Physics-Informed Neural Networks (PINNs) for Extended X-ray Absorption Fine Structure (EXAFS) Data Analysis: Fast and Accurate Local Atomic Structure Prediction

Latif U. Khan

Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME) P.O. Box 7, Allan 19252, Jordan

ABSTRACT

X-ray absorption fine structure (XAFS) is a crucial technique for probing the local atomic structure of advanced materials. Despite substantial global investment in XAFS beamlines at advanced synchrotron light sources, demand for this technique continues to grow rapidly due to its broad impact across scientific disciplines and industrial applications. However, the big data generated by modern XAFS beamlines at high-brilliance synchrotrons poses significant challenges for traditional XAFS data analysis methods. This can be addressed by developing machine learning-assisted data analysis techniques that leverage deep learning, particularly physics-informed neural networks (PINNs), to enable rapid, accurate, and automated interpretation of X-ray absorption fine structure (XAFS) data. PINNs model inserts the governing physics laws of EXAFS directly into the neural network training process, ensuring predictions that are both data-driven and physically consistent. The present work integrates PINNs with theoretical EXAFS simulations using FEFF and the XrayLarch package, enabling real-time prediction of local atomic structure from experimental EXAFS data even under noisy or data-limited conditions.

KEYWORDS: EXAFS, AI-deep learning model, PINNs, Real time prediction, Local atomic structure

1. INTRODUCTION

X-ray Absorption Spectroscopy (XAS) is an indispensable technique for unraveling the intricate local atomic and electronic structure of advanced materials [1], ranging from catalysts to battery electrodes, nanophotonic and quantum materials, which show remarkable applications in advanced technologies [2]. XAS provides element-specific probe to gain insight into the oxidation states and local coordination environment of the photoabsorber sites in materials' lattices. However, with the introduction of ultrafast and highly sensitive detection systems at the latest generation high-brilliance synchrotron sources, XAS beamlines produce vast amounts of data in minimal time. For example, quick XAFS beamlines [3] in *in-situ* XAFS experiments generate large volumes of data. This complexity of XAS data poses inherent challenges for traditional X-ray absorption fine structure (XAFS) data analysis techniques [4–6], creating a significant bottleneck in accelerating advanced materials discovery and characterization.

The integration of deep learning models in XAS data analysis provides a transformative force to unlock unprecedented capabilities in materials science, revolutionizing the interpretation of complex and larger datasets of XAFS into faster, smarter and more accurate insights [7]. Recent advancements in AI, particularly machine learning/deep learning [8], have opened new frontiers in scientific data analytics across a wide discipline of sciences, including X-ray scattering and X-ray absorption spectroscopy techniques [9]. These machine learning techniques are promising for tackling the challenges associated with XAS data analysis, where complexity of big data often exceeds the capacity of traditional analysis methods.

Artificial deep learning models [8] encompass a range of algorithms that enable computers to identify patterns and make predictions or decisions without being explicitly programmed. In this framework, supervised learnings [10] are widely used for tasks such as XAFS spectral fitting and

classification [7]. Deep learning neural networks with multiple hidden layers, learns hierarchical features from the large, complex big datasets of XAS that can be efficiently used for X-ray absorption fine structure (XAFS) prediction of unseen materials. For instance, convolutional neural networks (CNNs) and physics informed neural networks (PINNs) [11], for example, have shown significant potential in interpreting XAS spectra, automating feature extraction, and even predicting local structural parameters directly from the spectral data.

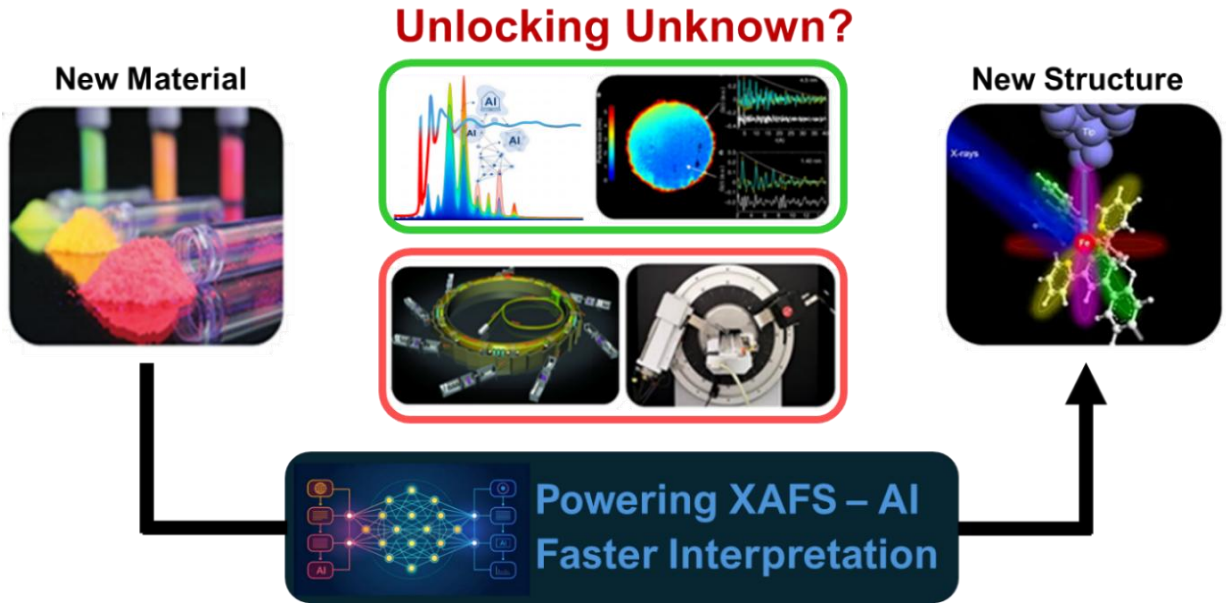

Figure 1. Powering XAFS data interpretation/analysis with artificial intelligence (AI) based machine learning/deep learning models for unlocking the structures of new materials.

In scientific literature, multiple efforts have been dedicated to applying the deep neural networks (DNNs) for simulating the XANES and EXAFS [9,10,12], predicting the oxidation state, the local chemical environment of the absorbing atom, three-dimensional (3D) structures, nanoparticle size and structural transformation in nanoparticles [10,12,13], such as single atom catalysts [7]. However, the fast prediction of the local atomic structure for the experimental EXAFS data of new advanced materials (Figure 1), with remarkable precision ultimate result using

pretrained AI based deep learning model and theoretical EXAFS database is highly demanded in the user's community of the XAS worldwide.

The integration of machine learning into XAS workflows (Figure 1) not only accelerates data processing but also enables real-time analysis, automated interpretation, and high-throughput screening, thereby facilitating rapid insights into local structure. These capabilities are crucial for advancing *in-situ* and *in-operando* XAFS experiments, where timely feedback is essential for guiding experimental design and decision-making. Herein, we trained a PINNs model [11] on the XAFS data of a wide number of advanced materials, which was integrated with FEFF code [4] and Xraylarch package [14] to predict the EXAFS efficiently for the photoabsorber site of new unseen materials.

2. METHODS AND RESULTS

2.1. Physics Informed Neural Networks (PINNs)

Physics-informed neural networks (PINNs) are a revolutionary class of neural networks that bridge the gap between machine learning and laws of physics [11]. They are trained to solve supervised learning tasks by integrating physics laws, typically expressed as general nonlinear partial differential equations (PDEs) directly into the neural network training process [15]. Unlike the traditional data-driven neural networks, which rely solely on input-output pairs to learn patterns/spectral features and try to generalize to new unseen data, PINNs are designed to respect known physical laws [15,16], ensuring that the learned solutions are consistent with the underlying physics. In many scientific and engineering domains, such as thermodynamics, fluid dynamics, and electromagnetism, the governing equations of a system are well established.

The purely data-driven approaches might ignore this valuable information, potentially leading to physically inconsistent or inaccurate predictions, especially with limited or noisy data. In such low-data regimes [15], conventional deep learning models (e.g., convolutional or recurrent neural networks) tend to overfit, lack robustness, and offer no guarantees of convergence. These data-driven models typically require large, high-quality datasets to perform reliably, and in the absence of sufficient data, they may overfit, produce unstable predictions, or yield physically inconsistent results. To address these challenges, physics-informed neural networks (PINNs) have emerged as a powerful alternative [11,15,16]. PINNs address these limitations by embedding physical constraints encoded as PDEs into the network' loss function. This guides the training process toward solutions that are both data-compliant and physically valid. As a result, PINNs can achieve better generalization, stability, and accuracy, particularly in scenarios involving complex and challenging problems. The fusion of physics laws and data makes PINNs a powerful tool for solving real problems where traditional models fall short.

Loss Functions in PINNs (Data and Physics Constraints): The key innovation in PINNs model lies in the composite loss function, which typically consists of two main components: *i)* data loss (MSE_u) that ensures the model's predictions to match the observed data (e.g., initial and boundary conditions), and *ii)* physics informed loss (MSE_f), which enforces the model to satisfy the physical laws (e.g., PDE residuals) at collocation points [11].

$$MSE = MSE_u + MSE_f \tag{1}$$

Data loss (MSE_u) [11]:

$$MSE_\mu = \frac{1}{N_\mu}\sum_{i=1}^{N_\mu}\left|\mu\left(t_\mu^i, x_\mu^i\right) - \mu_i\right|^2$$

$\{t_u^i, x_u^i, u^i\}_{i=1}^{N_u}$ are training data points from initial and boundary conditions, and $u(t,x)$ is the neural network's output evaluated at these data points.

Physics loss (MSE_f) [11]:

$$MSE_f = \frac{1}{N_f}\sum_{i=1}^{N_f}\left|f(t_f^i, x_f^i)\right|^2$$

Where, $\{t_f^i, x_f^i\}_{i=1}^{N_f}$ are collocation points, chosen in the domain where the physical constraint (PDE) is enforced. The $f(t,x)$ represents the PDE residual, defined by substituting the neural network's approximation $u(t,x)$ into the PDE. The function $f(t,x)$ can typically be computed via automatic differentiation, which enables the network to evaluate derivatives of $u(t,x)$ needed to define the PDE.

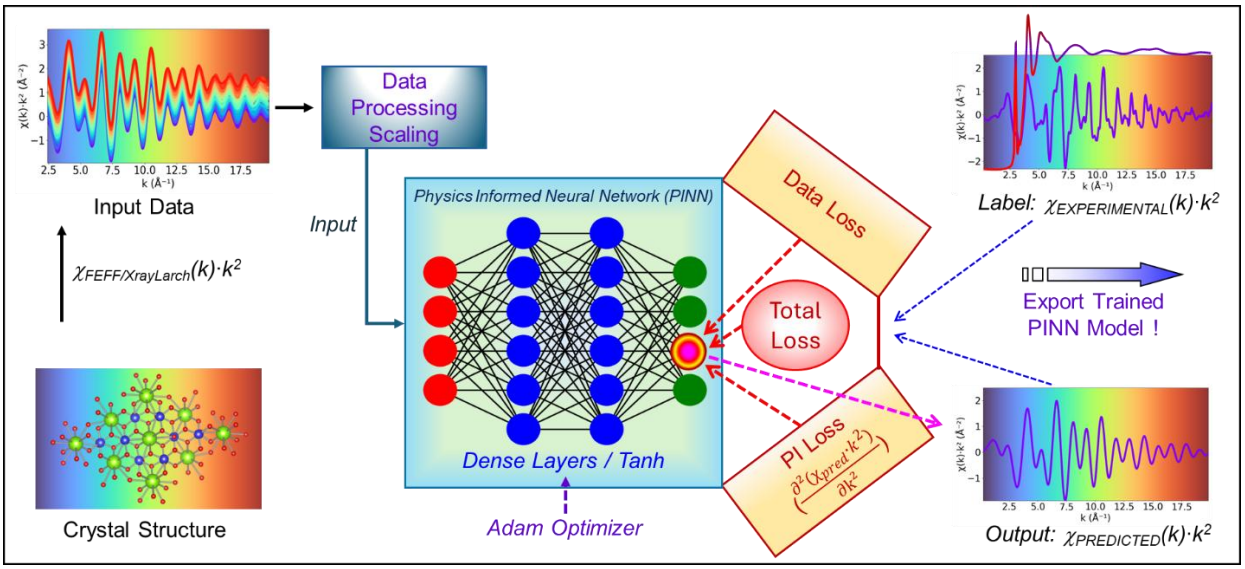

Figure 2. Step-by-step training of an artificial deep learning PINNs model on large EXAFS datasets for predicting theoretical $\chi(k)\cdot k^2$ and mapping it to unknown experimental EXAFS data [$\chi(k)\cdot k^2$], enabling rapid determination of local atomic structure in advanced materials.

2.2. Implementation of PINN Model in EXAFS Data Prediction

X-ray absorption fine structure (XAFS) is a powerful technique used to probe the local atomic structure around a specific element in a material's lattice [1]. The extended X-ray absorption fine structure (EXAFS) signal, typically denoted by χ(k), contains oscillatory features that encode information about interatomic distances, coordination numbers, and lattice's disorder [5]. The physical behavior of EXAFS is governed by well-known theoretical models, such as the EXAFS equation derived from quantum scattering theory. Traditional approaches to EXAFS data analysis involve fitting experimental EXAFS data with theoretical models (e.g., using FEFF calculations, such as Artemis from Demeter) [4,6,17], often requiring expert knowledge, and extensive manual tuning. These methods can be sensitive to noise and are limited in scalability when analyzing large datasets. Physics-informed neural networks provide a modern alternative by incorporating the known physics of EXAFS, such as the oscillatory behavior described by χ(k), and its Fourier-transformed counterpart χ(R) directly into the learning process. For example, a PINN can be trained to predict χ(k) from the input theoretical data after training, while ensuring that the predicted spectrum satisfies physical constraints, such as smoothness, energy conservation, and boundary conditions derived from the EXAFS equation 2 [1]. By embedding these constraints into the loss function, PINNs enable robust EXAFS spectrum prediction, denoising, and even reconstruction from incomplete data. This approach enhances accuracy, reduces reliance on large datasets, and offers a scalable, automated alternative for EXAFS analysis in complex materials systems.

$$\chi(k) = \sum_j S_0^2 \frac{N_j}{kR_j^2} \left| f_j\left(k, R_j\right) \right| e^{-2R_j/\lambda_j(k)} e^{-2k^2\sigma_j^2} \sin\left[2kR_j + \phi_j(k)\right] \qquad (2)$$

where, S_0^2 is the passive electrons' amplitude reduction factor, N_j is the number of backscattering neighboring atoms, R_j is the average distance between the central photoabsorber and the

backscatterers, $f_j(k)$ is the effective scattering amplitude, σ_j^2 is the correlated Debye-Waller factor, defining the mean square relative displacement (MSRD) along the equilibrium path length, $\phi_j(k)$ is the effective total phase shift, including contributions from the central atom and all scattering atoms, and λ_j is the mean free path of the photoelectron.

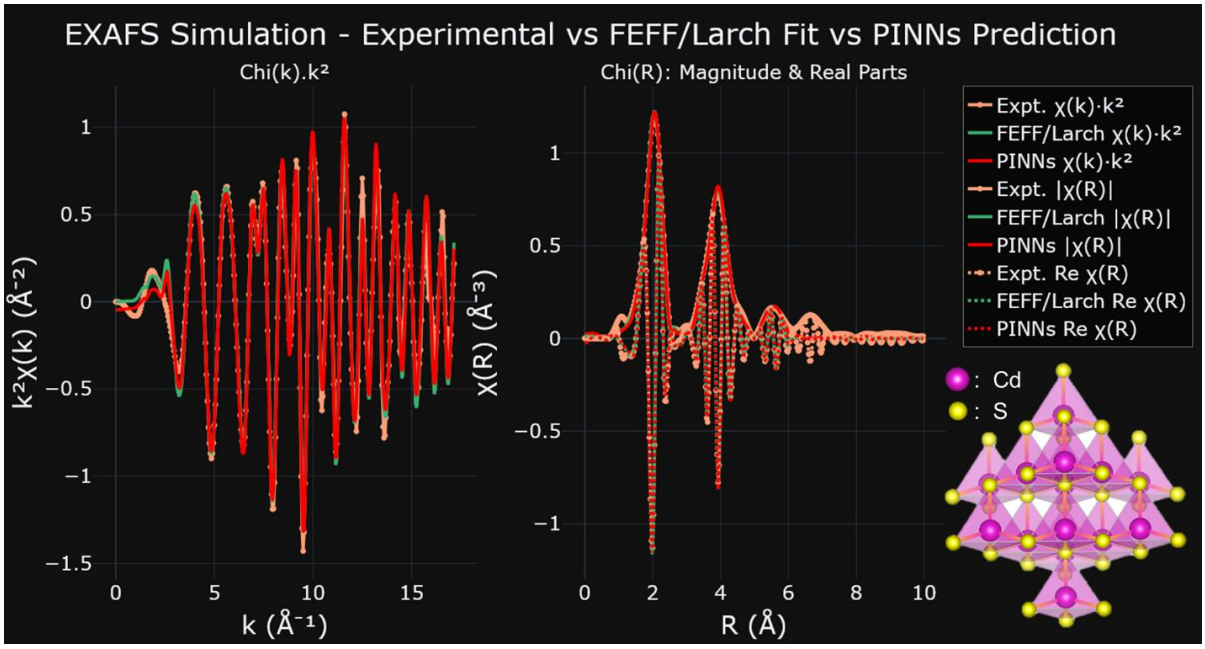

Figure 3. Comparison of the EXAFS fit performed by XrayLarch with PINNs prediction for the CdS NCs. The left panel shows the χ(k)·k² signals, and the right panel displays the magnitude and real part of the Fourier-transformed EXAFS signal [χ(R)]. The inset figure depicts the atomic structure of CdS used in the EXAFS simulation.

The application of a physics-informed neural network (PINN) architecture to the EXAFS data analysis is summarized as (Figure 2): The input datasets containing theoretical $\chi(k)\cdot k^2$ signals of the various metal sites in different materials (e.g., Cu in CuO) are derived from their crystallographic structures, by well-fitting their experimental $\chi(k)\cdot k^2$, using XrayLarch. These input data are then preprocessed and scaled before being fed into the neural network (PINNs),

comprising multiple dense layers with Tanh activation functions. The PINNs model is trained on the larger datasets of theoretical $\chi(k)\cdot k^2$ as input data with the corresponding experimental $\chi(k)\cdot k^2$ as labels, using an Adam optimizer to minimize a composite total loss, which is a weighted sum of the data loss and the physics-informed loss (Equation 3).

TensorFlow's automatic differentiation calculates the PI loss based on the second derivative of the network output (predicted $\chi(k)\cdot k^2$) with respect to the photoelectron wave number k, $\frac{\partial^2(\chi_{pred}\cdot k^2)}{\partial k^2}$. The PI loss is added to the total loss, forcing the model's output to be physically plausible by penalizing unrealistic oscillations or behavior in the XAFS spectrum [16].

$$L_{total} = \lambda_{data}\cdot L_{data} + \lambda_{PI}\cdot L_{PI} \quad (3)$$

Where L_{data} is the mean squared error (MSE_μ) between the predicted and true values, L_{PI} is the physics-informed loss (MSE_f). The weights, λ_{data} and λ_{PI} can be adjusted to balance the importance of fitting the data *versus* respecting the physical constraint. The training process tracks both the training and validation total losses, including the data loss and physics-informed loss components. The trained PINN model is then used to predict the $\chi(k)\cdot k^2$ output for the new unseen EXAFS data with improved smoothness and physical consistency, facilitating the accurate EXAFS data interpretation with possible structural model. Finally, the pretrained PINNs model on the big EXAFS data is exported and used together with Xraylarch in advanced programming language Python to probe and predict precisely the local atomic structure order (Figure 3) of the wide range of metal sites in advanced materials.

3. CONCLUSIONS

Globally, approximately 60 synchrotron facilities and free-electron lasers are currently in operation. It is estimated that 10–20% of advanced scientific experiments at these facilities utilize

XAFS techniques. Over the past three decades, billions of dollars have been invested in developing and operating XAFS capabilities at synchrotron radiation sources. The latest generation of high-brilliance synchrotron sources, equipped with advanced beamlines such as quick XAFS beamlines, can generate large volumes of XAS data within extremely short acquisition times. This rapid data generation poses significant challenges for traditional XAS data analysis tools (e.g., Athena/Artemis from the Demeter suite), which are not designed to efficiently handle or interpret such large datasets. Moreover, these conventional tools often require a high level of expertise and months of effort for quantitative XAFS data analysis, creating a substantial bottleneck in the timely characterization and discovery of advanced materials, ultimately slowing scientific progress and innovation. To overcome these limitations, there is a pressing need to develop an artificial intelligence-driven XAFS data analysis platform, utilizing deep learning neural networks implemented in modern programming environments such as Python. The present work of the author is a pioneering effort in successfully implementing PINN model to EXAFS data analysis, using Python and initial application to the prediction of theoretical EXAFS data with confirmation of local atomic structures for specific advanced materials, such as CdS, ZnS, ZnO, Cu, $LiCoO_2$, etc., applying in lighting/display devises, catalysis, renewable energy devices (e.g., batteries) etc.

Supporting Materials: Please find the Python codes in GitHub's link: https://github.com/khanlatif001/PINNmodel_EXAFS-Prediction for training of PINN model on EXAFS datasets and deploying to predicting EXAFS

REFERENCES

[1] L.U. Khan, Z.U. Khan, L. Blois, L. Tabassam, H.F. Brito, S.J.A. Figueroa, Strategy to Probe the Local Atomic Structure of Luminescent Rare Earth Complexes by X-ray Absorption

Near-Edge Spectroscopy Simulation Using a Machine Learning-Based PyFitIt Approach, Inorg Chem 62 (2023) 2738–2750. https://doi.org/10.1021/acs.inorgchem.2c03850.

[2] L.U. Khan, N. Jabeen, I. Jabbar, S. Jamil, A. Kanwal, Z. Akhter, M. Usman, M.Z. Abid, M. Harfouche, Investigating Local Structure of Ion-Implanted (Ni^{2+}) and Thermally Annealed Rock Salt CoO Film by EXAFS Simulation Using Evolutionary Algorithm, ACS Appl Energy Mater 4 (2021). https://doi.org/10.1021/acsaem.0c02676.

[3] S.J.A. Figueroa, A. Rochet, I. Ferreira Torquato, A.M. Espíndola, H. Rigamonti, B.C. Meyer, G. de M. Azevedo, QUATI beamline: QUick x-ray Absorption spectroscopy for TIme and space-resolved experiments at the Brazilian Synchrotron Light Laboratory, Radiation Physics and Chemistry 212 (2023) 111198. https://doi.org/10.1016/j.radphyschem.2023.111198.

[4] J.J. Rehr, J.J. Kas, F.D. Vila, M.P. Prange, K. Jorissen, Parameter-free calculations of X-ray spectra with FEFF9, Physical Chemistry Chemical Physics 12 (2010) 5503. https://doi.org/10.1039/b926434e.

[5] J.J. Rehr, R.C. Albers, Theoretical approaches to x-ray absorption fine structure, Rev Mod Phys 72 (2000) 621–654. https://doi.org/10.1103/RevModPhys.72.621.

[6] J.J. Rehr, J.J. Kas, M.P. Prange, A.P. Sorini, Y. Takimoto, F. Vila, Ab initio theory and calculations of X-ray spectra, C R Phys 10 (2009) 548–559. https://doi.org/10.1016/j.crhy.2008.08.004.

[7] T. Ying, R. Zhu, Y. Wang, Y. Zhu, Artificial Intelligence Guided Single Atom Catalysts: From Precise Design to Industrial-Grade Preparation, Adv Funct Mater n/a (2025) e15026. https://doi.org/10.1002/adfm.202515026.

[8] I.H. Sarker, Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Comput Sci 2 (2021) 160. https://doi.org/10.1007/s42979-021-00592-x.

[9] N. Marcella, Y. Liu, J. Timoshenko, E. Guan, M. Luneau, T. Shirman, A.M. Plonka, J.E.S. Van Der Hoeven, J. Aizenberg, C.M. Friend, A.I. Frenkel, Neural network assisted analysis of bimetallic nanocatalysts using X-ray absorption near edge structure spectroscopy, Physical Chemistry Chemical Physics 22 (2020). https://doi.org/10.1039/d0cp02098b.

[10] J. Timoshenko, A. Anspoks, A. Cintins, A. Kuzmin, J. Purans, A.I. Frenkel, Neural Network Approach for Characterizing Structural Transformations by X-Ray Absorption Fine Structure Spectroscopy, Phys Rev Lett 120 (2018). https://doi.org/10.1103/PhysRevLett.120.225502.

[11] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial

differential equations, J Comput Phys 378 (2019) 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

[12] J. Timoshenko, C.J. Wrasman, M. Luneau, T. Shirman, M. Cargnello, S.R. Bare, J. Aizenberg, C.M. Friend, A.I. Frenkel, Probing Atomic Distributions in Mono- and Bimetallic Nanoparticles by Supervised Machine Learning, Nano Lett 19 (2019). https://doi.org/10.1021/acs.nanolett.8b04461.

[13] E.-S. Jeong, I.-H. Hwang, S.-W. Han, Quantitative analysis of EXAFS data sets using deep reinforcement learning, Sci Rep 15 (2025) 17417. https://doi.org/10.1038/s41598-025-94376-5.

[14] M. Newville, Larch: An analysis package for XAFS and related spectroscopies, in: J Phys Conf Ser, 2013. https://doi.org/10.1088/1742-6596/430/1/012007.

[15] S. Cuomo, V.S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, F. Piccialli, Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next, J Sci Comput 92 (2022) 88. https://doi.org/10.1007/s10915-022-01939-z.

[16] L.D. McClenny, U.M. Braga-Neto, Self-adaptive physics-informed neural networks, J Comput Phys 474 (2023) 111722. https://doi.org/10.1016/j.jcp.2022.111722.

[17] B. Ravel, M. Newville, ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT, J Synchrotron Radiat 12 (2005) 537–541.