

PAPER • OPEN ACCESS

Machine learning for efficient grazing-exit x-ray absorption near edge structure spectroscopy analysis: Bayesian optimization approach

To cite this article: Cafer Tufan Cakir *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 025037

View the [article online](#) for updates and enhancements.

You may also like

- [Unravelling the role of iron and manganese oxides in colouring Late Antique glass by micro-XANES and micro-XRF spectroscopies](#)
Francesca Gherardi, Clément Hole, Ewan Campbell *et al.*
- [Machine-learning strategies for the accurate and efficient analysis of x-ray spectroscopy](#)
Thomas Penfold, Luke Watson, Clelia Middleton *et al.*
- [Low-dimensional systems investigated by x-ray absorption spectroscopy: a selection of 2D, 1D and 0D cases](#)
Lorenzo Mino, Giovanni Agostini, Elisa Borfecchia *et al.*



PAPER

OPEN ACCESS


RECEIVED
13 September 2023REVISED
19 March 2024ACCEPTED FOR PUBLICATION
23 April 2024PUBLISHED
10 May 2024

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Machine learning for efficient grazing-exit x-ray absorption near edge structure spectroscopy analysis: Bayesian optimization approach

Cafer Tufan Cakir^{1,*} , Can Bogoclu², Franziska Emmerling¹, Christina Strel³, Ana Guilherme Buzanich¹  and Martin Radtke^{1,*} 

¹ Federal Institute for Materials Research and Testing (BAM), Berlin 12489, Germany

² Research Group Mechanics of Adaptive Systems Faculty of Mechanical Engineering, Ruhr University Bochum, Bochum 44801, Germany

³ Vienna University of Technology, Atom Institute, Vienna 1020, Austria

* Authors to whom any correspondence should be addressed.

E-mail: cafer-tufan.cakir@bam.de and martin.radtke@bam.de

Keywords: XANES, GE-XANES, Bayesian optimization (BO)

Supplementary material for this article is available [online](#)

Abstract

In materials science, traditional techniques for analyzing layered structures are essential for obtaining information about local structure, electronic properties and chemical states. While valuable, these methods often require high vacuum environments and have limited depth profiling capabilities. The grazing exit x-ray absorption near-edge structure (GE-XANES) technique addresses these limitations by providing depth-resolved insight at ambient conditions, facilitating in situ material analysis without special sample preparation. However, GE-XANES is limited by long data acquisition times, which hinders its practicality for various applications. To overcome this, we have incorporated Bayesian optimization (BO) into the GE-XANES data acquisition process. This innovative approach potentially reduces measurement time by a factor of 50. We have used a standard GE-XANES experiment, which serve as reference, to validate the effectiveness and accuracy of the BO-informed experimental setup. Our results show that this optimized approach maintains data quality while significantly improving efficiency, making GE-XANES more accessible to a wider range of materials science applications.

1. Introduction

Layered material structures play a key role in various technological fields such as electronics, renewable energy and material degradation studies. For instance, thin film transistors (TFTs) in electronics utilize layered structures for enhanced charge transportation and flexibility [1], while thin-film photovoltaic devices in renewable energy employ strategic layering for optimized light absorption and electron-hole separation [2]. Similarly, the study of layered corrosion in compositionally complex alloys provides crucial insights into material degradation processes [3, 4]. Numerous examples, including those mentioned above and others, highlight the crucial role of layered structures in a wide range of research and application areas [5–11].

The conventional study of these complex layered structures employs advanced surface analysis techniques such as x-ray photoelectron spectroscopy (XPS), secondary ion mass spectrometry (SIMS), and Auger electron spectroscopy (AES) [12–18]. While these methods provide detailed insights, they have limitations, such as the need for high vacuum conditions and limited depth analysis.

To address these challenges, x-ray absorption near-edge structure (XANES) spectroscopy has emerged as a promising alternative. XANES operates under ambient conditions and presents a broader scope of applicability across various research contexts. This extends the possibilities for material analysis, particularly for studies requiring more flexible experimental conditions. Additionally, XANES enables time-resolved

measurements, providing dynamic insights into the evolving material structure and related properties under study [19, 20].

Grazing exit x-ray fluorescence spectroscopy (GEXRF) emerges as a non-destructive, depth-resolved, element-specific characterization technique, enabling the study of materials on the nanometer scale. Depending on the grazing exit angle of the fluorescence radiation, the information depth of the experiment can vary from tens to hundreds of nanometers. Such adaptability is important for the analysis of thin films, corrosion layers, and interfaces in layered materials, where the complex relationship between structure and properties at the nanoscale is critical [4].

For the collection of emitted XRF intensity as a function of grazing exit angle, two methodologies are presented: one involves using an energy-sensitive 1D detector, where the range of grazing emission angles is captured by modifying the detector's position; the other employs a position and energy sensitive area detector, enabling the capture of XRF intensities without necessitating detector scanning, as demonstrated in this study. Within the scanning-free approach, a trade-off is encountered between angular resolution and the intensity of emitted fluorescence radiation. The angular resolution of fixed area detectors is inherently dependent on the distance between the sample and the detector. Increasing this distance, increases the angular resolution but decreases the detected intensity by a factor proportional to the square of the distance (D^2) and as a result, experiment times are extended. For greater XRF intensity, the sample-detector distance can be reduced, but this reduces angular resolution and can lead to overlapping fluorescence signals from layered materials due to variance in emission angles from different depths [4].

It is well known in grazing exit setups that shallow emission angles carry information from the surface of the sample being analyzed, whereas as the emission angle increases, information from deeper levels of the material is obtained. The emission angle of the XRF mainly depends on the complex refractive index, and when different oxidation states are analyzed, the angular intensity profiles of layers with different states are close to each other. To resolve the angular intensity profiles of the different layers in such a scenario, good angular resolution is required [4].

In our study, we attempt to achieve high angular resolution and overcome long experimental times. Our study integrates machine learning with Grazing Exit XANES (GE-XANES) spectroscopy, exploiting the inherent strengths of XANES for an improved analytical approach [4]. This innovative strategy aims to overcome the limitations of traditional methods, thereby increasing the speed and efficiency of data acquisition, particularly in depth-resolved analyses of layered materials. By incorporating active learning, a subset of machine learning, we seek to refine the data acquisition process in GE-XANES. This integration allows for a more efficient and streamlined methodology, enabling the acquisition of comprehensive insights into layered structures relevant to multiple scientific disciplines.

Building on its proven track record in data analysis, image processing, and materials synthesis [21–32], machine learning offers promising ways to overcome obstacles in the GE-XANES data collection process. This integration of machine learning represents a new approach, as its potential benefits in facilitating the data collection process are yet to be fully explored.

Active learning has the potential to revolutionize traditional static data collection methods used in x-ray absorption spectroscopy analysis and other research areas [33, 34]. Active learning employs mathematical models to guide optimal data collection decisions. Bayesian optimization (BO), a particularly valuable technique when function evaluation is costly or time-consuming, can be embedded within an active learning framework [35–38]. BO incorporates a prior probability distribution over the objective function that can be updated as new data are collected [39]. This unique ability to quantify uncertainty assists in model interpretation and decision-making.

Advances in active learning are a good example of the broader trend of machine learning's transformative impact on experimental methodology. This shift towards more efficient, data-driven approaches is gaining considerable attention as it not only streamlines data collection processes but also increases the speed and effectiveness of research. Such advances are extremely useful in resource-constrained environments such as synchrotron facilities. The application of machine learning, and in particular its active learning subset, in the field of synchrotron radiation research has already yielded promising results, as evidenced by numerous studies [40–46].

Here, we present a novel integration of BO into GE-XANES analysis. This research aims to accelerate data collection, a much-needed development in the field. A layered reference sample consisting of a silicon wafer with a 500 nm layer of metallic chromium followed by a 300 nm layer of chromium oxide was used to test our approach. Through this innovative approach, it is potentially possible to reduce measurement time by a factor of 50 compared to conventional methodologies.

Our results suggest that this BO-based approach can be adapted to any sequential technique with prolonged measurement times, such as laboratory-based XANES, thereby potentially revolutionizing the field of experimental design. It also highlights the growing need for a faster, more efficient data acquisition

method across different technological domains, which has important implications for future scientific discovery and research.

2. Methods

The present study proposes an optimized data collection strategy for a novel, scanning free GE-XANES measurements in the medium/hard x-ray range. We used a monochromatic beam and an energy and position-sensitive pnCCD detector. By leveraging the BO process, we aim to reduce measurement times. Measurements were conducted at the μ Spot beamline at the BESSY II electron storage ring, operated by the Helmholtz-Zentrum Berlin für Materialien und Energie. Further details about the experimental setup can be found in Cakir *et al* [4].

2.1. Beamline

The mySpot beamline, mounted on the 7 T wavelength shifter at the BESSY II synchrotron radiation ring, is our main instrument of use. The synchrotron beam is first focused using a toroidal total external reflection mirror placed strategically in front of the main beam shutter. This mirror enables a high photon flux of 1010–1011 ph s⁻¹ at 6 keV, with a functional energy range of 4–30 keV. The concentrated beam then undergoes monochromatization via a double crystal monochromator, configured in Si (111), offering an energy resolution of 2×10^{-4} ($\Delta E/E$) [47, 48].

2.2. Detector

For the GE-XANES setup, we employed an energy-dispersive pnCCD detector for selective analysis of specific emission lines, ensuring improved data interpretation accuracy and precision. The detector boasts a sensitive area of 12.7×12.7 mm², comprising 264×264 pixels with an individual area of 48×48 μ m². The detector's quantum efficiency exceeds 95% between 3 keV and 10 keV and presents an energy resolution of 152 eV for Mn-K α (5.9 keV) at 450 kcps [49].

2.3. Sample description

In this study, a reference sample possessing a predetermined composition and layer thickness was employed to validate our results. The sample was prepared through a sequential magnetron sputtering technique, using high-purity chromium from an elemental target. The deposition of the metallic chromium and chromium-oxide layers occurred at room temperature, with the thickness controlled by the deposition rates. The metallic chromium layer was sputtered onto a Si-Wafer using a constant flow of Ar. In contrast, the chromium-oxide layer was deposited atop the chromium layer, using an Ar and O mixture. For a detailed overview of the sample preparation process, see Cakir *et al* [4].

2.4. BO for active learning

BO is a probabilistic framework for optimizing functions that are considered 'black-box' and may be linked to expensive and possibly noisy objectives. These objectives typically carry high evaluation costs, whether it is time, computational resources, or other expense types, such as long synchrotron radiation experiments [50]. BO requires a posterior distribution prediction, and Gaussian process (GP) is a popular choice for this goal.

2.4.1. GP

GP models enable us to understand the relationship between system inputs and outputs. They assume that objective function f is a realization from a distribution of functions,

$$p(f) = GP(f; \mu, \mathbf{K}). \quad (1)$$

Therefore, a GP represents an infinite collection of random variables and any finite subset of these also follows Gaussian distribution [51]. It follows that the marginal distribution of each function value is a univariate Gaussian distribution $f(x) \sim N(\mu(x), \sigma^2(x))$. The mean $\mu(x)$ and the variance $\sigma^2(x)$ functions can be derived by updating our prior with the observed data $D = X; y$ and marginalizing it out [52]:

$$\mu(x) = m(x) + \mathbf{K}(x, X) \mathbf{K}^{-1}(X, X) y \quad (2)$$

$$\sigma^2(x) = \mathbf{K}(x, x) - \mathbf{K}(x, X) \mathbf{K}^{-1}(X, X) \mathbf{K}(x, X)^T. \quad (3)$$

Here, $y = f(X)$ are the observed data points, m is a possibly parameterized mean function, which is set to zero through-out this paper, and \mathbf{K} is the covariance function often chosen from the list of parameterized

stationary kernel functions such as the Matérn kernel. The Matérn kernel is commonly employed by GPs to model a wide range of data relationships. It exhibits flexibility, capable of tracking both stable, smooth functions as well as abrupt changes and discontinuities. However, depending on the problem and data at hand, other kernel functions may prove more effective in different scenarios. The Matérn kernel, an extension of the radial basis function (RBF) kernel, introduces an additional hyperparameter, ν , which governs the smoothness of the resulting function. A smaller value of ν yields a less smooth approximation, while a larger ν causes the kernel to behave similarly to the RBF kernel. In this work, we use Matérn kernel $\mathbf{K}_{\text{Matérn}}$ with $\nu = 1.5$.

Additionally, we define as scaling factor s and homoscedastic noise σ_n resulting in the following kernel function:

$$\mathbf{K}(X, X) = s\mathbf{K}_{\text{Matérn}}(X, X) + \sigma_n^2\mathbf{I}. \quad (4)$$

Type-II maximum likelihood estimation was used to determine the hyperparameters l, s, σ_n where l is the length scale of the function, and $\sigma_n^2\mathbf{I}$ represents the White kernel. In this research, we estimated statistical errors by the square root of the total count, integrating this estimate into our model with a White kernel. We employed a kernel function that combines the Matérn and White kernels, enhancing our model's ability to accurately account for statistical errors.

2.4.2. Acquisition function

The foundational strategy in the BO process involves deriving a new measurement point using the posterior distribution, which is used to construct the acquisition function. We employed the upper confidence bound (UCB) [53] sampling function to determine the subsequent sample points iteratively and to find the peaks of the intensity function. The UCB balances exploration and exploitation to probe unexplored regions within the input space efficiently. The forthcoming sample point is selected by maximizing the UCB function. The equation for the UCB acquisition function is as follows:

$$UCB(x) = \mu(x) + \kappa\sigma(x). \quad (5)$$

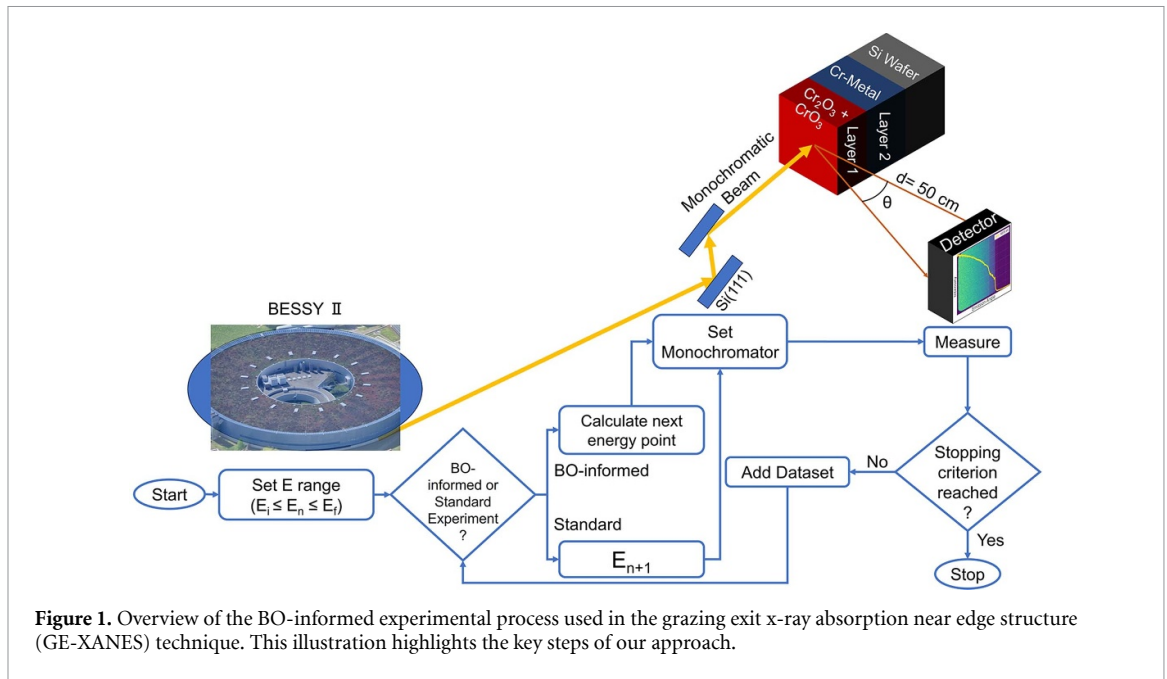
The equilibrium between exploitation ($\mu(x)$) and exploration ($\sigma(x)$) is maintained by weighing the variance function with a trade-off parameter (κ). The impact of the variance function on UCB is directly proportional to the trade-off factor, which remains constant throughout the optimization process. Nevertheless, as the sample size increases, the values derived from the variance function decrease due to the reduction in model uncertainty. This leads to an increased emphasis on the exploitation of the learned model, while the focus on exploration diminishes. Although the variance function could be directly employed as a utility function to predict the subsequent calculation points in maximum exploration mode [54], this might result in the omission of pre-peaks, a common phenomenon in XANES analyses. To circumvent this issue, we use the UCB function to calculate the next point instead of solely relying on the maximum of the variance function. The effect of the κ on UCB function is visualized in supplementary information.

3. Results & discussion

First, it must be emphasized that the standard experiment serves as the reference or ground truth in this analysis, a status supported by prior research and crucial for the comparative approach employed. This standard measurement is the best result experimentally achieved, as it was previously compared to know XANES spectra for the materials at stake [4]. This standard is vital for evaluating the effectiveness and accuracy of the BO-informed experimental method. The experimental setup and data acquisition processes for GE-XANES are illustrated in figure 1. The primary objective is to obtain acceptable data within the shortest possible time frame, which closely resembles the data derived from a 20 min per point standard experiment.

The reference sample used in this study is a silicon wafer with a 500 nm layer of metallic chromium, followed by a 300 nm layer of chromium oxide. To begin the scanning process, we define an energy range around the absorption edge of the chromium atom. This range spans 60 eV, specifically from 5975 eV to 6035 eV, covering the Cr $K\alpha$ absorption edge at 5989 eV.

To analyze the two distinct states of chromium (Cr), the sample-detector distance was set at 50 cm. This configuration offers a total angular range of 1.455° and a horizontal solid angle of 0.0055° per pixel. The angular intensity profile of the chromium oxide layer is compressed into approximately 40 of the 264 pixels, resulting in low count statistics in the XANES spectrum obtained from the surface layer. Consequently, longer measurement time is necessary to thoroughly analyze the surface layer.



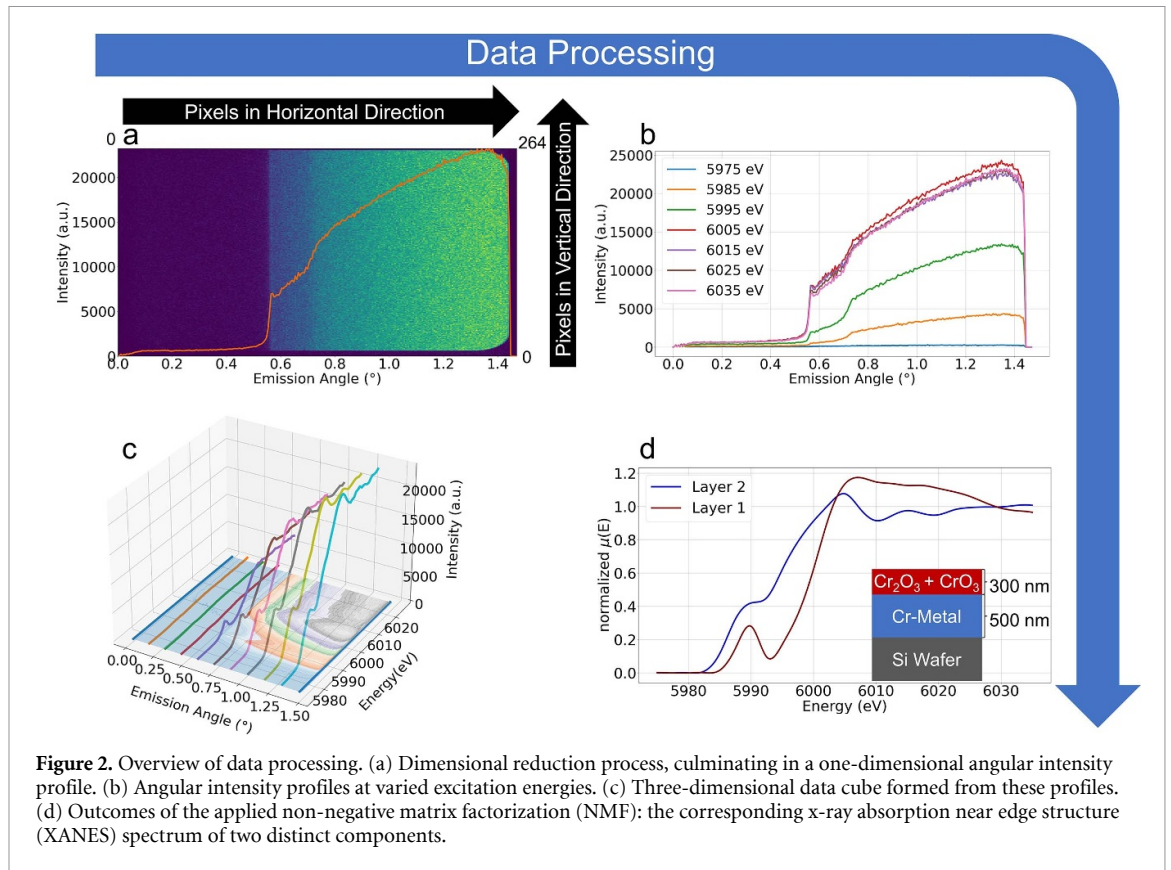
Two distinct data acquisition procedures are used to scan the defined energy range. In the standard procedure, a total of 60 points with 1 eV steps were scanned. The BO-informed method applies an intelligent approach to select energy points for measurement based on an algorithm, optimizing the energy scanning process down to 25 points. To obtain the desired energy values, the angle of the monochromator is aligned to the angle corresponding to the energy value sent from the code. The measurement process is triggered when the monochromator is aligned. The obtained data from both methods are compiled into a dataset. This process continues until a specific stopping criterion is reached, which is 60 measurement points in the case of the standard experiments and 25 measurement points in the experiments informed by BO. The decision to set the stopping criterion at 25 points in the BO-informed experiments will be discussed in the following paragraphs.

As shown in figure 1, with its unique two-dimensional structure, our detector records signal intensities over a specific range of angles corresponding to the emission angle. This two-dimensional design advantageously eliminates the need to physically scan different emission angles by moving the detector. It provides a scanning-free approach, saving considerable measurement time. Furthermore, the detector's sensitivity to the energy of incoming photons allows it to distribute them into 1024 channels according to their energy. This results in a detailed energy spectrum, providing a comprehensive analysis.

Transforming the three-dimensional data (pixel \times pixel \times channels) obtained from the detector into an XANES spectrum requires a series of processing steps. The initial goal is to condense the three-dimensional data collected from the analyzed energy range into a one-dimensional angular intensity profile.

In this study, we have a Cr oxide layer on top of Cr metal, so we can measure the most intense x-ray fluorescence line (Cr- K_{α}) free from other undesired contributions, such as scattering. As a result, we achieve higher accuracy by selectively capturing chromium's unique energy signature. We do this by initially defining a range of ± 10 channels surrounding the Gaussian peak corresponding to the element under study, thus reducing the data from three to two dimensions. Given our setup, the data of interest lies in the horizontal direction of this two-dimensional data set. Consequently, we sum the data in the vertical direction. This process not only amplifies our statistics but also further refines our data from two dimensions down to a one-dimensional angular intensity profile (see figure 2(a)).

The next step is to collect angular intensity profiles at various excitation energies, which are represented in figure 2(b). We can now combine all of this information as a 3D dataset, featuring emission angle, intensity and excitation energy (figure 2(c)). Following this transformation, we apply non-negative matrix factorization (NMF) to the three-dimensional dataset. The NMF process yields two matrices as outcomes. The first matrix elucidates the angular intensity profiles of the two distinct layers within our sample. Meanwhile, the second matrix presents the XANES spectrum corresponding to the angular intensity distribution derived from the first matrix (see figure 2(d)).



As mentioned above, the main advantage of BO-informed experiments lies in the intelligent selection of points in the energy range for analysis rather than statically scanning this range. We describe this process in detail in section 2.4. A flowchart of this process is shown in figure 3 (left panel).

To further illustrate this process, we chose to highlight the 1st, 3rd, 15th, and 17th iterations in figure 3 (right panel). These iterations have been chosen to illustrate the general improvement in the overall scanning process. The progression of the GP model's mean function over these iterations is shown in the right panel of figure 3. The lower graph shows the evolution of the UCB function derived from the selected iterations, highlighting the intelligent selection process inherent in the BO-informed experiment. The gold star on the lower graphs represents the next sampling point.

In BO, a precondition is the establishment of a 'prior function', which forms the basis for the iterative optimization process. This function is essentially a statistical representation of our beliefs about the function before viewing any data. To establish this prior function, we initially measured seven energy points evenly distributed within our energy range (see figure 3(a)). The measurements are denoted by red dots in figure 3 (right panel-top).

The mean function derived from the GP model is represented by the blue curve in figure 3(a). This curve is shaped based on the measurements from the initial seven energy points and serves as a probabilistic model for our function. The light green area surrounding the curve represents the $\pm 95\%$ confidence interval.

The BO-informed experimental process proceeds in iterations following the initial seven-point measurements. It uses the UCB, a function calculated using the mean and variance functions from the GP model, to guide the selection of energy points for measurement.

By the third iteration (figure 3(b)), the UCB function values at previously measured points have decreased. This reduction in UCB values and the associated confidence interval indicates an improvement in the GP model's prediction accuracy due to the additional data. The narrowing of the confidence interval suggests a reduction in uncertainty of predicted values, ultimately leading to a more precise estimation of the target variable values. This iterative improvement continues, with data points and model parameters updating each cycle.

By reaching the 17th iteration (figure 3(d)), the confidence interval has considerably narrowed for all but one point. Once this last point is measured, the maximum standard deviation between the points reduces to 0.005, which is a 20 fold reduction from the first iteration. Our empirical observations indicate that such a low standard deviation between points signifies that the model has already explored the given space and attained a sufficient level of precision. In our setup, we set the stopping criterion for the BO-informed

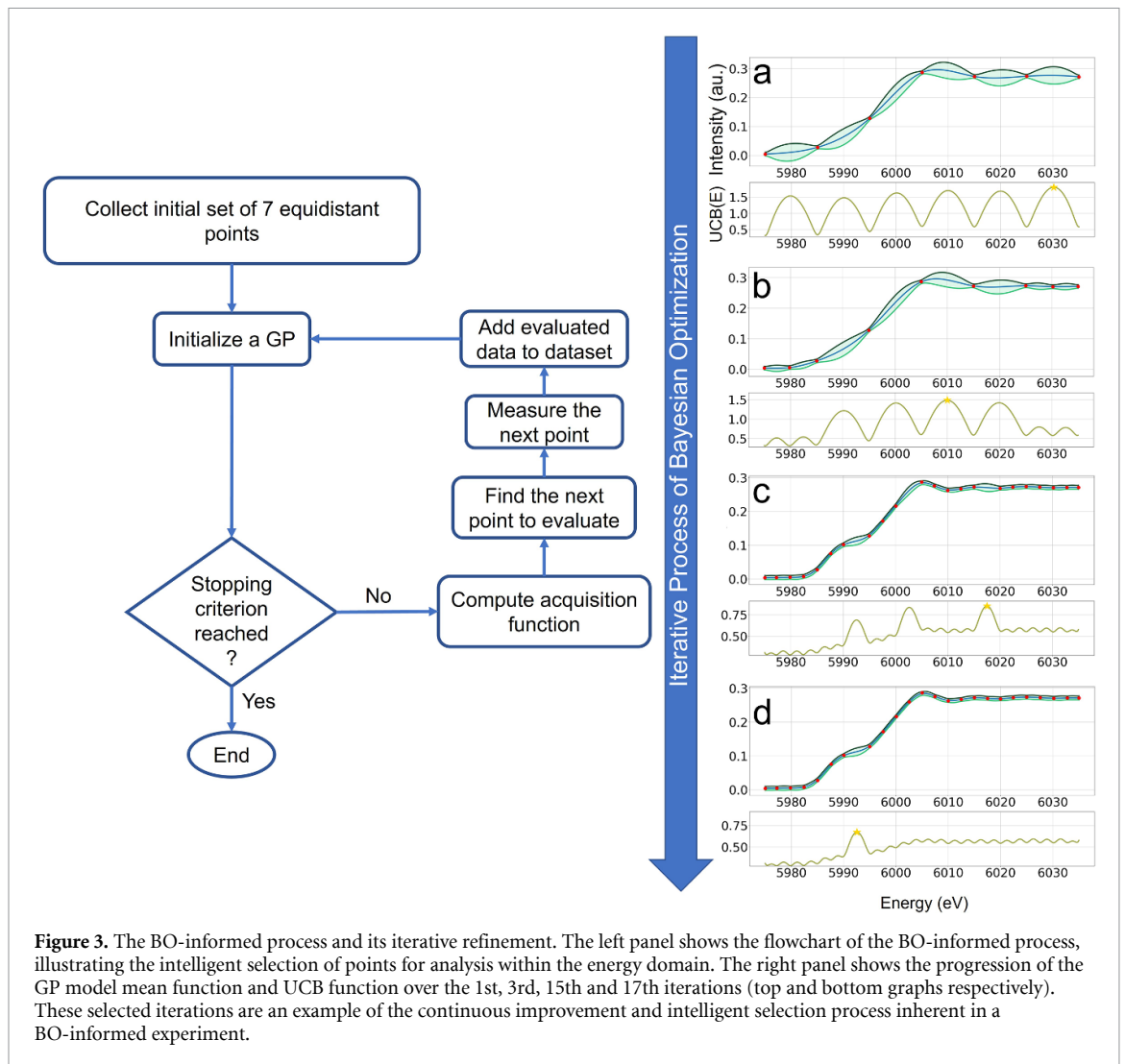


Figure 3. The BO-informed process and its iterative refinement. The left panel shows the flowchart of the BO-informed process, illustrating the intelligent selection of points for analysis within the energy domain. The right panel shows the progression of the GP model mean function and UCB function over the 1st, 3rd, 15th and 17th iterations (top and bottom graphs respectively). These selected iterations are an example of the continuous improvement and intelligent selection process inherent in a BO-informed experiment.

experiments at 18 iterations after the first seven measurements, resulting in a total of 25 points measured. However, the standard deviation between points could also serve as an alternative stopping criterion.

In conventional XANES analyses, a total count of at least 10^6 in the analyzed range is a common requirement. Given the setups inherently includes a low detection angle, meeting this condition for the first layer required extended experiment times. In the standard experimental approach, we more than fulfill this requirement ($\approx 7.5 \times 10^6$) with a total scan time of 20 h for the first layer. This was achieved by measuring 60 points, each lasting 20 min.

In particular, we ran two different BO-informed experiments with different point measurement durations of 5 min and 1 min. The results of the standard experiment and the two BO-informed experiments are shown in figure 4.

The left side in figure 4 shows the results of the standard experiment, while the right side shows the results of the BO-informed experiments, each performed with 1 min and 5 min per point measurements. The top graph shows the combined spectra, summed over all emission angles, before applying NMF analysis.

NMF analysis, which we previously detailed, was then applied to the data obtained from the standard and BO-informed experiments. The resulting first matrix, representing the angular intensity profiles of the two distinct layers, is presented in figure 4(b). Notably, even though the noise ratio increases as the measurement time decreases, the separation process of the two distinct components is carried out efficiently. Figure 4(c) displays the XANES spectra of the two layers derived from these angular intensity profiles. To facilitate a comprehensive comparison, we normalized the XANES spectra obtained for the different layers, and these results are presented in figure 5.

Figure 5, left panel, presents the results obtained from the first layer for the standard, BO-informed 5 min per point and 1 min per point experiments in the top graph. To assess the quality of the optimization process, the root mean square error (RMSE) was calculated between the outcomes of the standard

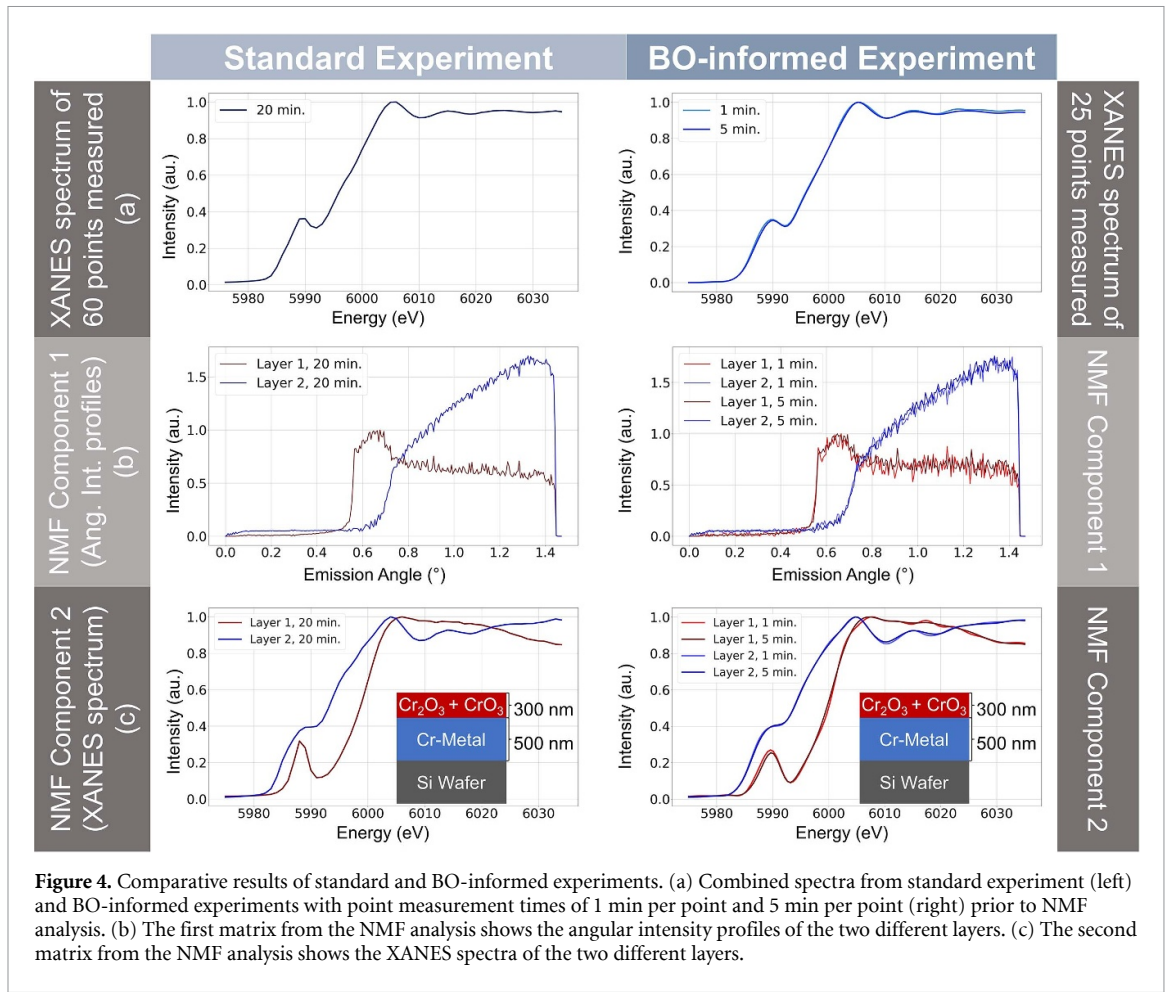


Figure 4. Comparative results of standard and BO-informed experiments. (a) Combined spectra from standard experiment (left) and BO-informed experiments with point measurement times of 1 min per point and 5 min per point (right) prior to NMF analysis. (b) The first matrix from the NMF analysis shows the angular intensity profiles of the two different layers. (c) The second matrix from the NMF analysis shows the XANES spectra of the two different layers.

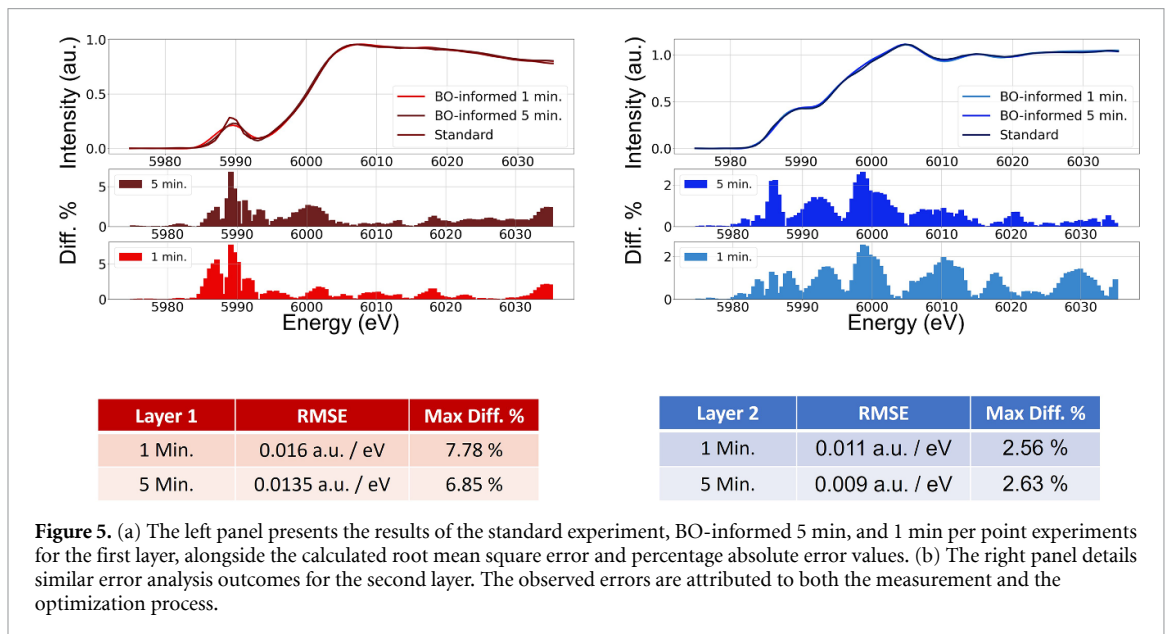


Figure 5. (a) The left panel presents the results of the standard experiment, BO-informed 5 min, and 1 min per point experiments for the first layer, alongside the calculated root mean square error and percentage absolute error values. (b) The right panel details similar error analysis outcomes for the second layer. The observed errors are attributed to both the measurement and the optimization process.

experiments, which involved acquiring 60 equidistant points, and the predictions made by the GP model trained with data from 25 points collected during BO-informed experiments lasting 5 min per point and 1 min per point, respectively. This resulted in a value of 0.016 a.u. eV⁻¹ for the 1 min per point BO-informed experiment. This value was reduced to 0.0135 a.u. eV⁻¹ for the 5 min per point BO-informed experiment, which can be attributed to the beneficial effect of a longer exposure time on the signal-to-noise ratio. When analyzing the percentage absolute error values between spectrums from standard experiment and BO-informed experiments, the maximum deviation from the standard experiment for the pre-peak region

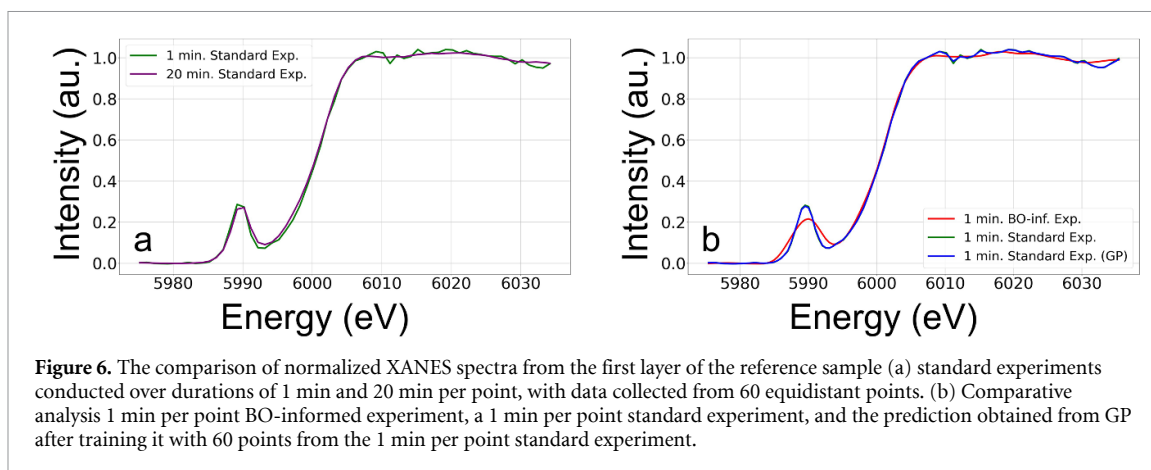


Figure 6. The comparison of normalized XANES spectra from the first layer of the reference sample (a) standard experiments conducted over durations of 1 min and 20 min per point, with data collected from 60 equidistant points. (b) Comparative analysis 1 min per point BO-informed experiment, a 1 min per point standard experiment, and the prediction obtained from GP after training it with 60 points from the 1 min per point standard experiment.

reached 7.78% and 6.85% in the 1 min and 5 min per point BO-informed experiments, respectively. However, for regions outside the pre-peak area, the 1 min and 5 min per point measurements showed an average maximum error of 1%.

The outcomes observed when a similar error analysis was applied to the second layer are genuinely encouraging (see figure 5, right panel). The RMSE between the standard experiment and the BO-informed 1 min per point experiment of the second layer has reduced to 0.011 a.u. eV^{-1} , with the maximum percentage difference also diminishing to 2.56%. When comparing the standard experiment with the BO-informed 5 min per point experiment, values approximating the BO-informed 1 min per point experiment are attained. Specifically, the RMSE for the BO-informed 5 min per point experiment in the second layer is 0.009 a.u. eV^{-1} , while the maximum difference is computed at 2.63%. Although the maximum difference is slightly higher in the BO-informed 5 min per point experiment, the RMSE suggests that the BO-informed 5 min per point experiment aligns more closely with the standard experiment than the BO-informed 1 min per point experiment. This close alignment as a result of the high-count statistics of the deeper levels. It is important to recognize that these deviations are due to the measurement and the optimization process.

An experiment was conducted to determine if the observed close alignment was due to BO or a GP. Figure 6(a) shows the comparison of normalized XANES spectra from the first layer of the reference sample, based on the exposure duration per point of the standard experiment, with data collected from 60 equidistant points as previously mentioned. The 1 min per point standard experiment captured the initial part of the spectrum (pre-peak region at 5975–6005 eV) but showed significant differences from the 20 min per point standard experiment in the region following the absorption edge (post-edge at 6005–6035). To quantify these differences, the RMSE was used, resulting in a value of 0.0157 a.u. eV^{-1} between the 1 min and 20 min per point standard experiments, and the RMSE between the prediction obtained from GP after trained with 25 points from 1 min BO-informed experiment and the 20 min per point standard experiment in the post-edge region is 0.005 a.u. eV^{-1} .

Figure 6(b) represents a comparison between two experimental procedures: a 1 min per point standard experiment and a 1 min per point BO-informed experiment. This comparison aimed to discern whether the close alignment of the BO-informed experiment's results with those of the established 20 min per point standard experiment was solely due to the application of a GP or an intrinsic aspect of the BO-informed methodology. In the 1 min per point standard experiment, a GP model was trained using 60 points collected from the 1 min per point standard experiment to assess its impact on handling noise and characterizing the data. However, even after applying the GP to the 1 min per point standard experiment data, the features of the curve do not match the ones from the 20 min per point standard experiment, while the 1 min per point BO-informed experiment do. The RMSE calculated between the 20 min per point standard experiment and the prediction obtained from GP after training it with 60 points from the 1 min per point standard experiment was 0.0124 a.u. eV^{-1} , indicating that while the GP contributes to data interpretation, it does not fully account for the features observed in the BO-informed experiment. The BO-informed approach is more effective at managing noise and keeping the features of the data from the 20 min per point standard experiment.

Despite using far fewer measurement points with shorter exposure time, the BO-informed experiments demonstrate their effectiveness. This suggests a more efficient method of data acquisition for GE-XANES and possibly conventional XANES applications. Our comparison with the reference standard experiment

highlights the potential and limitations of the BO-informed approach and provides insight into areas for further optimization and refinement.

However, as the results for the first layer indicate, the algorithm could benefit from further tuning to account for local maxima in the pre-peak region more effectively. One possible solution is to refine the optimization process by having the algorithm perform an NMF analysis on the initial angular intensity profile, subsequently determining the emission angles for the layers, and then focusing on the angle range of the desired layer during the optimization process or other solution can be using adaptive sampling functions such as mixed adaptive Sampling Algorithm (MASA).

Looking ahead, we will initiate further research to refine the BO-informed experimental process, focusing on the algorithm's sensitivity to local variations in the data. With iterative improvements, we anticipate this method could become even more accurate and efficient in producing accurate XANES spectra. Further studies could also explore the effect of different kernel functions in the GP model on the scanning results, in order to handle noisy and irregular patterns optimally and further improve the efficiency of BO-informed experiments.

4. Conclusion

The GE-XANES technique offers crucial insights into the local structure, oxidation states, and interlayers of thin films under ambient conditions. Nevertheless, the extensive duration required for experimental procedures has significantly limited its application within the field of materials science.

In response to this challenge, our study has showcased the effectiveness of BO in enhancing the data acquisition process for GE-XANES. In the example presented here, this represents a potential reduction in measurement time by a factor of 50. By integrating BO, we have effectively reduced the number of necessary data points and decreased the exposure time per point. These improvements have enabled more efficient conduct of XANES experiments, particularly in photon hungry scenarios. Such advancements have addressed the issue of lengthy acquisition times, which previously rendered GE-XANES analyses impractical for various applications.

The benefits of this methodological advancement are not restricted to GE-XANES but are also applicable to other laboratory-based x-ray absorption fine structure (XAFS) techniques and high-throughput experiments. This includes experiments associated with materials accelerator platforms, indicating a broad potential for impact across materials science research.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

GE-XANES experiments were performed at the mySpot beamline at the BESSY-II storage ring (Helmholtz Center Berlin). We thank the Helmholtz-Zentrum Berlin für Materialien und Energie for the allocation of synchrotron radiation beamtime. Cafer Tufan Cakir thanks the doctoral program of BAM for funding under the thematic project SURDIA (Surface Degradation Phenomena and Utilization of Innovative Alloy Systems). This manuscript is dedicated to the memory of Dr Uwe Reinholz (1956–2022), a valuable colleague and an outstanding scientist.

ORCID iDs

Cafer Tufan Cakir  <https://orcid.org/0000-0002-5033-9577>

Ana Guilherme Buzanich  <https://orcid.org/0000-0001-5543-9924>

Martin Radtke  <https://orcid.org/0000-0003-0014-8905>

References

- [1] Mirshojaeian Hosseini M J and Nawrocki R A 2021 A review of the progress of thin-film transistors and their technologies for flexible electronics *Micromachines* **12** 655
- [2] Green M A 2007 Thin-film solar cells: review of materials, technologies and commercial status *J. Mater. Sci. Mater. Electron.* **18** 15–19
- [3] Stephan-Scherb C, Schulz W, Schneider M, Karafiludis S and Laplanche G 2021 High-temperature oxidation in dry and humid atmospheres of the equiatomic CrMnFeCoNi and CrCoNi high- and medium-entropy alloys *Oxid. Met.* **95** 105–33
- [4] Cakir C T, Piotrowiak T, Reinholz U, Ludwig A, Emmerling F, Strelci C, Guilherme Buzanich A and Radtke M 2023 Exploring the depths of corrosion: a novel GE-XANES technique for investigating compositionally complex alloys *Anal. Chem.* **95** 4810–8

- [5] Yang Z, Sun P-F, Li X, Gan B, Wang L, Song X, Park H-D and Tang C Y 2020 A critical review on thin-film nanocomposite membranes with interlayered structure: mechanisms, recent developments, and environmental applications *Environ. Sci. Technol.* **54** 15563–83
- [6] Chung S, Cho K and Lee T 2019 Recent progress in inkjet-printed thin-film transistors *Adv. Sci.* **6** 1801445
- [7] Lei G, Lou C, Liu X, Xie J, Li Z, Zheng W and Zhang J 2021 Thin films of tungsten oxide materials for advanced gas sensors *Sens. Actuators B* **341** 129996
- [8] Park J W, Kang B H and Kim H J 2020 A review of low-temperature solution-processed metal oxide thin-film transistors for flexible electronics *Adv. Funct. Mater.* **30** 1904632
- [9] Moitzheim S, Put B and Vereecken P M 2019 Advances in 3D thin-film Li-ion batteries *Adv. Mater. Interfaces* **6** 1900805
- [10] Yao F-Z, Yuan Q, Wang Q and Wang H 2020 Multiscale structural engineering of dielectric ceramics for energy storage applications: from bulk to thin films *Nanoscale* **12** 17165–84
- [11] Vijayan K, Vijayachamundeswari S, Sivaperuman K, Ahsan N, Logu T and Okada Y 2022 A review on advancements, challenges, and prospective of copper and non-copper based thin-film solar cells using facile spray pyrolysis technique *Sol. Energy* **234** 81–102
- [12] Krishna D N G and Philip J 2022 Review on surface-characterization applications of x-ray photoelectron spectroscopy (XPS): recent developments and challenges *Appl. Surf. Sci. Adv.* **12** 100332
- [13] Finšgar M 2020 Electrochemical, 3D topography, XPS, and ToF-SIMS analyses of 4-methyl-2-phenylimidazole as a corrosion inhibitor for brass *Corros. Sci.* **169** 108632
- [14] Iatsunskiy I, Gottardi G, Micheli V, Canteri R, Coy E and Bechelany M 2021 Atomic layer deposition of palladium coated TiO₂/Si nanopillars: ToF-SIMS, AES and XPS characterization study *Appl. Surf. Sci.* **542** 148603
- [15] Morawski I and Nowicki M 2021 Properties of Au film on Ru (0001) revealed by AES, LEED, and DEPEs *Appl. Surf. Sci.* **569** 151100
- [16] Otto S-K, Moryson Y, Krauskopf T, Pepler K, Sann J, Janek J R and Henss A 2021 In-depth characterization of lithium-metal surfaces with XPS and ToF-SIMS: toward better understanding of the passivation layer *Chem. Mater.* **33** 859–67
- [17] Cornette P, Zanna S, Seyeux A, Costa D and Marcus P 2020 The native oxide film on a model aluminium-copper alloy studied by XPS and ToF-SIMS *Corros. Sci.* **174** 108837
- [18] Wang L, Mercier D, Zanna S, Seyeux A, Laurent-Brocq M, Perriere L, Guillot I and Marcus P 2020 Study of the surface oxides and corrosion behaviour of an equiatomic CoCrFeMnNi high entropy alloy by XPS and ToF-SIMS *Corros. Sci.* **167** 108507
- [19] Guilherme Buzanich A et al 2023 BAMline—a real-life sample materials research beamline *J. Chem. Phys.* **158** 244202
- [20] Lin S-C, Chang C-C, Chiu S-Y, Pai H-T, Liao T-Y, Hsu C-S, Chiang W-H, Tsai M-K and Chen H M 2020 Operando time-resolved x-ray absorption spectroscopy reveals the chemical nature enabling highly selective CO₂ reduction *Nat. Commun.* **11** 3525
- [21] Xiang S, Huang B, Li J, Liu Y, Marcella N, Routh P K, Li G and Frenkel A I 2022 Solving the structure of “single-atom” catalysts using machine learning–assisted XANES analysis *Phys. Chem. Chem. Phys.* **24** 5116–24
- [22] Martini A, Guda A, Guda S, Bugaev A, Safonova O and Soldatov A 2021 Machine learning powered by principal component descriptors as the key for sorted structural fit of XANES *Phys. Chem. Chem. Phys.* **23** 17873–87
- [23] Vermeulen M, McGeachy A, Xu B, Chopp H, Katsaggelos A, Meyers R, Alfeld M and Walton M 2022 XRFast a new software package for processing of MA-XRF datasets using machine learning *J. Anal. At. Spectrom.* **37** 2130–43
- [24] Kulow A, Buzanich A G, Reinholz U, Emmerling F, Hampel S, Fittschen U E A, Strelci C and Radtke M 2020 Comparison of three reconstruction methods based on deconvolution, iterative algorithm and neural network for x-ray fluorescence imaging with coded aperture optics *J. Anal. At. Spectrom.* **35** 1423–34
- [25] Moosavi S M, Jablonka K M and Smit B 2020 The role of machine learning in the understanding and design of materials *J. Am. Chem. Soc.* **142** 20273–87
- [26] Fujinuma N, DeCost B, Hatrick-Simpers J and Lofland S E 2022 Why big data and compute are not necessarily the path to big materials science *Commun. Mater.* **3** 59
- [27] Lee A, Sarker S, Saal J E, Ward L, Borg C, Mehta A and Wolverton C 2022 Machine learned synthesizability predictions aided by density functional theory *Commun. Mater.* **3** 73
- [28] Davariashitayani A, Kadkhodaie Z and Kadkhodaie S 2021 Predicting synthesizability of crystalline materials via deep learning *Commun. Mater.* **2** 115
- [29] Kuenneth C, Lalonde J, Marrone B L, Iverson C N, Ramprasad R and Paliana G 2022 Bioplastic design using multitask deep neural networks *Commun. Mater.* **3** 96
- [30] Mao K S, Gerczak T J, Harp J M, McKinney C S, Lach T G, Karakoc O, Nelson A T, Terrani K A, Parish C M and Edmondson P D 2022 Identifying chemically similar multiphase nanoprecipitates in compositionally complex non-equilibrium oxides via machine learning *Commun. Mater.* **3** 21
- [31] Samarakoon A, Tennant D A, Ye F, Zhang Q and Grigera S A 2022 Integration of machine learning with neutron scattering for the Hamiltonian tuning of spin ice under pressure *Commun. Mater.* **3** 84
- [32] Kusne A G et al 2020 On-the-fly closed-loop materials discovery via Bayesian active learning *Nat. Commun.* **11** 5966
- [33] Zhang Y, Xie R, Long T, Günzing D, Wende H, Ollefs K J and Zhang H 2023 Autonomous atomic Hamiltonian construction and active sampling of x-ray absorption spectroscopy by adversarial Bayesian optimization *npj Comput. Mater.* **9** 46
- [34] Teixeira Parente M, Brandl G, Franz C, Stuhr U, Ganeva M and Schneidewind A 2023 Active learning-assisted neutron spectroscopy with log-Gaussian processes *Nat. Commun.* **14** 2246
- [35] Jones D R, Schonlau M and Welch W J 1998 Efficient global optimization of expensive black-box functions *J. Glob. Optim.* **13** 455–92
- [36] Settles B 2011 From theories to queries: active learning in practice *Active Learning and Experimental Design Workshop in Conjunction with AISTATS 2010, JMLR Workshop and Conf. Proc.* pp 1–18
- [37] Burnaev E and Panov M 2015 Adaptive design of experiments based on Gaussian processes *Statistical Learning and Data Sciences: 3rd Int. Symp., SLDS 2015 (Egham, UK, 20–23 April 2015) Proceedings 3*
- [38] Liu H, Xu S, Ma Y, Chen X and Wang X 2016 An adaptive Bayesian sequential sampling approach for global metamodeling *J. Mech. Des.* **138** 011404
- [39] Houthby N, Huszár F, Ghahramani Z and Lengyel M 2011 Bayesian active learning for classification and preference learning (arXiv:1112.5745)
- [40] Rakotondrajoa A and Radtke M 2020 Machine learning based quantification of synchrotron radiation-induced x-ray fluorescence measurements—a case study *Mach. Learn.: Sci. Technol.* **2** 025004
- [41] Starostin V, Pithan L, Greco A, Munteanu V, Gerlach A, Hinderhofer A and Schreiber F 2022 End-to-end deep learning pipeline for real-time processing of surface scattering data at synchrotron facilities *Synchrotron Radiat. News* **35** 21–27

- [42] Campbell S I, Allan D B, Barbour A M, Olds D, Rakitin M S, Smith R and Wilkins S B 2021 Outlook for artificial intelligence and machine learning at the NSLS-II *Mach. Learn.: Sci. Technol.* **2** 013001
- [43] Konstantinova T, Maffettone P M, Ravel B, Campbell S I, Barbour A M and Olds D 2022 Machine learning enabling high-throughput and remote operations at large-scale user facilities *Digit. Discov.* **1** 413–26
- [44] Ueno T and Iwasawa H 2022 Measurement informatics in synchrotron radiation x-ray spectroscopy *Synchrotron Radiat. News* **35** 3–8
- [45] Shahriari B, Swersky K, Wang Z, Adams R P and De Freitas N 2015 Taking the human out of the loop: a review of Bayesian optimization *Proc. IEEE* **104** 148–75
- [46] Noack M M et al 2021 Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities *Nat. Rev. Phys.* **3** 685–97
- [47] Paris O, Li C, Siegel S, Weseloh G, Emmerling F, Rieseemeier H, Erko A and Fratzl P 2007 A new experimental station for simultaneous x-ray microbeam scanning for small- and wide-angle scattering and fluorescence at BESSY II *J. Appl. Crystallogr.* **40** s466–70
- [48] Zizak I 2016 The mySpot beamline at BESSY II *J. Large-Scale Res. Facil.* **2** A102
- [49] Scharf O et al 2011 Compact pnCCD-based x-ray camera with high spatial and energy resolution: a color x-ray camera *Anal. Chem.* **83** 2532–8
- [50] Mockus J B and Mockus L J 1991 Bayesian approach to global optimization and application to multiobjective and constrained problems *J. Optim. Theory Appl.* **70** 157–72
- [51] Schulz E, Speekenbrink M and Krause A 2018 A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions *J. Math. Psychol.* **85** 1–16
- [52] Rasmussen C E and Williams C K 2006 *Gaussian Processes for Machine Learning* vol 1 (Springer)
- [53] Garnett R 2023 *Bayesian Optimization* (Cambridge University Press)
- [54] Sacks J, Welch W J, Mitchell T J and Wynn H P 1989 Design and analysis of computer experiments *Stat. Sci.* **4** 409–23