Keller K Kirchglqsner and B J Matkowsky C S Peskin. *Applied Mathematical Sciences Advisors*. Vol. 139 (cited on pages 9, 10).

[18] Nawaf Bou-rabee. 'Geometric integrators and the Hamiltonian Monte Carlo method'. In: 2018 (2019), pp. 113–206. DOI: 10.1017/S0962492917000101 (cited on pages 11–14, 28).

[19] Michael Betancourt. 'A Conceptual Introduction to Hamiltonian Monte Carlo'. In: () (cited on pages 12, 29).

[20] V. A. Solé et al. 'A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra'. In: *Spectrochimica Acta - Part B Atomic Spectroscopy* 62.1 (2007), pp. 63–68. DOI: 10.1016/j.sab.2006.12.002 (cited on page 21).

[21] Koji Ohta and Hatsuo Ishida. 'Matrix formalism for calculation of electric field intensity of light in stratified multilayered films'. In: *Applied Optics* 29.13 (1990), p. 1952. DOI: 10.1364/ao.29.001952 (cited on page 23).

[22] Steven G. Johnson. 'The NLopt nonlinear-optimization package'. In: *http://github.com/stevengj/nlopt* () (cited on page 27).

[23] Mjd Powell. 'The BOBYQA algorithm for bound constrained optimization without derivatives'. In: *NA Report NA2009/06* (2009), p. 39. DOI: 10.1.1.443.7693 (cited on page 27).

[24] P. Kaelo and M. M. Ali. 'Some variants of the controlled random search algorithm for global optimization'. In: *Journal of Optimization Theory and Applications* 130.2 (2006), pp. 253–264. DOI: 10.1007/s10957-006-9101-0 (cited on page 27).