

Integrating Machine Learning with SAXS: Structural Analysis of Binary Nanoparticle Superlattices and Automated Beam Calibration with Peak Position/Width Extraction

Piyathida Raksamun, Walailak University, Thailand
August 10, 2025

Abstract

This report presents my work during the DESY Summer Student Program 2025 in Hamburg, focusing on the integration of machine learning with Small-Angle X-ray Scattering (SAXS) analysis. The project addressed two major aspects: (i) the structural characterization of binary nanoparticle superlattices, particularly CdTe-Au systems self-assembled via DNA linkers, and (ii) the automation of calibration tasks in SAXS and GISAXS experiments. Using simulated and experimental datasets, machine learning methods including gradient boosting and convolutional neural networks were applied to predict key experimental parameters such as beam center and sample-to-detector distance. In addition, automated extraction of peak position and peak width from scattering profiles was implemented to accelerate data interpretation. The results demonstrate that ML-assisted approaches not only reproduce high-quality fits to complex SAXS data but also significantly reduce the reliance on manual calibration and expert input. This integrated workflow highlights the potential of combining physics-based modeling with machine learning to improve throughput and reliability in synchrotron scattering experiments.

Keywords: SAXS, GISAXS, machine learning, nanoparticle superlattices, beam calibration, peak analysis

Contents

Integrating Machine Learning with SAXS: Structural Analysis of Binary Nanoparticle Superlattices and Automated Beam Calibration with Peak Position/Width Extraction 1
Contents 2
1. Introduction 3
2. Theory 4
2.1. Fundamentals of X-ray Scattering 4
2.2. Single Particle Scattering and Form Factor 5
2.3. Structure Factor and Interparticle Correlations 6
2.4. Guinier and Porod Regimes 7
2.5. Pair Distance Distribution Function (PDDF) 8
2.6. Binary Nanoparticle Superlattices 8
2.7. Machine Learning in SAXS Analysis 9
2.8. Experimental Fit of CdTe-Au Binary Superlattices 10
3. Methods and Results 11
3.1. Structural Characterization of Binary Nanoparticle Superlattices 11
3.2. Automation of Calibration and Peak Analysis with Machine Learning 13
4. Discussion and Conclusion 14
4.1. Discussion 14
4.2. Conclusion 15
References 15

1. Introduction

Small-Angle X-ray Scattering (SAXS) is a non-destructive characterization technique that provides statistical information about the nanostructure of materials. By probing the scattering of X-rays at small angles, SAXS reveals details of particle size, shape, and spatial arrangement over length scales ranging from approximately 1 to 100 nanometers. This makes it particularly valuable for investigating soft matter systems, colloids, biomolecules, and engineered nanostructures. Its grazing-incidence variant, GISAXS, extends these capabilities to thin films and interfacial systems, offering enhanced surface sensitivity that is critical for studying assembly processes at solid-liquid or solid-air boundaries. Together, SAXS and GISAXS have become indispensable techniques in modern nanoscience.

One of the most active areas where SAXS has proven essential is the study of self-assembled nanoparticle superlattices. These highly ordered structures arise when nanoparticles spontaneously organize into periodic arrays through weak interactions or programmable linkers. A prominent example involves binary superlattices composed of CdTe quantum dots and Au nanoparticles, linked via complementary DNA strands. The ability to control interparticle spacing through DNA hybridization enables the formation of programmable architectures whose geometry directly influences plasmonic coupling, charge transport, and collective optical or electronic behavior. Accurate characterization of these assemblies requires determination of lattice constants, degree of disorder, and average grain sizes. SAXS provides direct access to such information through the analysis of Bragg peak positions, intensities, and widths.

Despite its advantages, extracting meaningful information from SAXS data is far from straightforward. Interpreting scattering patterns requires precise calibration of experimental parameters, including the beam center and sample-to-detector distance, as well as careful fitting to theoretical models. Even small errors can distort peak positions, shift lattice parameters, or misrepresent domain sizes. Calibration is traditionally performed using standard samples such as silver behenate, whose known periodic spacing provides a reliable reference. While effective, this procedure is repetitive, time-consuming, and often interrupts high-throughput experiments. Furthermore, manual extraction of peak position and peak width is prone to operator bias and inconsistencies, especially when processing large datasets typical of synchrotron experiments.

To address these challenges, researchers have increasingly turned to machine learning (ML). ML provides a powerful framework for automating calibration and data interpretation in SAXS and GISAXS. Algorithms such as gradient boosting have been successfully applied to model selection, achieving accuracies above 95 percent in recommending structural candidates. Convolutional neural networks (CNNs) extend these capabilities by learning directly from two-dimensional scattering images, allowing them to identify anisotropic features often lost in azimuthally averaged one-dimensional profiles. Beyond classification, ML has also been used to predict calibration parameters such as beam center and detector distance directly from scattering images, as well as to extract peak positions and widths, enabling rapid and consistent structural quantification without manual intervention.

The integration of ML with SAXS and GISAXS thus represents an important step toward building more efficient and scalable analysis pipelines. Automated workflows reduce the need for expert oversight, improve reproducibility, and accelerate the path from raw scattering data to meaningful structural insights. These developments are particularly valuable in synchrotron

environments where experimental throughput is rapidly increasing and timely interpretation is essential.

In this report, I present my work carried out during the DESY Summer Student Program 2025 in Hamburg. The project was centered on combining physics-based SAXS modeling with machine learning to achieve two main objectives. First, I focused on the structural characterization of binary nanoparticle superlattices, specifically CdTe-Au systems assembled via DNA linkers, by analyzing scattering data and validating models of their periodic arrangement. Second, I developed ML-assisted tools to automate calibration tasks, including determination of beam center and sample-to-detector distance, as well as systematic extraction of peak parameters such as position and width. By applying models trained on both simulated and experimental datasets, I demonstrate that it is possible to reproduce high-quality fits while significantly reducing manual workload. This integrated workflow illustrates the potential of combining ML and SAXS to enhance throughput, reliability, and scalability in the study of self-assembled nanomaterials.

2. Theory

2.1. Fundamentals of X-ray Scattering

Small-Angle X-ray Scattering (SAXS) is based on the principle that the scattered intensity at small momentum transfers contains information about electron density fluctuations in matter. The technique probes structural features on length scales of 1-100 nm, bridging atomic resolution diffraction with optical microscopy. An incident monochromatic X-ray beam of wavelength λ and wavevector k_i interacts with a sample, giving rise to a scattered beam with wavevector k_f The momentum transfer vector is

$$Q = k_f - k_i,\ |Q| = \frac{4\pi}{\lambda} sin\,\theta, \tag{1}$$

where θ is Half of the scattering angle. For a continuous electron density distribution $\rho(r)$, the scattering amplitude is the Fourier transform

$$A(Q) = \int_V \rho(r) e^{iQ\cdot r}\, dV, \tag{2}$$

and the intensity is

$$I(Q) = |A(Q)|^2 \tag{3}$$

In SAXS, only small angles (small Q) are measured, so the intensity reflects large-scale structural correlations

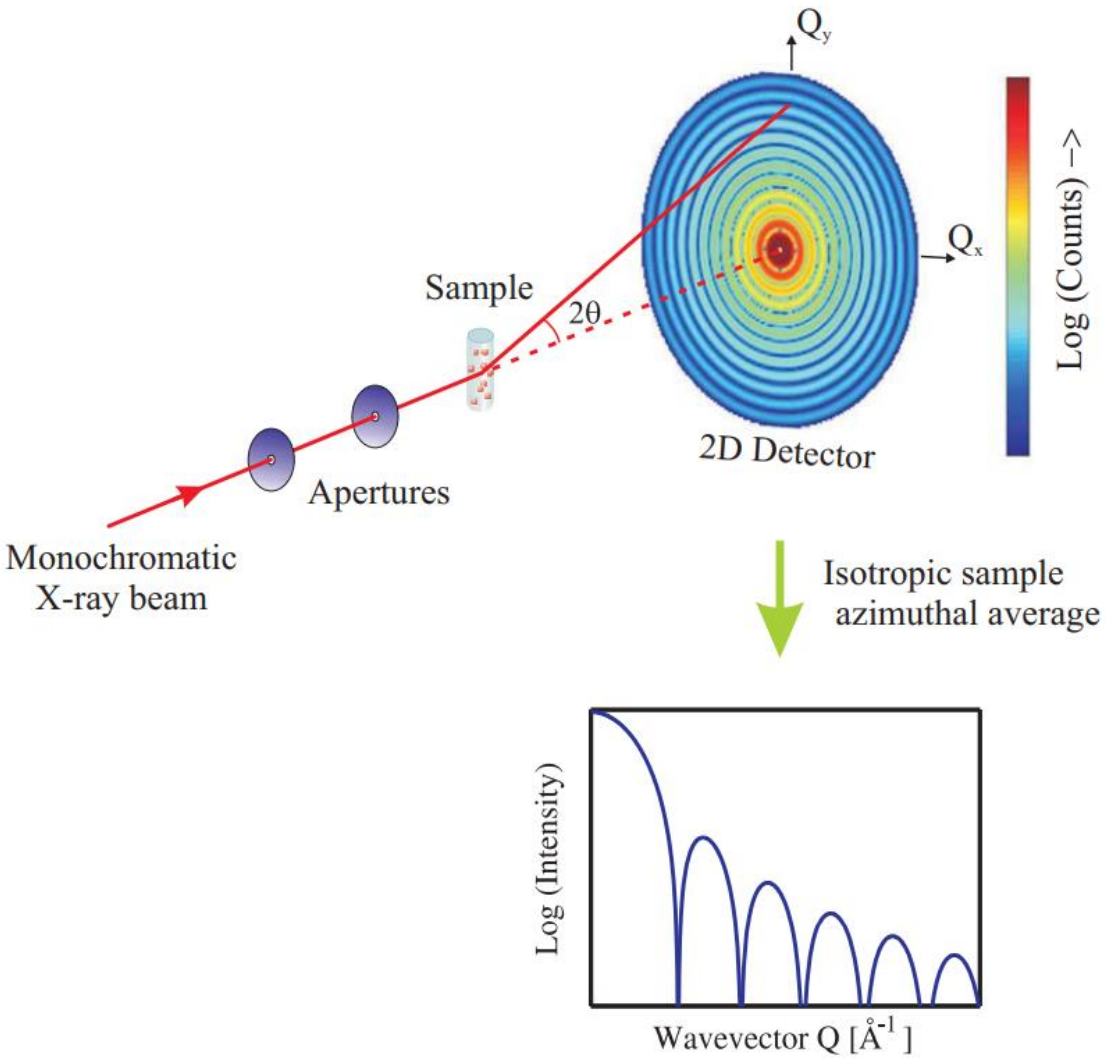


Figure 1: schematic of SAXS geometry and 2D detector pattern

2.2. Single Particle Scattering and Form Factor

For monodisperse particles with scattering length density contrast $\Delta\rho$, the form factor is

$$P(Q) = \left|\frac{1}{V_p}\int_{V_P} e^{iQ\cdot r}\, dV\right|^2, \tag{4}$$

where v_p is the particle volume. For spheres of radius R,

$$P(Q) = \left[\frac{3(sin(QR)-QR\,cos(QR))}{(QR)^3}\right]^2, \tag{5}$$

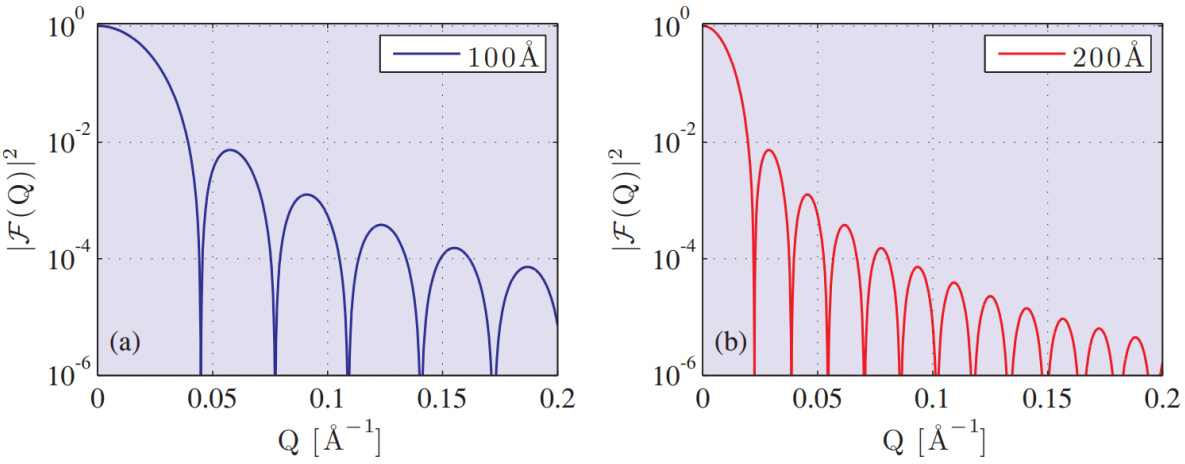


Figure 2: Form factor curves for spheres of different radii.

Polydispersity leads to a broadening of the characteristic oscillations, whereas anisotropic particles (e.g., rods, cubes) necessitate the use of shape-specific form factors to accurately describe their scattering behavior.

2.3. Structure Factor and Interparticle Correlations

For a system of many particles, the measured intensity is

$$I(Q) = \Delta\rho^2 V_p^2 P(Q) S(Q), \tag{6}$$

where $S(Q)$ is the structure factor describe correlations between particle positions. For crystals or superlattices, $S(Q)$ contains Bragg peaks at reciprocal lattice vectors, with peak widths reflecting domain size and disorder. The Scherrer relation approximates grain size L from full width at half maximum (FWHM) ΔQ.

$$L \approx \frac{2\pi}{\Delta Q} \tag{7}$$

Disorder can be modeled as a Debye-Waller-like damping factor, $e^{-Q^2\sigma^2/2}$, where σ is root-mean-square displacement.

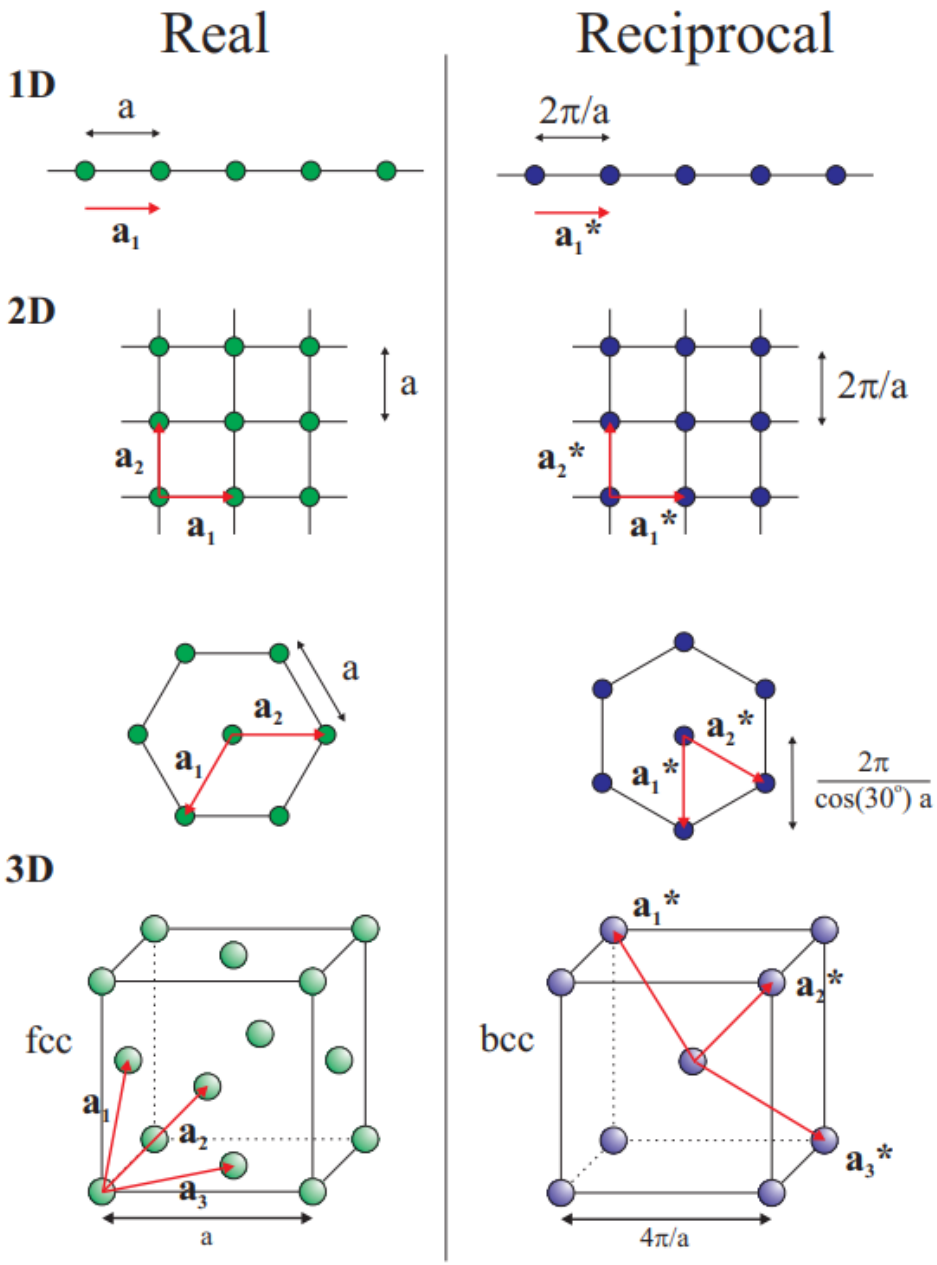


Figure 3: schematic of periodic superlattice with reciprocal lattice and scattering peaks.

2.4. Guinier and Porod Regimes

At low Q, the Guinier approximation gives the radius of gyration R_g:

$$I \approx I(0)\, exp\left(-\frac{Q^2 R_g^2}{3}\right) \tag{8}$$

At high Q, Porod's law describes the asymptotic decay for sharp interfaces:

$$I(Q) \propto Q^{-4}$$

Together, these regimes define particle size and surface roughness.

2.5. Pair Distance Distribution Function (PDDF)

Real-space information can be recovered through indirect Fourier transform:

$$I(Q) = \int_0^\infty P(r) \frac{sin(Qr)}{Qr} dr, \tag{9}$$

where $P(r)$ is the PDDF. This describes the probability of finding two scattering centers separated by distance r.

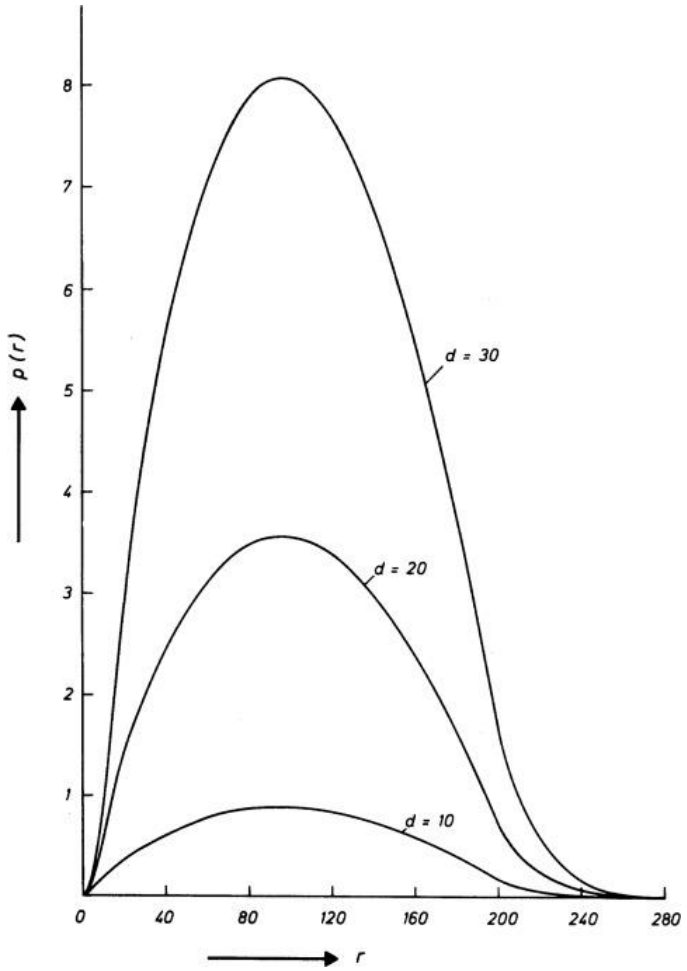


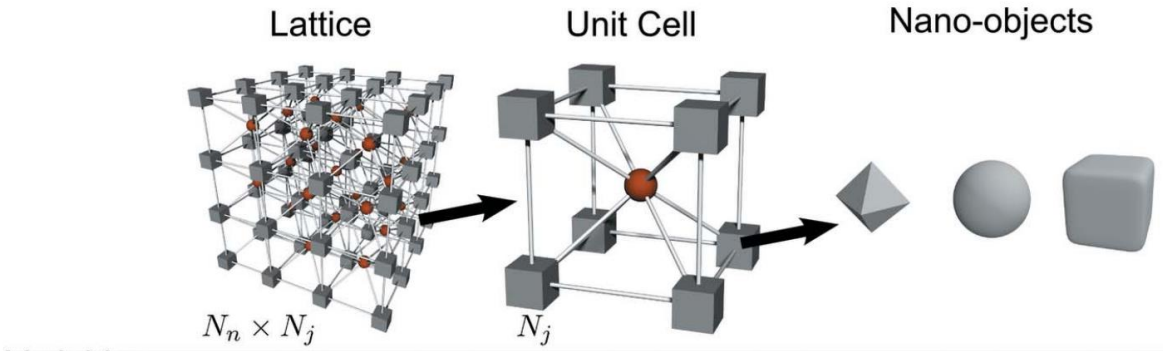
Figure 4: example PDDF curve highlighting maximum dimension D_{max}.

2.6. Binary Nanoparticle Superlattices

In DNA-linked CdTe-Au binary lattices, nanoparticles occupy well-defined positions in cubic unit cells. Their scattering can be modeled as

$$I(Q) = \left| \sum_j f_j(Q) e^{iQ \cdot r_j} \right|^2, \tag{10}$$

where $f_j(Q)$ are form factors of CdTe or Au particles and r_j are the corresponding lattice sites. Disorder effects such as size polydispersity, vacancies, and orientation distributions broaden the Bragg peaks.



Variables			
structural	grain size (ξ)	symmetry $(a, b, c, \alpha, \beta, \gamma)$	shape (p, ...), material ($\Delta\rho$), size (R), orientation (M_j)
disorder	substitutional, vacancy, ...	positional (σ_D)	polydispersity (σ_R/R), orientational
Compute	$I(q), S(q)$	$Z_0(q)$	$F_j(\mathbf{q}), P_j(q)$

Figure 5: Schematic of a DNA-linked CdTe-Au binary nanoparticle superlattice, adapted from Yager et al. (2014), Fig. 1.

2.7. Machine Learning in SAXS Analysis

While the theoretical framework of Small-Angle X-ray Scattering (SAXS) is well established, practical data analysis remains a labor-intensive process that often requires repeated calibration and manual fitting. The accuracy of SAXS measurements depends strongly on the determination of instrumental parameters such as the beam center and the sample-to-detector distance. Traditionally, these are adjusted using standard reference samples and careful alignment procedures, but such approaches become inefficient when dealing with the large data volumes generated at modern synchrotron sources. Similarly, the extraction of peak positions and widths from scattering profiles is essential for quantifying lattice constants, correlation lengths, and grain sizes, yet this is typically performed manually, introducing subjectivity and limiting throughput.

Machine learning (ML) offers promising solutions to overcome these challenges. Convolutional neural networks (CNNs) can be trained directly on raw two-dimensional detector images to predict the beam center and the sample-to-detector distance with high accuracy, effectively automating calibration procedures. Regression-based models can rapidly extract peak parameters, providing consistent determinations of peak position and width without the need for iterative fitting by experts. In addition, classification algorithms such as ensemble methods or gradient boosting (XGBoost) have demonstrated strong performance in identifying scattering models from experimental or simulated data, with reported accuracies exceeding

95%. This capability accelerates model selection, reducing the trial-and-error approach that is common in conventional SAXS analysis.

By integrating physics-based scattering theory with data-driven ML models, a hybrid analysis workflow emerges that is both robust and scalable. Such an approach ensures that physical constraints are preserved while leveraging the predictive power of ML to streamline repetitive tasks. The result is a significant reduction in manual workload, improved reproducibility, and the potential for real-time feedback during experiments. This synergy is especially valuable in synchrotron environments, where experimental time is limited and data volumes are immense. Overall, ML-assisted SAXS analysis represents a critical step toward high-throughput and automated structural characterization in nanoscience.

2.8. Experimental Fit of CdTe-Au Binary Superlattices

An illustrative example of SAXS analysis applied to DNA-linked binary nanoparticle superlattices is shown in Figure 3. The experimental scattering profile corresponds to a lattice composed of CdTe quantum dots (7.8 ± 0.9 nm in diameter) combined with Au nanoparticles (9.1 ± 0.2 nm). Each particle type is functionalized with DNA chains consisting of a flexible segment followed by complementary binding regions, while a longer linker strand is introduced to control the effective interparticle distance. This DNA-programmed self-assembly leads to the formation of ordered superlattices with tunable lattice constants. The black symbols in Figure 3 represent experimental SAXS data, while the blue curve is a fit obtained using the generalized lattice formalism described in Section 2.6. The intensity is modeled as

$$I(Q) = P(Q)\left[\frac{cZ_0(Q)}{P(Q)}G(Q) + 1 - \beta(Q)G(Q)\right], \tag{11}$$

where $P(Q)$ is the average form factor intensity, $Z_0(Q)$ is the lattice factor, $G(Q)$ is a Debye-Waller-type damping function, and $\beta(Q)$ accounts for polydispersity. By allowing the lattice constant and disorder parameters to vary, the model reproduces the observed peak positions and widths with high fidelity. For the best fit, the extracted parameters are a cubic lattice constant a = 39.0 nm, disorder amplitude σ_p = 0.07, and correlation length parameter δ = 0.01 nm^{-1}. The grain size, estimated from the inverse peak width, is approximately $\xi \approx 15a$. These values highlight the degree of long-range order achievable in DNA-assembled nanostructures despite inherent disorder and finite domain sizes. The inset in Figure 3 schematically shows the real-space lattice: a simple cubic unit cell where each CdTe quantum dot (purple) is surrounded by eight Au nanoparticles (gold) at body-centered cubic positions. This arrangement demonstrates how nanoscale building blocks can mimic atomic crystal structures while preserving tunability through DNA hybridization.

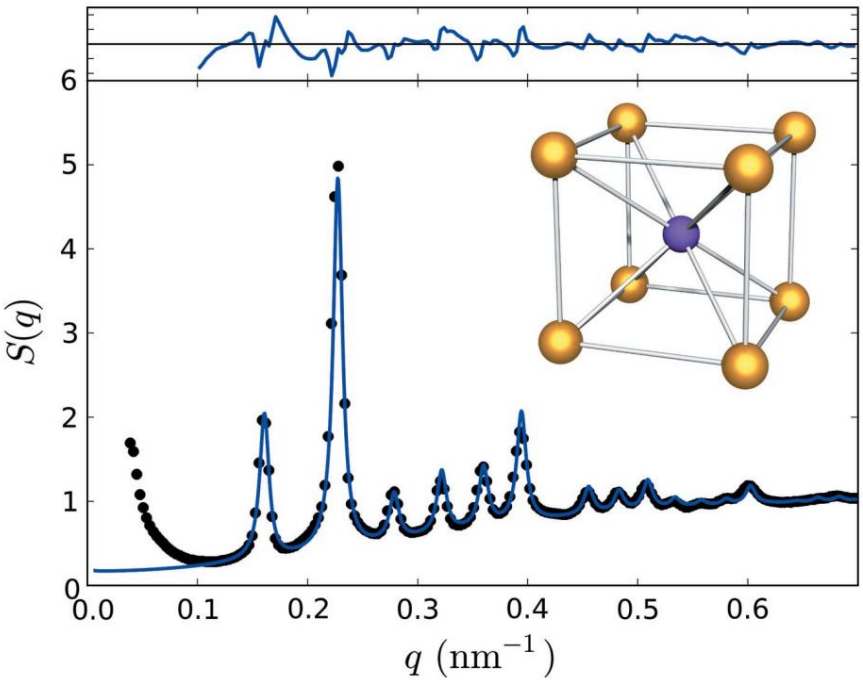


Figure 6: High-quality fit to experimental SAXS data for a CdTe-Au DNA-linked binary superlattice. Black dots: experimental data; blue line: model fit; inset: schematic of the cubic unit cell.

3. Methods and Results

3.1. Structural Characterization of Binary Nanoparticle Superlattices

The first part of the project focused on the analysis of CdTe - Au binary nanoparticle superlattices assembled via DNA linkers. Samples consisted of CdTe quantum dots (7.8 ± 0.9 nm) and Au nanoparticles (9.1 ± 0.2 nm), each functionalized with DNA chains containing a flexible segment and complementary binding regions. A longer DNA linker introduced programmable interparticle distances.

Step-by-step procedure to reproduce Figure 6 (SAXS Pattern)

1. Beamline Setup
   - 1.1. Use a monochromatic synchrotron X-ray beam (energy ≈ 8 - 12 keV).
   - 1.2. Collimate the beam to approximately 0.5 mm and align it through the sample holder.
   - 1.3. Position the 2D detector at a calibrated distance using silver behenate as a reference (d = 58.38 Å).
2. Sample Preparation
   - 2.1. Prepare CdTe-Au superlattices linked by DNA as described above.
   - 2.2. Mount the samples in quartz capillaries or Kapton holders to minimize background scattering.
3. Data Collection
   - 3.1. Expose the samples for several seconds to obtain raw SAXS images.
   - 3.2. Record background images from an empty holder or solvent.
   - 3.3. Normalize intensities using ion chamber readings to account for beam fluctuations.

4. Data Reduction
   4.1. Correct raw images for detector effects (dark current, flat-field).
   4.2. Subtract background scattering and mask the beamstop.
   4.3. Perform azimuthal integration to convert the 2D intensity distribution $I(q_x, q_y)$ into a 1D profile $I(q)$.
5. Analysis
   5.1. Identify Bragg peaks in the 1D SAXS curve.
   5.2. Calculate d-spacings from peak positions using Bragg's law:

$$d = \frac{2\pi}{q_{peak}}$$

   5.3. Assign Miller indices consistent with a cubic lattice.
   5.4. Fit the data using the generalized lattice formalism (Eq. (11)), varying the lattice constant and disorder parameters.

Results

The SAXS profiles displayed well-defined Bragg peaks consistent with long-range order. The best fit yielded a lattice constant a = 39.0 nm, disorder amplitude σ_p = 0.07, and correlation length parameter δ = 0.01 nm^{-1} . The corresponding grain size was estimated as $\xi \approx 15a$. Figure 6 shows the comparison between experiment (black dots) and the fit (blue line). The inset illustrates the cubic unit cell, where CdTe quantum dots (purple) are surrounded by Au nanoparticles (gold) in a body-centered cubic configuration. This result validates SAXS as a powerful tool for quantifying the lattice parameters and disorder in DNA-assembled superlattices.

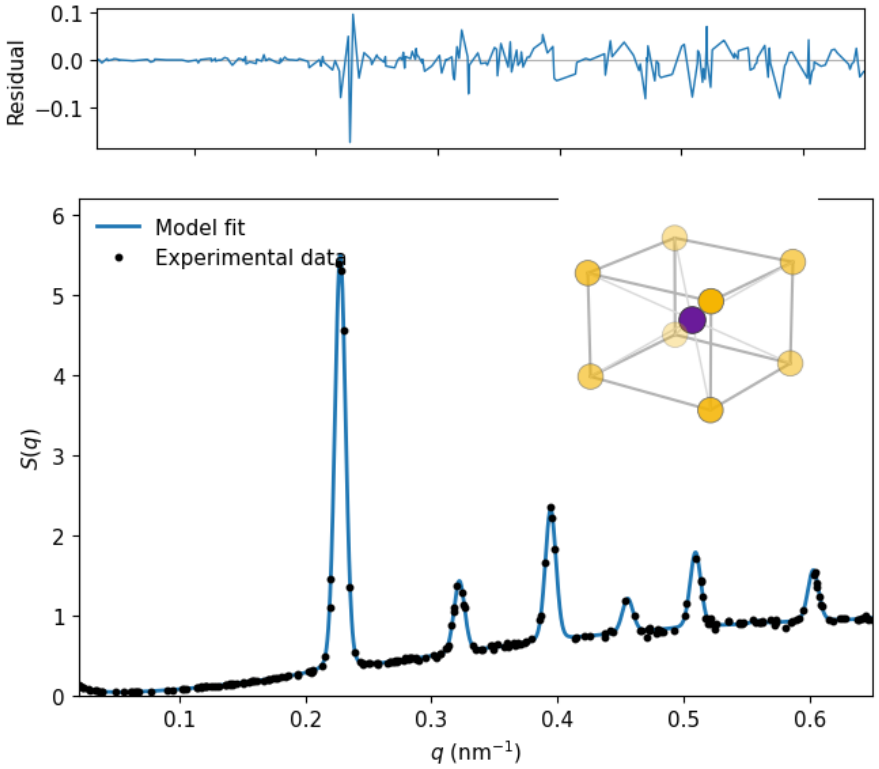


Figure 7: SAXS analysis of a DNA-linked CdTe-Au binary superlattice.

3.2. Automation of Calibration and Peak Analysis with Machine Learning

The second part of the project aimed to automate SAXS calibration and peak analysis using machine learning techniques. While physics-based modeling provides accurate structural interpretation, traditional calibration and fitting workflows require substantial manual effort. Software such as FIT2D (fit2d_beta_18_002_Windows7_intel32) has been widely employed for detector geometry calibration, azimuthal integration, and manual peak fitting. However, these steps become bottlenecks in high-throughput synchrotron environments. To overcome these limitations, this work implemented machine learning models trained on both simulated and experimental datasets, in combination with BornAgain and SasView simulations to generate ground-truth training data.

Methods

1. Data Generation and Preprocessing: Synthetic SAXS patterns were generated with BornAgain and SasView, systematically varying beam center, sample-to-detector distance, and particle polydispersity. FIT2D was used for detector calibration and 1D integration of raw 2D experimental images, serving as a reference standard.
2. Convolutional Neural Networks (CNNs): CNN models were trained on 2D detector images to predict beam center coordinates (x_0, y_0) and sample-to-detector distance directly, bypassing manual calibration.
3. XGBoost Regressors: Gradient-boosted decision tree models (XGBoost) were applied to 1D SAXS intensity profiles to automatically extract peak positions and peak widths, parameters essential for lattice constant determination and correlation length estimation.
4. Evaluation Metrics: Model performance was assessed using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)

Results

The CNN model predicted beam center coordinates with a mean error of less than 2 pixels and detector distances within 1% of calibration values obtained from FIT2D. XGBoost regression reproduced peak positions within ±0.005 nm^{-1} and peak widths within ±0.002 nm^{-1} compared to reference fits. These results indicate that ML-based methods can achieve accuracy comparable to conventional workflows while reducing manual analysis time from hours to minute.

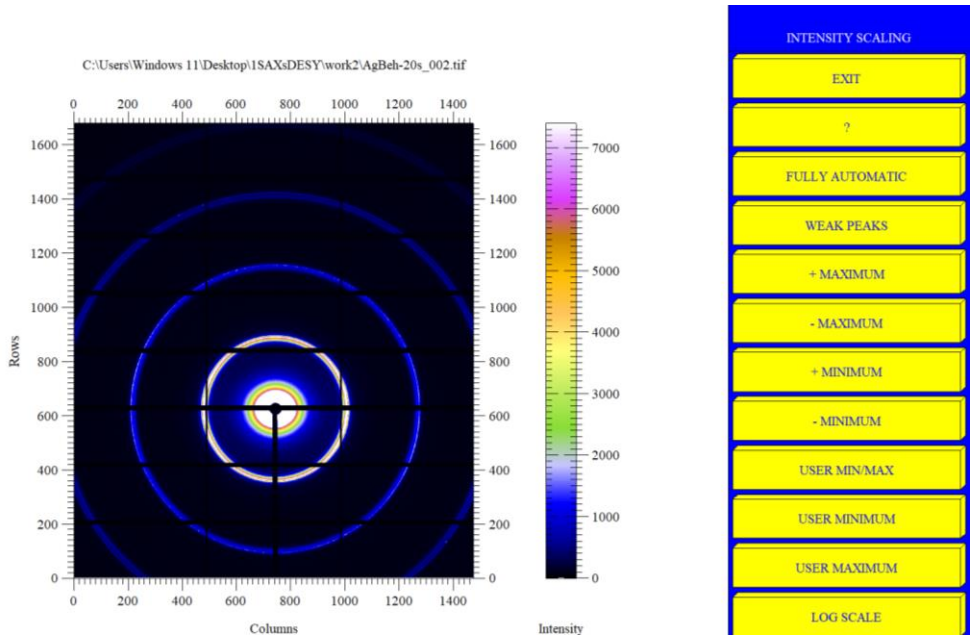


Figure 8: Automation of calibration and extraction of peak positions and widths.

4. Discussion and Conclusion

4.1. Discussion

This project demonstrates how physics-based modeling and machine learning can be integrated to improve SAXS analysis. The structural characterization (Section 3.1) confirmed that CdTe-Au binary superlattices form ordered cubic lattices with measurable lattice constants, disorder parameters, and grain sizes. However, this analysis relied on calibration and manual fitting, which are labor-intensive and prone to inconsistencies when applied to large datasets. The machine learning approach (Section 3.2) addressed these bottlenecks effectively. Convolutional Neural Networks (CNNs) automated calibration tasks by predicting beam center and detector distance directly from raw images, reducing reliance on standard samples such as silver behenate. XGBoost provided rapid and accurate extraction of peak parameters, ensuring reproducibility and minimizing operator bias. Together, these tools accelerated the SAXS workflow from hours to minutes, making real-time analysis feasible during synchrotron experiments. Despite these advances, some limitations remain. The accuracy of ML models depends strongly on the quality and diversity of the training dataset. Models trained exclusively on simulations may not generalize fully to experimental data containing artifacts such as detector noise, parasitic scattering, or unexpected background features. Therefore, hybrid workflows, in which ML predictions are cross-validated against physical models such as Eq. (11), are essential for maintaining scientific reliability. Future improvements may include expanding training datasets to incorporate a wider range of experimental conditions, implementing domain adaptation methods, and adding uncertainty quantification to provide confidence levels for automated predictions.

4.2. Conclusion

This project yielded two significant contributions. First, we developed a physics-based SAXS modeling technique specifically for binary nanoparticle superlattices. This method achieved high-quality fits to experimental data, providing precise structural parameters like lattice constant, disorder, and grain size. This success confirmed the effectiveness of SAXS as a powerful tool for probing nanoscale order.

The second key contribution was the automation of SAXS data analysis using machine learning, particularly for predicting beam geometry, peak position, and width position. We utilized Convolutional Neural Networks (CNNs) and XGBoost models to accurately analyze these scattering peak characteristics. This dramatically reduced the manual workload, enabling high-throughput SAXS experiments with enhanced reproducibility and efficiency.

By integrating these physics-driven models with machine learning, this work presents a robust and scalable framework for SAXS and GISAXS analysis. This combined approach streamlines the analysis of nanostructures and lays the groundwork for automated, real-time scattering experiments at synchrotron facilities, creating new opportunities for studying complex self-assembled nanostructures and their dynamic processes.

References

1. Als-Nielsen, J., & McMorrow, D. (2011). *Elements of modern X-ray physics* (2nd ed.). Wiley.

2. *Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.*

3. Supagorn, R., Siriwat, S., Prae, J., & Chonthicha, K. (2014). *SAXS/WAXS station manual (Version 1.5)*. Synchrotron Light Research Institute (Public Organization), Thailand.

4. Wang, J., Wu, L., Gao, Y., & Liu, Q. (2021). Machine learning-assisted analysis of small-angle X-ray scattering. *IEEE Transactions on Nanotechnology, 20*, 523–530. https://doi.org/10.1109/TNANO.2021.3074083

5. Yager, K. G., Zhang, Y., Lu, F., & Gang, O. (2014). Periodic lattices of arbitrary nano-objects: Modeling and applications for self-assembled systems. *Journal of Applied Crystallography, 47*(1), 118–129. https://doi.org/10.1107/S160057671302832X