



UNIVERSITÉ DE MONTPELLIER

FACULTÉ DES SCIENCES

COMMISSARIAT A L'ENERGIE ATOMIQUE ET
AUX ENERGIES ALTERNATIVES

**Report of the Alternance program for the Masters
second year in Computational Physics**

Author :

Nischal Dhungana

Under the supervision of :

Guillaume Freychet

June 2024

Acknowledgements

Contents

1	Introduction	4
2	Context	6
2.1	Host organisation	6
2.1.1	History	6
2.1.2	Sectors of activity	7
2.1.3	Future and oreintation strategy	8
2.2	Project description	9
2.3	Objectives	10
3	CD-SAXS	12
3.1	Introduction	12
3.2	Scattering Model	13
3.3	Experimental setup	14
3.4	Fitting Algorithm	14
3.4.1	Covariance Matrix Adaptation Evolution Strategy (CMAES)	16
3.4.2	Uncertainty estimation by Monte Carlo Markov Chain(MCMC) method	18
3.4.3	Overall view	19
4	My contribution	22
4.1	Introduction	22
4.2	Design	22
4.2.1	Components	22
4.2.2	Relationships between components	24
4.2.3	Why this design?	25
4.3	Simulation Models	26
4.3.1	Stacked Trapezoid Model	27
4.4	Implementation of MCMC	27
4.5	Optimisation and parrellelisation	27
4.6	On the fly uncertainty estimation	27
5	Future Prospects	28

1 Introduction

The continuous miniaturization of microelectronic components, driven by Moore’s Law, has led to a significant reduction in transistor size and increased chip complexity. This rapid advancement has presented new challenges in the field of metrology, the science of measurement. Existing metrology techniques, such as Optical Critical Dimension (OCD) and Critical Dimension Scanning Electron Microscope (CDSEM), are reaching their limits in terms of resolution and accuracy as feature sizes shrink to the nanometer scale.

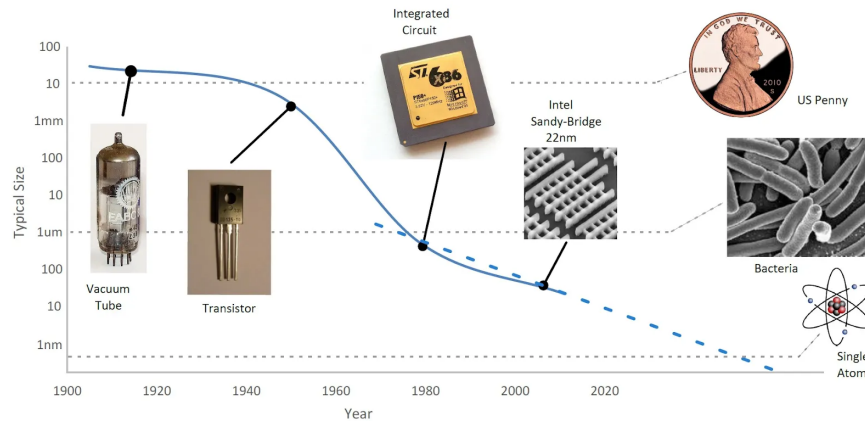


Figure 2: Evolution of microelectronics and the need for advanced metrology techniques [1].

To address these challenges, a new metrology technique called Critical Dimension Small Angle X-ray Scattering (CDSAXS) is being developed. CDSAXS utilizes short-wavelength X-rays ($\lambda \approx 0.05 - 5nm$), to probe the internal structure of materials, providing high-resolution measurements of critical dimensions (CDs) with greater accuracy than conventional methods. CEA-Leti, a leading research institute in microelectronics, is actively involved in the development of CDSAXS technology.

This work-study project focused on the development of a coherent software for the fit and analysis of CDSAXS data. The software aims to streamline the data processing workflow and enhance the accuracy of CD measurements. The project involved a comprehensive understanding of CDSAXS theory, data collection procedures, and fitting algorithms.

The report begins with an overview of the context of the project, highlighting the evolution of microelectronics and the need for advanced metrology techniques. It then delves into the CDSAXS technique, explaining the principles, data collection, fitting, and analysis. Then the subsequent section describes the software development process, outlining the software’s functionalities and design. Finally, the report concludes with a summary of the project’s achievements and outlines potential future directions.

2 Context

2.1 Host organisation

The French Alternative Energies and Atomic Energy Commission (CEA) stands as a cornerstone of the nation’s research landscape. Its multifaceted expertise encompasses a broad spectrum of fields, including nuclear energy, renewable energy, technological research for industry, material sciences, health and life sciences, and defense and security. The CEA’s network of research centers spans across France, each with its unique specializations and areas of excellence. Among these, CEA Grenoble holds a prominent position, where I have the privilege of pursuing my work-study program. I am part of the Leti Institute, a research center dedicated to microelectronics and nanotechnologies. More specifically, I was with “Materials and Structures Properties Laboratory” (MSPL) which is under “Technology Platforms Department” (TPFD), one of six different departments of Leti.



Figure 3: CEA Grenoble Campus

2.1.1 History

The French atomic energy commission, CEA, was born in 1945 after World War II. Its mission was to develop nuclear expertise for France. Pioneering scientists like Frédéric Joliot-Curie

and Francis Perrin led the way in building research reactors and nuclear power plants.

The CEA didn't stop at just nuclear energy. In the 1960s, they began to diversify into new areas like renewable energy, micro and nanotechnologies, defense, and healthcare. This diversification led to the creation of specialized research centers, including the future innovation hub, CEA Grenoble.

It was founded in 1956 by Nobel laureate physicist Louis Néel. He saw the scientific potential of the Grenoble region and his vision proved to be true. The center grew rapidly in the following decades, attracting talent and investment from around the world.

It came to be known as France's "atomic capital" due to its research reactors. However, their influence went far beyond nuclear. They developed their first integrated circuit in 1965, launching their journey into micro and nanotechnologies. They also played a key role in creating Minatec, the first European hub for excellence in this field. In addition, they became a leader in renewable energy research with the Institut national de l'énergie solaire (Ines).

Today, it is a research powerhouse with over 2,500 researchers and technicians. Their campus houses specialized institutes in various fields, from healthcare to digital technologies. It's also the headquarters for CEA Tech, the technological branch of the CEA with over 4,500 researchers across France.

2.1.2 Sectors of activity

CEA Grenoble plays a pivotal role in the nation's economic and technological advancement through its groundbreaking research and innovations across diverse fields.

- **Energy and Sustainability:** Supporting current and future nuclear power, exploring solar, hydrogen for carbon neutrality (2050), researching SMRs and thermonuclear fusion.
- **Digital Technologies:** Contributing through the SPIN program for spintronics (frugal, agile, sustainable computing) aligned with France 2030 plan.
- **Healthcare:** Distinguished research in biology and biotechnology for health, addressing current and future challenges. (e.g., Laboratoire de biologie et biotechnologie pour la santé)
- **Defense:** Traditionally significant role, developing cutting-edge technologies for national security and defense (less documented for Grenoble).



Figure 4: CEA Grenoble campus before and after

CEA Liten for example also serves as an innovation hub for new energy technologies and nanomaterials, emphasizing energy diversification and renewable energy integration. Their research encompasses solar photovoltaics, energy storage, and transportation (hydrogen and fuel cells).

2.1.3 Future and orientation strategy

CEA stands as a powerhouse for innovation in France. Spanning energy, healthcare, defense, and digital technologies, the CEA pushes boundaries by collaborating with universities and industry on ambitious R&D projects.

Their focus is clear: develop transformative solutions for global challenges. This includes renewable energy sources, innovative healthcare systems, advanced defense solutions, and disruptive digital technologies. Sustainability is paramount, with research prioritizing energy efficiency and minimizing environmental impact.

The CEA partners with leading institutions to accelerate technology transfer and create open innovation ecosystems. These partnerships combine expertise, resources, and networks, leading to breakthrough technologies with significant economic and social value.

CEA Grenoble exemplifies this innovative spirit. They focus on cutting-edge technologies like AI, advanced materials, nanotechnologies, and quantum technologies. Their research aims to revolutionize industries and improve lives, from developing next-generation batteries to creating innovative medical devices and advancing digital technologies.

The challenges they tackle are vast: climate change, the energy transition, emerging diseases, cybersecurity, and technological sovereignty. The CEA goes beyond just solutions; they strive to influence public policy and raise awareness. Their research is guided by a long-term vision, anticipating future challenges and preparing for them through innovation.

The CEA is a vital force in shaping a sustainable future. Their commitment to innovation promises clean energy sources, advanced healthcare, robust national security, and a cutting-edge digital landscape. By working collaboratively and addressing global challenges head-on, the CEA positions itself as a leader in the global scientific and technological landscape.

2.2 Project description

CEA Grenoble, a frontrunner in material metrology and characterization, recognizes the potential of CD-SAXS for analyzing nanostructures. To this end, CEA has actively invested in CD-SAXS development for several years.

During my work-study program, I contributed to this initiative by developing a Python application for data fitting and analysis obtained through CD-SAXS. This technique relies on an inverse algorithm that translates scattering intensity data into a relevant real-space structure. The algorithm simulates the experiment using a model and iteratively compares the simulated data with the experimental data until a good fit is achieved. A robust codebase is essential for efficient CD-SAXS data processing.

Previously, a rudimentary collection of functions developed at Brookhaven served as the foundation for CD-SAXS analysis. However, it lacked the coherence of a well-structured application. The code for various data simulation models was similarly disorganized. My primary task was to develop a user-friendly Python application that integrates these functions, streamlining data fitting and result analysis.

Furthermore, the existing code suffered from slow execution speeds due to a lack of optimisation. To address this, I implemented optimisations and parallelisation techniques, significantly improving the code's efficiency.

2.3 Objectives

The initial aim of this project was commendable - to improve the existing CD-SAXS data analysis codebase. However, as we delved deeper, the project's objectives evolved into a series of well-defined, targeted enhancements. This iterative approach ensured that our efforts addressed the specific needs of researchers at CEA Grenoble.

Here's a breakdown of the key objectives that emerged:

- **Unleashing Computational Power:**

- **Vectorization for Server Optimization:** Standard code often struggles to fully leverage the parallel processing capabilities of modern servers. We identified the need to vectorize the code, allowing it to efficiently utilize the computational power available on CEA's powerful servers. This significantly boosted the application's overall processing speed.
- **GPU Acceleration - Pushing the Limits:** Recognizing the ever-increasing capabilities of Graphics Processing Units (GPUs), we explored the potential of integrating GPU acceleration into the code. This would potentially unlock even faster performance, enabling us to tackle increasingly complex datasets.

- **User-Centric Design:**

- **Building a Coherent Application:** The existing code, while functional, lacked the user-friendliness and intuitiveness necessary. We prioritized creating a well-structured, modular application. This would not only simplify data fitting and analysis but also make future feature additions and maintenance easier and more efficient.
- **Addition of different simulation models in same code base:** The original codebase had a only one simulation model, the stacked trapezoid model. We aimed to integrate overlay and rounded trapezoid model into a single, cohesive codebase. This consolidation would streamline the process of selecting and applying different models, enhancing the overall user experience.

- **Quantifying Uncertainty - Confidence in Results:**

- **MCMC for Uncertainty Estimation:** A critical component missing from the original code was the ability to estimate the uncertainty associated with the fitted parameters. We addressed this by implementing a Markov Chain Monte Carlo (MCMC) algorithm. This powerful technique allowed us to generate a statistically

representative set of solutions, enabling us to accurately estimate the level of uncertainty in the fitted parameters.

- **Real-Time Uncertainty Estimation - Saving Valuable Time:**

- **On-the-Fly Uncertainty:** Taking the concept of uncertainty estimation further, we aimed to integrate this feature into the data acquisition process itself, specifically during synchrotron measurements. This would allow researchers to estimate the uncertainty in real-time. By setting a desired uncertainty threshold, the system could potentially stop the experiment once this level of certainty is achieved. This innovation has the potential to save valuable synchrotron beamtime, a precious resource for researchers.

This refined set of objectives ensured that the project delivered valuable enhancements tailored to the needs of CEA researchers. The project not only improved the code's performance but also transformed it into a user-friendly and powerful tool for analyzing CD-SAXS data with greater confidence.

3 CD-SAXS

3.1 Introduction

Miniaturizing transistors, the building blocks of integrated circuits, is getting tougher for the semiconductor industry. Shrinking their size and spacing (pitch) brings not only manufacturing hurdles but also a metrology problem. Precisely measuring these features during production is crucial for high-quality chips. Existing in-line metrology techniques, like optical critical-dimension (OCD) scatterometry and critical-dimension scanning electron microscopy (CD-SEM), are nearing their limits [2, 3]. OCD struggles with limitations inherent to light and the ever-shrinking features. CD-SEM offers valuable insights but is restricted by the sampling area. To overcome these obstacles, the industry is exploring X-ray-based metrology. X-rays have much shorter wavelengths than the features being measured, allowing for more precise analysis. Additionally, they are sensitive to variations in composition, providing a richer data set.

Early research on X-ray characterization of patterned nanostructures used reflection methods like X-ray diffraction (XRD) and grazing-incidence small-angle X-ray scattering (GISAXS). These techniques demonstrated X-ray’s sensitivity to features’ shape and spacing. Furthermore, X-rays can probe buried features due to their sensitivity to composition via electron density. For instance, a GaInAs/InP multilayer was studied with high-resolution XRD, revealing sensitivity to both the grating and strain between layers [4].

For sub-100 nm features, small-angle scattering methods like GISAXS become more practical. GISAXS examines nanostructures across large areas, making it a potential metrology tool for semiconductors. It uses X-rays near the critical angle of the probed film resulting in a large sampling area and statistically significant data. This large area allows for faster measurements, enabling in situ kinetic studies.

However, GISAXS limitations include a large spot size (50-100 mm wide by 5-10 mm long) and computational challenges for complex nanostructure modeling. This limits its use for semiconductor metrology to simple, large-area patterns like memory arrays (Hofmann et al., 2009; Scholze et al., 2011). Logic devices require smaller probing areas due to test structure size and the complexity of the multicomponent, 3D nanostructures. [5]

Unlike other X-ray techniques, CDSAXS uses a transmission geometry, enabling a much smaller spot size compared to methods like GISAXS. This allows for more precise measurements on smaller features. Studies have shown CDSAXS’s effectiveness in characterizing the shape and spacing of nanometer-sized patterns [6].

CDSAXS utilises variable-angle transmission scattering. By rotating the sample, it can probe the vertical profile of the nanostructures. This allows for reconstructing their shape and composition in two or even three dimensions. We can think of it as a diffraction experiment for single crystals, but instead of a crystal, the periodic array of nanostructures acts like one. This technique excels at reconstructing intricate shapes smaller than 15 nm and with spacings around 30 nm, dimensions crucial for the semiconductor industry. [7]

3.2 Scattering Model

We can represent the diffraction of a collimated X-ray beam by:

$$I(\mathbf{Q}) = \Omega |F(\mathbf{Q})|^2, \quad (1)$$

where $I(\mathbf{Q})$ represents the scattered intensity as a function of the scattering vector \mathbf{Q} , \mathcal{F} is a constant independent of \mathbf{Q} , and $F(\mathbf{Q})$ is the Fourier transform of a function describing the mass distribution within the nanoimprinted pattern.

This relationship is considered valid within the limitations of the CDSAXS geometry, which includes a transmission geometry and a low probability of multiple scattering. Unfortunately, the conjugate product in the equation leads to a loss of phase information, making it impossible to analytically extract $F(\mathbf{Q})$ from $I(\mathbf{Q})$. Therefore, the primary method for determining feature dimensions involves constructing a real-space model of the pattern's cross-section. The Fourier transform of this model is then fitted to the experimental CDSAXS data.

During my work-study program, we were mainly concerned with lines of nanostructures (see figure 6). The cross-section of these can be represented as a stack of trapezoids. We

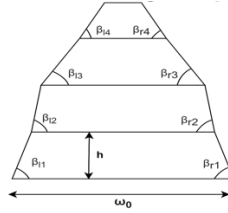


Figure 5: Cross-section of a nanostructure line represented as a stack of trapezoids.

can calculate the analytical fourier transform of a trapezoidal shape to use for fitting can be calculated is given by expression [6]:

$$F(q_x, q_z) = \frac{1}{q_x} \left[-\frac{m_1}{t_1} e^{-iq_x(\frac{\omega_0}{2})} \left(1 - e^{-ih(\frac{q_x}{m_1} + q_z)} \right) + \frac{m_2}{t_2} e^{-iq_x(\frac{\omega_0}{2})} \left(1 - e^{-ih(\frac{q_x}{m_2} + q_z)} \right) \right] \quad (2)$$

where,

$$m_1 = \tan(\beta_1), m_2 = \tan(\pi - \beta_r),$$

$$t_1 = q_x + m_1 q_z, t_2 = q_x + m_2 q_z$$

so,

$$I_0(\mathbf{q}) = |F(q)|^2 \quad (3)$$

An additional decay of scattered intensity $I(Q_x)$ is expected beyond that predicted by the trapezoidal model. This arises from the distribution of periodicities within the sample. This distribution can be caused by two factors:

- **Random variations in average line position:** In this case, the line width remains constant, but the average position of the lines fluctuates slightly across the sample.
- **Variations in line width:** Here, the line width itself varies, which also affects the periodicity.

Both factors indicate a degree of long-range order within the pattern. Additionally, they provide insights into specific types of line edge roughness [8].

To account for this distribution, we introduce an effective Debye-Waller factor, similar to the one used for fluctuations in crystal lattices.

Hence,

$$I(\mathbf{q}) = I_0(\mathbf{q}) \exp(-q^2 DW^2) \quad (4)$$

where DW is the Debye-Waller factor.

3.3 Experimental setup

3.4 Fitting Algorithm

While CDSAXS excels at detecting deviations from a perfect grating pattern in buried structures, it requires additional processing to convert the raw data into a meaningful real-world structure. This process involves using an inverse algorithm, which essentially translates the scattered intensity information back into the original structure's characteristics.

However, there's a catch. Traditional optimization methods used for refinement often fall short when dealing with complex internal structures with numerous parameters. These methods rely on iteratively simulating scattering data and comparing it to the measured data. Unfortunately, this approach can be very time-consuming, especially for intricate structures.

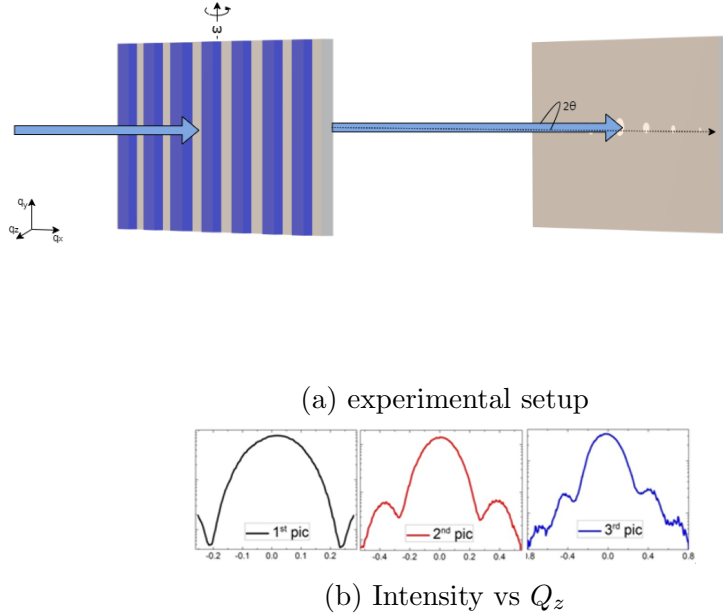


Figure 6: In a is the schematic of the CD-SAXS instrument layout. The X-ray beam (thick solid lines) is transmitted through the patterned sample. Scattered intensity (I) is measured by a 2D detector as a function of scattering angle (2θ) and converted to $I(Q_x)$, where Q_x is defined in the text. Measurements are performed at various sample incidence angles (θ'). In b After conversion from the (Q_x, ω) plane, the intensities from a trapezoidal cross section would appear as predicted in the model calculation on b, plotted as I as a function of the Fourier component, Q_z .

Another challenge arises from the possibility of "degenerate" solutions. These occur when multiple structural models can produce the same scattering data, making it difficult to pinpoint the true structure. This is a common issue in scattering analysis.

Therefore, the ideal scenario for CDSAXS analysis involves an optimization algorithm that can consistently and rapidly converge on the best possible fit for the data. While some prior knowledge about the underlying structure can accelerate the process, such information isn't always readily available. This highlights the need for more efficient algorithms that can handle complex structures even with limited prior knowledge.

Previous research has explored various algorithms to determine the optimal set of parameters for a model that best fits the measured CDSAXS data. These parameters essentially describe the actual structure of the nanostructure being analyzed.

One approach utilizes a Markov chain Monte Carlo (MCMC) algorithm. However, this method requires a good initial guess for the structure's parameters and limitations on their

search range. Additionally, it necessitates multiple independent runs to ensure the algorithm converges on the correct solution. While this approach can be effective, the need for tight parameter bounds might overlook potential fabrication errors in the sample.

Another strategy involves massive computing resources with parallelization and highly refined grid-based models. This method, known as reverse MCMC, offers greater accuracy but is limited by the availability of such computational power.

Genetic and evolutionary algorithms have emerged as promising alternatives. These methods mimic biological evolution, with the model parameters acting as the "genetic code." Starting with randomly generated parameters, these algorithms iteratively refine them through a "mixing strategy" over multiple generations until the optimal set is found. This approach excels at searching large parameter spaces with wide bounds, making it suitable for complex structures. [9]

3.4.1 Covariance Matrix Adaptation Evolution Strategy (CMAES)

One such algorithm is the Covariance Matrix Adaptation Evolution Strategy (CMAES). This method is particularly well-suited for high-dimensional optimization problems, making it ideal for complex nanostructure analysis. CMAES operates by maintaining a population of candidate solutions, with each iteration generating new candidates based on the previous generation's performance. By adapting the covariance matrix of the candidate solutions, CMAES can efficiently explore the parameter space and converge on the optimal solution.

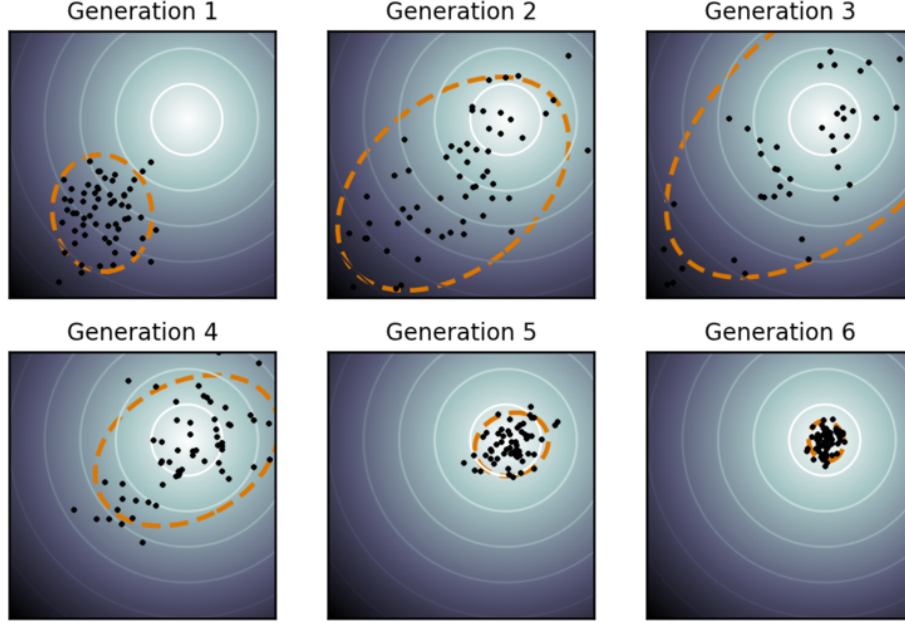


Figure 7: Illustration of CMAES algorithm. The algorithm maintains a population of candidate solutions, with each iteration generating new candidates based on the previous generation's performance.(image taken from wikipedia CMAES page)

The code for CD-SAXS generates a set of possible solutions for the parameters of the model. Then we calculate the analytical Fourier transform of the model and compare it with the experimental data. To compare the two, we use mean-absolute error log:

$$\Xi = \frac{1}{N_q - 1} \sum_{\mathbf{q}} |\log_{10} I_{\text{Sim}}(\mathbf{q}) - \log_{10} I(\mathbf{q})| \quad (5)$$

where $I_{\text{Sim}}(\mathbf{q})$ is the simulated intensity and $I(\mathbf{q})$ is the experimental intensity.

We call it the goodness of fit. The algorithm then tries to minimize this quantity by adjusting the parameters of the model.

In an article [9], researchers investigated the efficiency of various algorithms for reconstructing various nanostructures using X-ray scattering data. Their findings specifically highlighted the advantages of the CMAES. Compared to other methods like Markov Chain Monte Carlo (MCMC) and Differential Evolution (DE), CMAES demonstrated significantly faster convergence times when analyzing an experimental structure. Notably, CMAES achieved a solution that matched the quality of previous studies in approximately 1 to 2 orders of magnitude less time than MCMC and less than an order of magnitude time than DE. This speed advantage held true regardless of the specific objective function used to evaluate the goodness of fit. These results suggest that CMAES offers a powerful tool for analyzing complex nanos-

structures with X-ray scattering data, particularly when dealing with limited computational resources or tight time constraints.

3.4.2 Uncertainty estimation by Monte Carlo Markov Chain(MCMC) method

The CMAES algorithm provides a single best-fit solution for the nanostructure parameters. However, it's essential to understand the uncertainty associated with these parameters. This uncertainty relates to the different possible combinations of parameters that could result in a similar goodness of fit. For instance, decreasing slightly height of one trapezoid and increasing the height of other one can result in a similar goodness of fit. To address this, we can use MCMC algorithm to explore and find all the sets of population that can result in the same goodness of fit.

This confidence interval then can be calculated from this population of parameters. This interval gives us an idea of the range of possible values for each parameter that could still provide a good fit to the data. The lower and upper bounds of this interval can be used to estimate the uncertainty associated with each parameter.

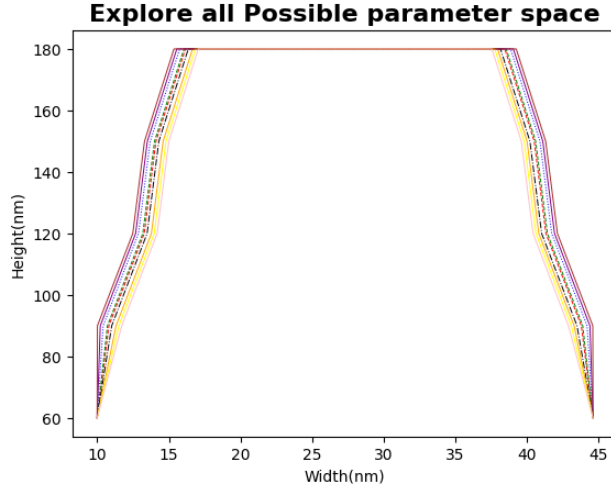


Figure 8: MCMC algorithm exploring all possible sets of parameters that can result in the same goodness of fit.

After determining the best-fit model structure, the researchers of this article [10] employed a MCMC algorithm to calculate the uncertainties associated with the model parameters. This algorithm generates a population of models that can be evaluated to assess the uncertainty in a set of parameters. The probability of a given model being accepted into the population depends on how well the simulated scattering profile it generates matches the experimentally measured one.

Here is the overview of there algorithm:

1. **Seeding:** The algorithm initializes with the model (M) exhibiting the best known goodness-of-fit, denoted GFB.
2. **Proposal generation:** Random perturbations are applied to each parameter within the model, resulting in a new candidate model (M_i) and its corresponding goodness-of-fit GF (GF_i).
3. **Acceptance for better fit:** If $GF_i < GF_{i-1}$, then M_i is automatically accepted into the population (and GFB is updated to $GFB = GF_i$ if the new model has a better fit).
4. **Metropolis-Hastings acceptance for worse fit:** If $GF_i > GF_{i-1}$, the probability (P_i) of accepting M_i is calculated using Eq. (6):

$$P_i = \exp(-0.5 \cdot (GF_i - GFB)) \quad (6)$$

A random number α between 0 and 1 is then generated. If $\alpha < P_i$, M_i is accepted; otherwise, it is rejected, and a new proposal is generated from M_{i-1} .

5. **Equilibrium and resampling:** Steps 2-4 are repeated until the model population reaches equilibrium. To avoid correlations between accepted models, the population is resampled (e.g., every 50 steps in this case).
6. **Uncertainty calculation:** The uncertainties are calculated from the final accepted model population and represent 95% confidence intervals. The step size for generating proposals was optimized to achieve an acceptance probability between 0.25 and 0.35, a range known to yield the fastest convergence.

I used a very similar algorithm but with some modifications to increase the efficiency, we will discuss this in the next section.

3.4.3 Overall view

The process begins with an initial guess for the parameters of the model, which describe a geometric structure with specific width and height parameters. The model data is then transformed into the frequency domain using a Fourier Transform, and this transformed data is compared with the experimental data to optimize the fit by adjusting the parameters and reducing the error between the simulated and experimental data.

The error between the experimental and simulated data is quantified using an error metric that measures the goodness-of-fit by comparing the logarithms of the simulated and experimental intensity profiles. If the error is within a tolerable range, the fit is considered

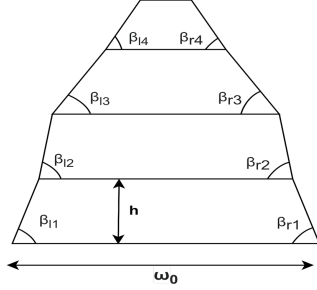
acceptable, and the best-fit model is extracted. This involves identifying the parameters that provide the best match to the experimental data.

To quantify the uncertainties associated with the fitted parameters, a Monte Carlo approach is employed. This involves exploring the possible parameter space through a series of stochastic simulations. The MCMC method is used to sample the parameter space, ensuring that the distribution of the sampled parameters reflects their likelihood given the data. During the MCMC process, initial samples are generated from the best-fit parameters, new samples are proposed by perturbing the current parameters, and the Metropolis-Hastings criterion is used to decide whether to accept or reject the new samples based on their goodness-of-fit compared to the previous samples. This process is repeated until a sufficient number of samples are collected, ensuring the sampled parameter distribution converges to the true posterior distribution.

The final step involves calculating the error bars (uncertainties) for the fitted parameters based on the Monte Carlo samples, providing 95% confidence intervals and a robust measure of the reliability of the fitted model parameters. By combining CMAES for initial parameter optimization and MCMC for uncertainty quantification, the algorithm offers a robust approach to model fitting, ensuring both optimal parameter estimation and reliable uncertainty analysis.

Here is a figure that shows the overall view of the CD-SAXS algorithm:

1. Start with best guess parameter

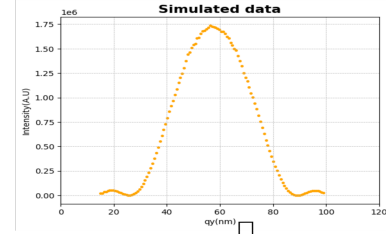


$$m_1 = \tan(\beta_1) \quad m_2 = \tan(\pi - \beta_r)$$

$$t_1 = q_x + m_1 q_z$$

$$F(q_x, q_z) = \frac{1}{q_x} \left[-\frac{m_1}{t_1} e^{-i q_x \left(\frac{\omega_0}{2}\right)} \left(1 - e^{-i h \left(\frac{q_x}{m_1} + q_z\right)}\right) + \frac{m_2}{t_2} e^{-i q_x \left(\frac{\omega_0}{2}\right)} \left(1 - e^{-i h \left(\frac{q_x}{m_2} + q_z\right)}\right) \right]$$

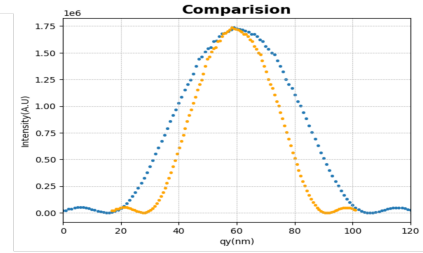
2. Fourier Transform



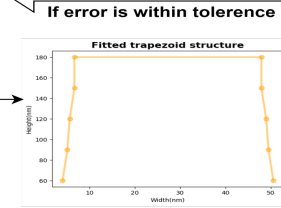
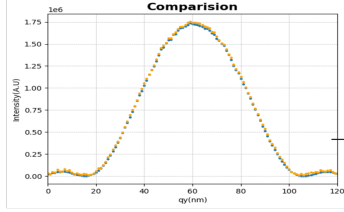
Compare with Experimental Data

3. Error Calculation

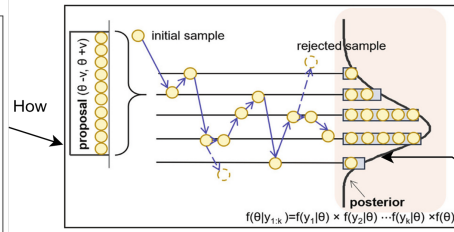
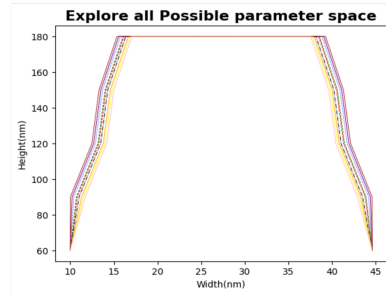
$$\Xi_{G,C} = \frac{1}{N_q - 1} \sum_{\vec{q}} \left| \log_{10} I_{\text{Sim},G,C}(\vec{q}) - \log_{10} I_{\text{Tar}}(\vec{q}) \right|$$



4. Extract The Best Fit



5. Calculate Error Bars With Monte Carlo



$$Z \in g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{if } z \in [\frac{1}{a}, a] \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Metropolis Hastings } P_i = e^{-0.5(GF_i - GF_B)}$$

$$\text{Stretch Move } P_i = e^{-Z^{1-N}(GF_i - GF_B)}$$

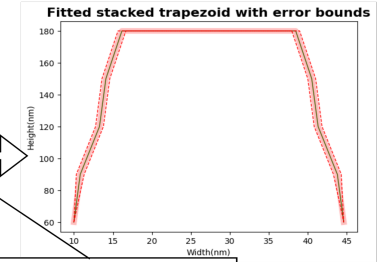


Figure 9: This algorithm integrates CMAES for initial parameter optimization and MCMC for uncertainty quantification. By combining these methods, the algorithm provides a robust approach to model fitting, ensuring both optimal parameter estimation and reliable uncertainty analysis.

4 My contribution

4.1 Introduction

After outlining the context of the work-study program and describing the technique I worked on, it is finally time to discuss my contributions to the project. First, I will explore the design I implemented and explain why this specific design was chosen. Next, we will examine the simulation models I developed and their applications. Following this, I will detail the optimization and parallelization of the code that was done to enhance its performance. Finally, we will discuss how this optimized code could be utilised to estimate uncertainty in real-time during synchrotron experiments.

4.2 Design

To gain a clear understanding of how this system works, let's visualize the design with the help of a UML diagram (see Figure 10). This diagram offers a roadmap, highlighting the various components and their interactions. We'll then delve deeper to explore each component's role in the system.

4.2.1 Components

Fitter:

This class is a crucial component of the design. It includes the *cmaes* function for estimating the best-fit parameters and the *mcmc* function for assessing the uncertainty in the fit. The class takes a simulation model and experimental data as input. When the *cmaes* function is called, it returns the best-fit parameters. Subsequently, the *mcmc* function can be invoked to provide statistical information about the best fit, including the uncertainties in the parameters.

Residual:

This class calculates the residuals between the experimental data and the model. Currently, we use the log-likelihood as the residual function, but it can be easily extended to other residual functions. The *Fitter* class calls this class and provides the relevant model. The *Residual* class then uses the model's *simulate_diffraction* function to calculate the model diffraction pattern and compare it with the experimental data.

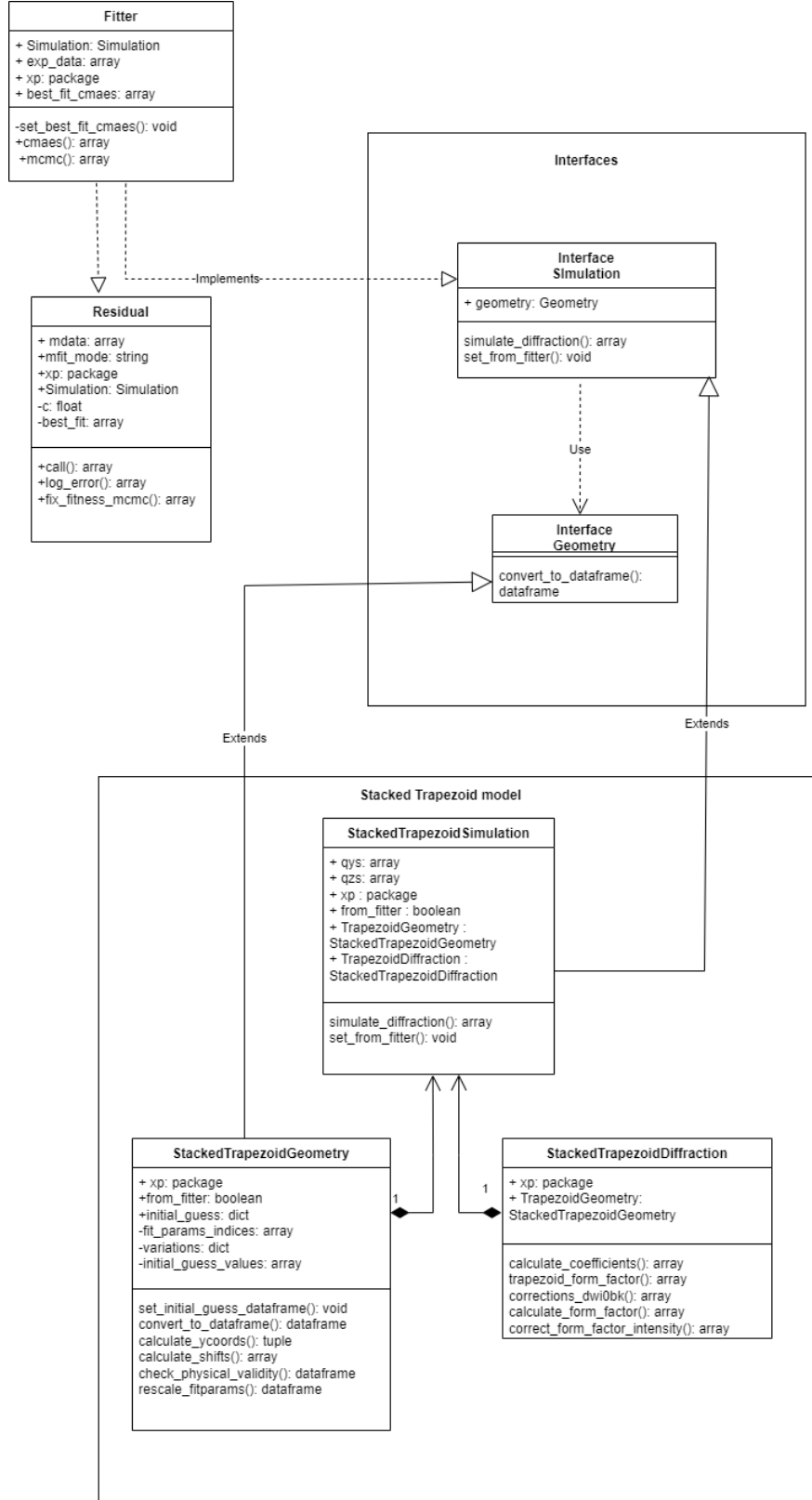


Figure 10: UML diagram of the design for CD-SAXS simulation application.

Simulation interface:

This is the base class for all simulation models. The functions and classes used in it should be implemented in all simulation models.

Geometry interface:

Like Simulation interface. This is the base class for all geometry models.

Model:

In this UML diagram (Figure 10), we present the implementation of the stacked trapezoid model. Central to this model is the *StackedTrapezoidSimulation* class, which is a composite class integrating *StackedTrapezoidGeometry* and *StackedTrapezoidDiffraction* classes. The *StackedTrapezoidGeometry* class handles all geometrical calculations and stores the relevant information, while the *StackedTrapezoidDiffraction* class is responsible for all diffraction-related calculations.

These classes work together to simulate the physics and generate data that can be compared with experimental results. Throughout the project, additional models are being developed and they will be discussed later. The stacked trapezoid model serves as a prototype, illustrating how other models will be implemented. Each model will adhere to the base interfaces *Simulation* and *Geometry*.

4.2.2 Relationships between components

This system is designed to simulate and analyze diffraction patterns using a modular approach. At its core, the system comprises several key components: the *Fitter*, *Residual*, and interfaces for *Simulation* and *Geometry*, along with specific model implementations like the *StackedTrapezoidSimulation*. The *Fitter* class orchestrates the fitting process by using the *cmaes* function to estimate the best-fit parameters and the *mcmc* function to assess the uncertainty in the fit. It takes in a model and experimental data, utilising the *Residual* class to compute the difference between the experimental data and the model's simulated data. The *Residual* class leverages the model's *simulate_diffraction* function to generate the diffraction pattern, which it then compares with the experimental data.

The model, meaning specific implementations of the interfaces, is designed to operate independently, allowing users to utilize a particular model for simulations without needing to engage with the fitting component. This autonomous functionality ensures that users can easily perform simulations solely with the model of their choice. This design choice enhances flexibility and usability, as it decouples the simulation process from the fitting procedures,

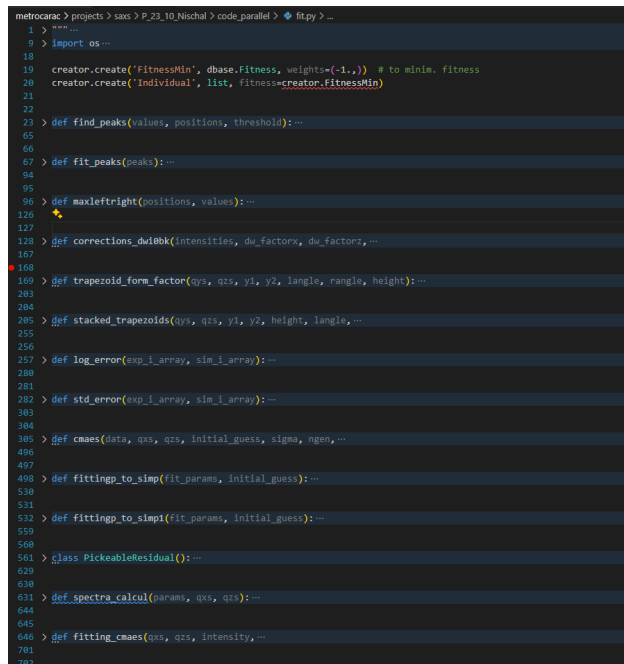
making it more accessible for users who may only need to run simulations. The usefulness and implications of this design choice will be discussed shortly.

4.2.3 Why this design?

Let's first discuss the shortcomings of the previous design before explaining why the current design was chosen.

Previous design:

The previous design adopted a monolithic approach, where the fitting and simulation components were tightly coupled. This meant that any changes to the fitting algorithm (e.g., adding a new optimization technique) would necessitate modifications to the simulation code as well. This interdependence made it difficult to introduce new functionalities or modify existing ones without potentially impacting other parts of the system.



```

metrocarac > projects > sax > P_23_10_Nischal > code_parallel > fitpy > ...
1 > """...
2 >
3 > import os ...
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19 creator.create("FitnessMin", base.Fitness, weights=(-1.0,)) # to min. fitness
20 creator.create("Individual", list, fitness=creator.FitnessMin)
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Figure 11: Monolithic design of the previous version. This script combined all the functions for the stacked trapezoid model in one file.

Furthermore, individual models within the system implemented their own fitting and simulation functionalities. This resulted in a significant amount of code duplication, leading to inconsistencies between models. Moreover, adding new models required rewriting these functionalities from scratch, hindering the system's scalability. This approach proved inefficient for managing a growing number of models. These two limitations combined made it

challenging to optimize or parallelize the code effectively. Since the fitting and simulation components were intertwined, it was difficult to isolate and optimize individual sections for improved performance.

Similarly the codes for the different models each implemented their own fitting and simulation components, leading to code duplication and inconsistencies across models. This design was not scalable, as adding new models required rewriting the fitting and simulation components for each model.

Current design:

The current design improves upon the previous version by decoupling the fitting and simulation components, resulting in a modular structure. The *Simulation* interface serves as the connector between fitting and simulation, standardizing the process of fitting any model.

Also, by enabling simulation models to be developed independently of the fitting component, new models can be introduced without modifying the existing codebase. Developers can focus on developing new models without concerning themselves with the fitting process. This design choice significantly enhances the system’s flexibility and scalability.

The decoupling of components also facilitates optimization and parallelization by allowing developers to focus on specific areas of improvement without cross-component interference. For example, the simulation component can be optimized to run faster and more efficiently through advanced numerical methods and parallel computing techniques. This is particularly beneficial for handling large datasets or complex simulations that demand significant computational resources. By isolating the simulation component, developers can experiment with and implement various optimization strategies, ensuring that the component runs as efficiently as possible.

4.3 Simulation Models

During my project there were three main models that are planned to be integrated with the application. In this section I will discuss the idea behind each model and how they were or are in the process of being implemented in the code. For each model, We will also shortly discuss the lithography technique and/or fault that it is designed to simulate.

4.3.1 Stacked Trapezoid Model

4.4 Implementation of MCMC

4.5 Optimisation and parallelisation

4.6 On the fly uncertainty estimation

5 Future Prospects

References

- [1] Soham Chatterjee. *Beginner's Guide to Moore's Law*. URL: <https://medium.com/@csoham358/beginners-guide-to-moore-s-law-3e00dd8b5057> (visited on 07/01/2021).
- [2] G. F. Lorusso and D. C. Joy. "Experimental resolution measurement in critical dimension scanning electron microscope metrology". In: *Scanning* 25.4 (2003), pp. 175–180. DOI: <https://doi.org/10.1002/sca.4950250403>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sca.4950250403>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sca.4950250403>.
- [3] Weidong Yang et al. "Line-profile and critical-dimension monitoring using a normal incidence optical CD metrology". In: *IEEE Transactions on Semiconductor Manufacturing* 17.4 (2004), pp. 564–572. DOI: 10.1109/TSM.2004.835728.
- [4] T. Baumbach, D. Lübbert, and M. Gailhanou. "Strain and Shape Analysis of Multilayer Surface Gratings by Coplanar and by Grazing-Incidence X-Ray Diffraction". In: (2000). DOI: 10.1063/1.372409. URL: <https://publica.fraunhofer.de/handle/publica/197409>.
- [5] Guillaume Freychet. "Analyses morphologiques et dimensionnelles de nanostructures organisées par diffusion centrale des rayons X". Theses. Université Grenoble Alpes, Oct. 2016. URL: <https://theses.hal.science/tel-01612242>.
- [6] Daniel Sunday et al. "Determining the shape and periodicity of nanostructures using small-angle X-ray scattering". In: *Journal of Applied Crystallography* 48 (Oct. 2015). DOI: 10.1107/S1600576715013369.
- [7] Ronald Jones et al. "Pattern fidelity in nanoimprinted films using critical dimension small angle x-ray scattering". In: *Journal of Microlithography Microfabrication and Microsystems* 5 (Jan. 2006). DOI: 10.1117/1.2170550.
- [8] Jérôme Reche. "Nouvelle méthodologie hybride pour la mesure de rugosités sub-nanométriques". Theses. Université Grenoble Alpes, Oct. 2019. URL: <https://theses.hal.science/tel-02520554>.
- [9] Adam F Hannon et al. "Advancing X-ray scattering metrology using inverse genetic algorithms". In: *Journal of micro/nanolithography, MEMS, and MOEMS : JM3* 15.3 (2016), p. 034001. DOI: 10.1117/1.JMM.15.3.034001.
- [10] Daniel F Sunday et al. "Evaluation of the effect of data quality on the profile uncertainty of critical dimension small angle x-ray scattering". In: *Journal of Micro/Nanolithography, MEMS, and MOEMS* 15.1 (2016), p. 014001. DOI: 10.1117/1.JMM.15.1.014001.