

A Universal Deep-Learning Framework for Materials X-ray Absorption Spectra

Shubha R. Kharel,^{1,*} Fanchen Meng,² Xiaohui Qu,² Matthew R. Carbone,^{1,†} and Deyu Lu^{2,‡}

¹*Computing and Data Sciences Directorate, Brookhaven National Laboratory, Upton, New York 11973, USA*

²*Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973, USA*

(Dated: November 14, 2024)

X-ray absorption spectroscopy (XAS) is a powerful characterization technique for probing the local chemical environment of absorbing atoms. However, analyzing XAS data presents with significant challenges, often requiring extensive, computationally intensive simulations, as well as significant domain expertise. These limitations hinder the development of fast, robust XAS analysis pipelines that are essential in high-throughput studies and for autonomous experimentation. We address these challenges with OmniXAS, a framework that contains a suite of transfer learning approaches for XAS prediction, each uniquely contributing to improved accuracy and efficiency, as demonstrated on K-edge spectra database covering eight 3d transition metals (Ti – Cu). The OmniXAS framework is built upon three distinct strategies. First, we use M3GNet [*Nat. Comput. Sci.*, 2, 718 (2022)] to derive latent representations of the local chemical environment of absorption sites as input for XAS prediction, achieving up to order-of-magnitude improvements over conventional featurization techniques. Second, we employ a hierarchical transfer learning strategy, training a universal multi-task model across elements before fine-tuning for element-specific predictions. Models based on this cascaded approach after element-wise fine-tuning outperform element-specific models by up to 69%. Third, we implement cross-fidelity transfer learning, adapting a universal model to predict spectra generated by simulation of a different fidelity with a much higher computational cost. This approach improves prediction accuracy by up to 11% over models trained on the target fidelity alone. Our approach significantly boosts the throughput of XAS modeling by orders of magnitude as compared to first-principles simulations and is extendable to XAS prediction for a broader range of elements. The proposed transfer learning framework is generalizable to enhance deep-learning models that target other properties in materials research.

I. INTRODUCTION

X-ray absorption spectroscopy (XAS) is a widely used materials characterization technique in a broad array of scientific research fields [1–3], such as condensed matter physics, materials science, chemistry and biology. XAS is element-specific and its near-edge region, known as X-ray absorption near-edge structure (XANES), contains rich information of the local chemical environment of the absorbing site (e.g., oxidation state, coordination number and local symmetry). Therefore, XANES measurements provide important insights into the structural and electronic properties of the sample, which are needed for mechanistic understanding of the underlying physical and chemical processes.

However, XANES analysis is non-trivial, as the spectral function is a complex representation of the underlying atomic structure and electronic structure associated with the short-range order. Traditionally, extracting the information from XANES spectral features heavily relies on empirical fingerprints [4] from well-established experimental standards and/or first-principles simulations. These traditional XANES analysis approaches require strong domain expertise and, when combined with first-principles simulations, can be computationally demanding, creating a practical barrier for many researchers.

On the other hand, rapid advances in synchrotron X-ray facilities enables XAS measurements with high time and energy resolution. For example, using a quick XAS scanning method, a spectrum can be measured with 10 ms time resolution [5]; the energy resolution in high-energy-resolution fluorescence-detected (HERFD) XAS can reach below the core-hole lifetime broadening [6]. Such experimental instrumental development creates the demand of XANES analysis of large scale temporal data. In addition, the deployment of autonomous experimentation pipelines [7–9] makes a strong case for real-time data analysis. To address these emerging challenges, a robust data-driven XANES analysis approach is needed to lower the barrier to entry for non-experts, reduce computational cost and accelerate throughput. In practice, a data-driven XANES pipeline requires multiple key building blocks, including (a) workflow software for high-throughput XANES simulations, (b) large and diverse simulated XANES spectra databases, and (c) tailored machine learning (ML) models that capture the structure-spectrum relationship.

In the past several years, significant progress has been made in this field. Systematic multi-code XANES benchmarks [10] were carried out to quantify the effects of key approximations and implementations in simulation, as well as to determine the converged parameters for spectral simulations. Workflow software [11–13] has also been developed to standardize the spectral input file generation and ensure data reproducibility. Concurrently, the corpus of open-access, publicly available simulated XANES spectra databases [14–20] continue to grow, thus

* skharel@bnl.gov

† mcarbone@bnl.gov

‡ dlu@bnl.gov

facilitating the development of ML models for XANES analysis.

ML models trained to predict spectra from atomistic structure, i.e. spectroscopy surrogate models, are of particular importance, since once trained, they can bypass expensive first-principles simulations. In the same spirit of neutral network potentials (NNPs) [21], which are used to predict energies and forces from atomistic geometries, ML models can be trained on first-principles data to predict XANES spectra (or more broadly, excited state properties, e.g. spectral functions) with the accuracy of *ab initio* methods but at a fraction of the computational cost. Spectroscopy ML surrogates therefore make it feasible to tackle complex materials with up to thousands atoms in a supercell. Several ML models have been developed for XANES, such as graph neural network (GNN) and multilayer perceptron (MLP) models for C, N, and O K-edge XANES [22, 23] of the QM9 molecular database [24], C K-edge XANES of amorphous carbon [25, 26], K-edge XANES of Fe compounds [27] and 3d transition metal complexes [18]. As these models represent a limited scope in materials and chemical space (e.g., finite systems and a small sub-space of the extended systems), there is still an unmet need to develop *general* ML models to predict XANES in a broad materials space.

In this work, we address the need for a generalizable ML framework in the XANES domain. Our approach leverages transfer learning and domain adaptation techniques to develop universal models. By adopting M3GNet-derived latent features [28] to predict K-edge XANES spectra for a range of 3d transition metals, we demonstrate that this method not only outperforms traditional featurization techniques but also effectively bridges elements and fidelities, providing a scalable solution for XANES analysis across a wide range of materials. The OmniXAS framework lays the groundwork for future applications in high-throughput and real-time XAS analysis, significantly reducing the reliance on computationally expensive simulations.

II. METHOD

In this work, we demonstrate how a variety of deep learning models can be used to accurately predict the XANES spectra of a large class of materials. Fig. 1 highlights our workflow of the OmniXAS framework, which includes data acquisition, data curation, ML model architecture, and model training and evaluation.

To create a database of paired structures and the corresponding XANES spectra, as shown in Fig. 1a we pulled structural data from the Materials Project [27], generated spectral simulation input files using the Lightshow package [10, 13], and performed spectral simulations, post-processing and data curation (Sec. II A). Two levels of theory were used in the spectral simulation: the real-space multiple scattering method implemented in

FEFF9 [29] and the excited-electron core-hole potential method implemented in the Vienna Ab initio Simulation Package (VASP) [30]. This module ultimately produces the structure-spectrum pairs which are used as the training, validation and testing data in our ML models.

Next, multiple transfer learning ML models are developed for site-spectrum prediction of each element, where the featurization of the local structure of the absorbing site was taken from the M3GNet [28] via a feature transfer process (Sec. II B- II C and Fig. 1b). We underscore that a universal model can be developed using the data of all eight elements and subsequently fine-tuned on specific element to further improve the model performance (Sec. II B and Fig. 1c). Details of the model architectures are explained in Sec. II D. Training and evaluation of these models, such as partitioning the data into training, validation and testing splits, are described in Sec. II E.

A. Data Acquisition and Curation

Before ML models can be trained, structure and spectral data need to be sourced, curated, and standardized. We used the Lightshow Python package [10, 13] to pull all available Materials Project structural data of pure compounds, and oxides containing at most 0, 1 or 2 other unique atom types via Pymatgen [31] for Ti, V, Cr, Mn, Fe, Co, Ni and Cu. FEFF9 and VASP spectral input files were also written using Lightshow.

1. Input File Generation

FEFF calculations were performed in the real space, and the random phase approximation was used for core-hole screening. Self-consistent field and full multiple scattering lengths were set to 7 and 9 Å, respectively. In VASP spectral simulations, both ground state and core-hole excited state were calculated under the non-spin polarized configuration. We used the GW-type pseudopotential, in order to obtain a good description of high-energy scattering states [10]. A full core hole was included in spectral simulations and the core electron was placed at the bottom of the conduction band. The k -point mesh used in the Brillouin zone sampling is an important convergence parameter of the VASP spectral calculation, which was determined using the effective crystal size method [10] with the *cutoff* parameter set at 24 Å in Lightshow. The total number of bands (n_b) included in the spectral calculation depends on system size and the chosen energy range. Here we estimated n_b based on the uniform electron gas model [10] with the *e_range* parameter set to 30 eV (20 eV) for Ti (Cu) in Lightshow. The computational cost grows linearly with n_b . To make the high-throughput calculations tractable, we only considered materials with n_b less than 2200 in VASP simulations.

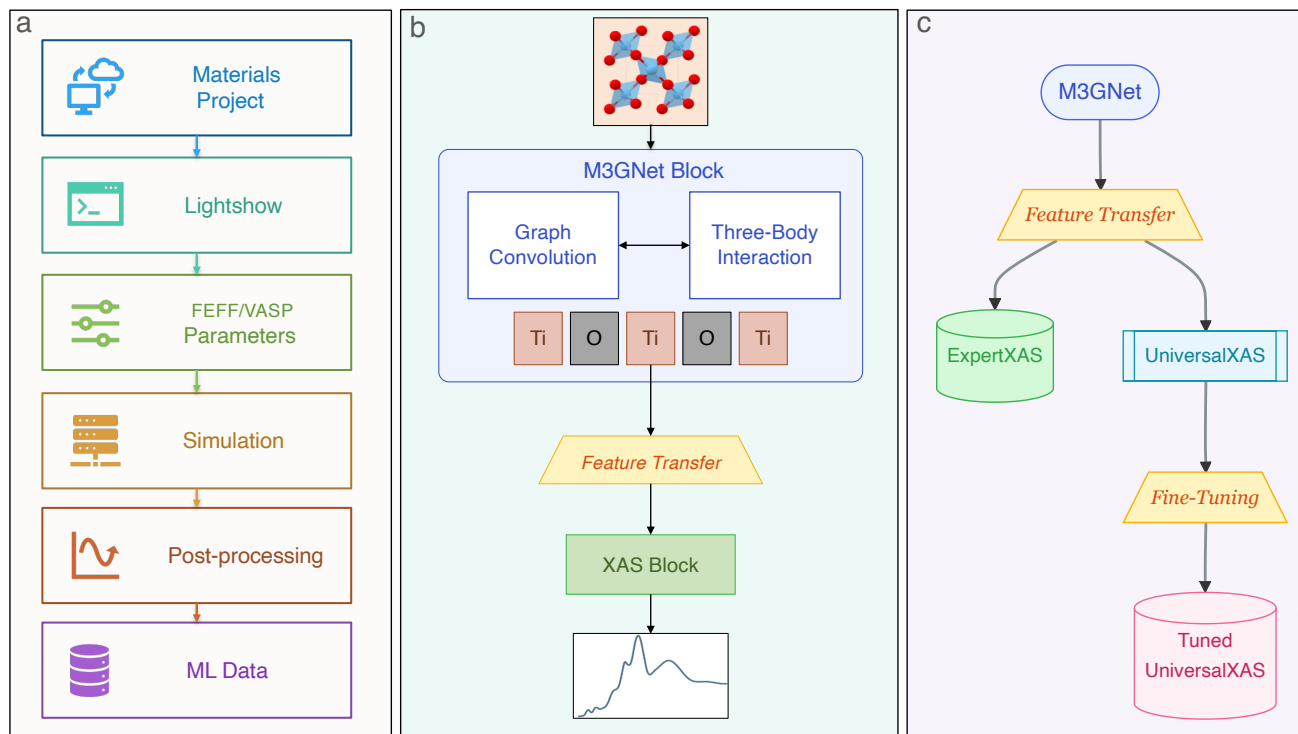


FIG. 1. Schematic of the OmniXAS framework. a) Data Curation: Structural data are sourced from the Materials Project, from which FEFF and VASP spectral input files are generated locally using Lightshow. FEFF and VASP spectral simulations are performed using these input files, and the results are screened and processed into machine learning-ready data. b) XAS Model: Materials structure information is processed through frozen M3GNet blocks using a series of graph convolutions with three-body interaction updates. From these operations, only the latent state at the node of the absorption site is passed into trainable neural networks for predicting site-specific XAS spectra. c) Cascaded Transfer Learning: The workflow diverges into two paths for training the XAS-block, resulting in three variants of XAS models. In one path, individual models for each data subset (ExpertXAS) are trained. In the other, a single model that predicts FEFF spectra for all elements (UniversalXAS) is developed. Further along, knowledge transfer from the UniversalXAS model is applied through fine-tuning, producing another set of specialized models for each subset that we call Tuned-UniversalXAS.

2. Removal of Unphysical Spectra and Outliers

In the first data standardization step, site-specific spectra that failed sanity checks were discarded, when the FEFF calculations did not converge with the default input. Next we adopt a statistical approach to remove outliers. Define $\boldsymbol{\mu}^{(i)} = [\mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_M^{(i)}]$ the i -th simulated spectrum (i.e., ground truth of the ML models) discretized on M energy grid points. The absorption coefficient at grid point j averaged over N_{tot} spectra is given by

$$\bar{\mu}_j = \frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} \mu_j^{(i)}, \quad (1)$$

and the corresponding standard deviation (without bias correction, i.e., using N_{tot} instead of $N_{tot} - 1$ in the normalization [32, 33]) is given by

$$\sigma_j = \sqrt{\frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} (\bar{\mu}_j - \mu_j^{(i)})^2}. \quad (2)$$

Any spectrum with $|\mu_j^{(i)} - \bar{\mu}_j| > c \sigma_j$ for any $j = 1, \dots, M$ is removed from the database. In other words, spectra where the absorption coefficient lies outside of the envelope set by a chosen multiple of the standard deviation are removed. For FEFF and VASP spectra, we used standard deviation thresholds of $c = 2.5$ and 5 , respectively. These threshold choices were determined empirically to remove the significant outliers observed in spectra heat maps while retaining most of the data.

3. Rescaling and Edge Alignment

The raw output of the FEFF spectra under the default normalization was re-scaled to obtain the absolute cross section with default broadening. The output of the VASP spectra is the imaginary part of the macroscopic dielectric constant. To compare results between FEFF and VASP, we converted the imaginary part of the dielectric constant to cross section [10]. Raw VASP spectra were calculated purposely with a small broadening of 0.05 eV. This allows one to further broaden the spectra in

post-processing as needed. We applied a 0.89 (3.46) eV core-hole lifetime broadening to Ti (Cu) VASP spectra, such that the spectral resolutions in FEFF and VASP are comparable.

Relative edge alignment was applied to VASP site-specific spectra using the Δ SCF method [10, 34]. After that, a constant empirical shift of 5108.59 (9492.80) eV was applied to VASP Ti (Cu) spectra to align with reference experimental spectrum of anatase TiO_2 [35] (cuprite Cu_2O [36]). In Sec. II C, we need to further align FEFF and VASP spectra, and an additional constant shift of -5.5 (7.0) eV was applied to VASP Ti (Cu) spectra such that the average spectra in VASP align with its FEFF counterparts. After post-processing, we consider the same energy range of 35 eV for all the spectra spanned on a uniform energy grid of $\delta E = 0.25$ eV from different starting energies determined empirically for each element (see Supplementary Table S1). This results in a 141-dimension vector that represents each spectrum. Values of the relative energy ΔE are defined on this uniform energy grid as $\Delta E_j = \delta E \times j$, where j is the grid index.

The final dataset after curation and standardization consists of 75,691 site-specific spectra across all eight elements from two simulation methods, which we refer to as the 3d transition metal XANES dataset, “3dtm_xanes_ml2024”. The detailed data size breakdown is summarized in Fig. 2 and Supplementary Table S2.

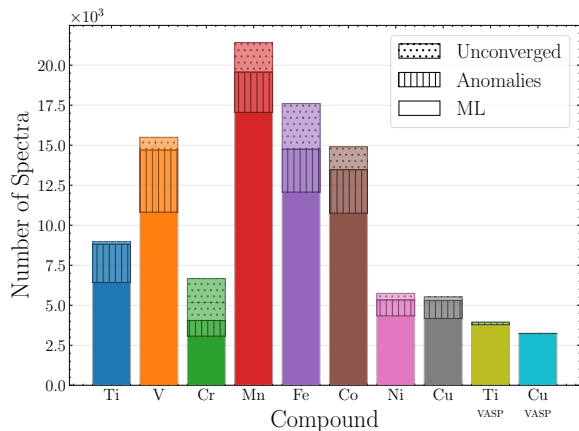


FIG. 2. Number of spectra generated for each element from FEFF and VASP simulations. The highest bars represent the total number of spectra, with partitions indicating the portions removed during different data cleaning stages (unconverged and anomalies) and the remaining ML-ready data.

B. XAS Prediction Hypotheses

We propose two key hypotheses for XAS prediction that guide the rest of the methodologies for ML model design:

- *Feature Transfer Hypothesis*: The structural features in ML models trained for materials interatomic potentials provide sufficient transferable knowledge for XAS prediction models.
- *Universal XAS Model Hypothesis*: A universal XAS model, trained on diverse elements, can extract core knowledge of the structure-spectrum relationship that is also transferable to improve element-specific predictive models.

These two hypotheses lead us adopt a cascaded transfer learning approach when developing our ML models.

C. Transfer Learning

Transfer learning draws inspiration from human cognition, where knowledge acquired from one task can be applied to another [37]. This knowledge transfer approach is adopted in machine learning by leveraging a pre-trained source model to improve the training efficiency and performance of a distinct target model.

Like in human learning, transfer learning is effective when the learning tasks of source and target models are related, if not identical. This method has found success in many fields [38], such as computer vision [39–41], natural language processing [42–44], and speech recognition [45], offering benefits like reduced training time, improved performance, and better generalization. Due to the high computational cost to generate XAS spectra using first-principles band-structure codes, constructing large XAS datasets with different levels of theory remains a bottleneck. Effective transfer learning provides a practical route to leverage data from different fidelities and significantly enhance the performance of XAS ML models.

Transfer learning is categorized based on approaches used, level of similarity of source and target tasks as well as that of input (domain) data [46]. In this paper, we focus on two of such transfer learning categories: *inductive-transfer-learning* and *domain-adaptation*, for which we use the *feature-transfer* and *fine-tuning* approach, respectively.

Based on the categorization proposed by Pan and Yang [46], transfer learning is classified according to the approaches used, the level of similarity between source and target tasks, and the similarity of input (domain) data. In this paper, we focus on two of the categories they define: *inductive-transfer-learning* and *domain-adaptation*. For these categories, we employ the *feature-transfer* and *fine-tuning* approaches, respectively.

1. Inductive Transfer Learning via Feature-Transfer

Our methods to investigate the “Feature Transfer Hypothesis” fall under the category of *inductive transfer*

learning [46] in the ML terminology, as the tasks between source and target models, while distinct, share similarities in predicting physical properties of materials based on their graph representations. To harness this potential, we selected a pre-trained interatomic potential model called M3GNet [28] as our transfer source, and hypothesize that it contains valuable, transferable knowledge for XAS prediction.

Among several approaches that can be taken for inductive transfer learning, we implemented *feature-based transfer learning*. This methodology is guided by the principle that representation of input in early layers of neural networks should be similar across models trained for different (but similar) tasks. One important note to this approach is that XAS prediction is focused mostly on local properties compared to M3GNet, which predicts both global properties (total energy and stress) as well as local properties (atomic forces). This intricacy suggests that only a subspace of the full latent representation of M3GNet is needed in knowledge transfer onto XAS prediction. Our design and implementation details for this model adaptation are described in Sec. IID.

2. Domain Adaptation via Fine-tuning

To further investigate the “Universal XAS Model Hypothesis”, we complement our inductive transfer learning approach with domain adaptation. This method is particularly effective when the task remains consistent but the input data distribution shifts—a common scenario when predicting XAS across diverse elements. Our domain adaptation strategy is implemented in two stages: a) the development of a universal XAS model, and b) fine-tuning universal models with element-specific XAS data.

We begin by developing a universal XAS prediction model designed to operate across eight 3d-transition metals. This model is trained on an extensive dataset generated from FEFF simulations, encompassing a wide array of structural and site configurations. The fundamental premise underlying this approach is rooted in the intuition that while XAS spectra exhibit element-specific characteristics, the core physical principles governing X-ray absorption remain sufficiently similar across elements, which lead to general spectral trends beyond a single element.

For example, since 1920 it has been established that the edge position in the XANES spectra strongly correlates with the oxidation state of the absorbing atom [47–50]. The absorbing atom can exhibit a positive (negative) edge shift depending on whether it is oxidized (reduced), which can be understood as the shielding effects of valence electrons on the core-electron energy levels. Another important trend in 3d transition metal K-edge XANES is the relation between the pre-edge features and the cation coordination number and local symmetry [4, 51]. This relation is governed by the dipole-

forbidden $1s \rightarrow 3d$ transition in the pre-edge of the 3d transition metal. The pre-edge intensity vanishes in a perfect octahedron, but becomes more evident, when the inversion symmetry is broken that allows the mixing of p and d orbitals. Overall, the pre-edge feature is pronounced in early 3d transition metals (e.g., Ti) [4], but becomes less obvious in late 3d transition metals [52], as filled 3d orbitals in the latter block the $1s \rightarrow 3d$ transitions. Liang *et al.* constructed a rank-constrained adversarial autoencoder to disentangle individual structure-spectrum relationship in the same material space as this study [53]. In addition to oxidation state and coordination number, the authors identified distinct spectral features associated with a set of less well-known local structure descriptors, such as oxygen coordination number, standard deviation of nearest neighbor (NN) bond angles, and minimum O-O distance on the edges of the NN polyhedron [53]. The distinct spectral features are not limited to edge position and pre-edge intensity but also include more subtle ones, such as main edge position and intensity, post-edge position and intensity and curvature of the pit, based on the terminology of Guda *et al.* [17].

Given the domain knowledge on spectral trends beyond single element, we believe that ML models can learn to capture not just element-specific features, but also essential and transferable knowledge for XANES prediction, creating a robust foundation for element-specific adaptations. Following the training of the universal XAS model, we perform targeted fine-tuning for specific elements of interest. This process allows the model to adapt its learned representations to the unique spectral signatures of individual elements while preserving the broader understanding gained from the diverse training set.

Based on the description above, we introduce the following model names used throughout the rest of the text as shown in Fig. 1 as well. We also collectively refer them as *XASModels*.

- *ExpertXAS*— These models are trained to predict site-specific spectra for a particular element using only the corresponding subset of data from either FEFF or VASP simulations.
- *UniversalXAS*— This model is trained to predict spectra for all eight 3d-transition metals using the complete set of data from FEFF simulations.
- *Tuned-UniversalXAS*— These models are created by fine-tuning the UniversalXAS model using only the corresponding subset of data from either FEFF or VASP simulations.

D. XASModels

In this section, we introduce our model for site-specific XAS prediction. As outlined in Sec. IIC1, we employ a feature-based transfer learning approach, leveraging the

pretrained M3GNet model. We first briefly summarize the core components of M3GNet. Then, we explain how XASModels are constructed by combining two key elements: frozen feature extraction layers from M3GNet and new trainable neural network blocks designed specifically for XAS prediction, which we call *XAS blocks*.

1. M3GNet block

M3GNet is a graph deep learning model designed to predict the inter-atomic potential, forces, and stresses from the material graph representation [28]. It was trained on a comprehensive dataset from the Materials Project, comprising over 17 million entries and encompassing 89 elements from the periodic table. The efficiency M3GNet was demonstrated by the study that predicted almost 2 million stable materials from a total of approximately 32 million candidate structures [28].

Inspired by the significance of three-body interactions in inter-atomic potentials, M3GNet distinguishes itself in its design from other graph deep learning models by embedding three-body interaction within the graph edge features. Specialized “many-body to bond” neural blocks are used to update latent features, incorporating three-body interactions among atoms based on their precise atomic coordinates. These “many-body to bond” blocks are sandwiched between graph convolution layers. A pair of these “many-body to bond” and graph convolution layers is collectively referred to as an *M3GNet block*.

As information flows through M3GNet, it passes through three of these M3GNet blocks before final decoding using the readout block. We use the latent space representation from the *final* M3GNet block to encode structural information for XAS prediction.

2. Transfer-Features

The M3GNet architecture is designed to predict both global and local material properties, with its latent space at each layer containing information relevant to both. In contrast, XASmodels focus solely on local properties. This fundamental difference needs to be taken into account when attempting to leverage M3GNet’s capabilities for XAS prediction through transfer learning. The key lies in identifying an appropriate subset of M3GNet’s latent space that strikes a delicate balance. This subset must be localized, capturing the specific site of interest for XAS prediction to ensure specificity of results. Simultaneously, it needs to be comprehensive enough not to discard potentially valuable information (e.g., intermediate range structural information beyond the first near-neighbor shell) that is known to contribute to XAS features. A significant complexity in this process stems from M3GNet’s “black box” nature – a common issue in machine learning where advanced models have internal mechanisms that are not easily interpretable [54, 55].

This lack of transparency makes it difficult to identify which aspects of the model’s latent space are most crucial for accurate XAS prediction. To address these issues, we adopt the following approach for latent space selection: we extract node features from the graph after final graph convolution layer and just before M3GNet’s readout block, focusing on the site where the absorption element information was initially encoded. Our rationale is that M3GNet’s training on both local and global property prediction may have naturally produced a latent space that can be segmented into clusters containing rich local information surrounding the sites where atomic data was initially encoded, while still preserving some global context relevant to XAS prediction. While intuitively appealing, it is important to note that the effectiveness of this approach needs to be validated, a task we undertake in Secs. III B and IV B.

The M3GNet block can be represented as a function that takes in the material structure’s graph representation \mathcal{S} and generates features in latent space, i.e.

$$\mathbf{H}_0 = f(\mathcal{S}) \in \mathbb{R}^{n \times 64}. \quad (3)$$

Here, $f : \mathcal{S} \rightarrow \mathbb{R}^{n \times 64}$ represents the M3GNet function that transforms an input structure \mathcal{S} containing n atoms into a matrix \mathbf{H}_0 . Each row of \mathbf{H}_0 encapsulates the features around the location where an individual atom i was initially encoded in the graph.

Our focus is the site-specific spectrum, so we extract the feature vector for the i -th absorbing atom as:

$$\mathbf{h}_0^{(i)} \in \mathbb{R}^{64}. \quad (4)$$

This vector, corresponding to the i -th row of \mathbf{H}_0 , serves as the input to the subsequent MLP layers and is henceforth referred to as *transfer-features*.

3. XAS block

The XAS block module harnesses the power of MLPs to transform M3GNet-extracted transfer-features into accurate XAS spectra. At their core, MLPs are composed of interconnected neuron layers that sequentially process and refine input data. As transfer-features flow through the XAS block, each layer applies complex nonlinear transformations, enabling the network to capture intricate relationships between transfer-features and resulting XAS spectra. The universal approximation theorem of MLPs, a fundamental principle in machine learning, mathematically establishes that a properly constructed XAS block can arbitrarily closely approximate any continuous function mapping transfer-features to XAS spectra, provided sufficient network depth and optimal parameter configuration.

Each layer of the MLP applies a carefully chosen sequence of operations: linear transformation, batch normalization, activation, and dropout. This sequence maintains the network’s generalization capabilities and

enhances learning. For hidden layers ($l = 1, 2, \dots, L - 1$), the computation is defined as:

$$\mathbf{h}_l = \text{Dropout}_l(\text{SiLU}(\text{BatchNorm}_l(\text{Linear}_l(\mathbf{h}_{l-1})))) \quad \forall l \in \{1, 2, \dots, L - 1\}. \quad (5)$$

This formulation encapsulates the layer-wise transformation of features, with each operation serving a specific purpose in the learning process. The final output layer differs slightly to others by using Softplus,

$$\hat{\boldsymbol{\mu}} = \mathbf{h}_L = \text{Softplus}(\text{Linear}_L(\mathbf{h}_{L-1})). \quad (6)$$

Here, $\hat{\boldsymbol{\mu}} \in \mathbb{R}^{141}$ represents the predicted XAS spectrum across 141 energy grid points.

Each component of the MLP serves a specific purpose in the learning process, contributing to the model’s overall effectiveness:

- Linear layers form the backbone of the MLP, enabling the network to learn complex mappings between input and output spaces:

$$\text{Linear}_l(\mathbf{h}) = W_l \mathbf{h} + \mathbf{b}_l. \quad (7)$$

Here, the weight matrix, W_l , and bias vector, \mathbf{b}_l , are learnable parameters that the network adjusts during training to capture the underlying patterns in the data.

- Batch normalization plays a crucial role in stabilizing the training process. By standardizing the inputs to each layer, it helps mitigate the internal covariate shift problem, allowing for faster and more stable training [56].
- The SiLU activation function introduces non-linearity into the network, which is essential for capturing complex relationships. SiLU has been shown to outperform ReLU in many scenarios [57]:

$$\text{SiLU}(\mathbf{x}) = \mathbf{x} \odot \sigma(\mathbf{x}), \quad (8)$$

where $\sigma(\mathbf{x}) = (1 + e^{-\mathbf{x}})^{-1}$ is the element-wise sigmoid function and \odot is the element-wise product. The smooth nature of SiLU aids in the propagation of gradient during training.

- Dropout serves as a powerful regularization step, preventing overfitting by randomly deactivating a fraction of nodes during training. This encourages the network to learn more robust features and reduces its reliance on any single node [58]. We set dropout rate of 0.5 for all Dropout layers in XAS-block.
- The Softplus activation in the final layer ensures that the output respects the physical constraints of XAS spectra:

$$\text{Softplus}(\mathbf{x}) = \log(1 + e^{\mathbf{x}}), \quad (9)$$

where all operations are applied element-wise. This function guarantees positive outputs, which is crucial as XAS intensities cannot be negative.

The flexibility of the MLP allows for extensive optimization of both architectural and training parameters. Key architectural variables include network depth (number of layers) and width (neurons per layer), which directly impact the model’s capacity to learn complex patterns. Crucial training hyperparameters, including batch size, influence optimization stochasticity and generalization ability.

To efficiently navigate this vast hyperparameter space, we employ Bayesian optimization technique implemented in Optuna Package [59]. This approach systematically explores potential configurations, using validation set performance to guide the search for optimal combinations of architectural and training parameters. The exploration of the search space and the resulting optimized configurations for all XASModel instances are detailed in Supplementary Table S3-S4.

E. XASModel Training

In this section, we describe our methodology for training XASModels. We introduce our data partitioning scheme, which gives more consistent split sizes across data sets of different sizes and material distributions. This also ensures that there is no material information leakage between them. We then detail the training steps taken to ensure robust model performance.

1. Balanced Material Splitting

For training, we first divide the data into training, validation, and test sets, used respectively for model fitting, hyperparameter tuning, and final performance evaluation. The goal of our XAS prediction models is to be generalizable to unseen *materials*, not just unseen *absorbing sites*. To achieve this, we employ materials splitting, where all XAS spectra from a single material’s absorption sites are grouped together, ensuring they always fall into the same split. This ensures no data from any single material appears across multiple splits (e.g., a material with two atoms, in which one atom’s data is in the training set and the other atom in the testing set), preventing data leakage and overly optimistic performance estimates. We call this approach *materials splitting*, which provides a more stringent assessment of a model’s generalizability to new compounds.

Materials splitting raises an issue in maintaining consistent proportions across training, validation, and testing sets. While we target specific ratios (e.g., 80/10/10), the uneven distribution of XAS data across materials complicates achieving these exact proportions. Some materials in our dataset have more unique site-XAS spec-

tra than others, creating an inherent imbalance. Consequently, when splitting by material, the actual data proportions in each set may slightly deviate from the intended ratios. This discrepancy is especially noticeable in datasets with a wide variance in the number of site-XAS spectra per material or in smaller datasets. Such imbalances can potentially impact the consistency of model training and evaluation, necessitating careful consideration during the data preparation phase.

We designed a heuristic algorithm to address the issue of attaining desired proportions in the splits while maintaining materials splitting. The algorithm takes a greedy approach when assigning unique materials to groups. It begins by sorting the unique materials by their number of unique sites in the descending order, then iteratively assigns them to groups. This sorting scheme ensures that materials with the largest number of sites are handled first, allowing for more effective balancing of the splits later on. For each unique material, the algorithm calculates how adding it would affect each group’s proximity to its target size. It then assigns the material to the group where this addition would most effectively reduce the gap between current and target sizes. This process continues until all the materials are allocated.

2. Training

We trained the XASModels with the Adam optimizer [60] using the mean squared error (MSE) loss function,

$$J \equiv \frac{1}{N} \sum_{i=1}^N J^{(i)}, \quad (10)$$

where

$$J^{(i)} = \frac{1}{M} \sum_{j=1}^M (\mu_j^{(i)} - \hat{\mu}_j^{(i)})^2 \quad (11)$$

and N is the number of spectra in the training set. $\mu_j^{(i)}$ and $\hat{\mu}_j^{(i)}$ represent the ground truth and the predicted absorption coefficient, respectively, at the j -th energy grid point of the i -th target spectrum.

Before the start of each training run, we used an automated search [61] to determine a good initial learning rate. To prevent overfitting of models, we terminated training whenever the minimum validation loss did not improve for 50 consecutive epochs. The maximum number of epochs was set to 1000, which was verified to be sufficient to allow for convergence in training of all model variants across all element datasets.

When generating Tuned-UniversalXAS models, we face two distinct fine-tuning scenarios. The first involves fine-tuning the UniversalXAS model with element-specific FEFF datasets, where the training data is a subset of what was used to train the original UniversalXAS model. The second involves fine-tuning with

VASP datasets, which introduces both a different fidelity and variance, demanding more substantial model adaptation. Given these different adaptation requirements, we systematically evaluated dropout regularization using two configurations: maintaining the original dropout rate of 0.5 (consistent with ExpertXAS and UniversalXAS models) and disabling the regularization by setting it to 0. This approach allows us to determine the optimal approach for each type of Tuned-UniversalXAS model.

III. PRELIMINARY ANALYSIS AND DATA EXPLORATION

Data exploration serves as an important indicator of the overall data quality of the dataset and helps to uncover potential connections between features and targets. This is a necessary step before training ML models.

The ML-data features and targets frequently exist in high-dimensional spaces, rendering data exploration challenging. Our work, for instance, employs 64-dimensional feature-transfer vectors and 141-dimensional target vectors for each data point. The number of spectra in dataset for each element range from 3,067 (Cr) to 17,052 (Mn). To address the challenge of high data dimension, we use two approaches for preliminary data analysis and exploration: a) direct visualization of the spectral data using heatmaps and b) dimensionality reduction-assisted analysis of both feature and target data.

We employ Uniform Manifold Approximation and Projection (UMAP) to visualize our feature and target data [62]. UMAP is a powerful dimensionality reduction technique that effectively preserves the topology of both local and global data structures. It first constructs a high-dimensional representation using a fuzzy topological framework, specifically a weighted k -neighbor graph where edge weights represent the probabilities of connections between points. Then a low-dimensional representation is constructed with the same method, which is optimized by minimizing the cross-entropy between the high- and low-dimensional fuzzy representations through a stochastic gradient method. Due to its stochastic nature, UMAP projections are inherently non-deterministic and can vary based on parameters like the number of neighbors and minimum distance. While these variations can lead to different details in clustering outcomes, even a single projection that reveals clear clustering provides strong evidence for the existence of clusters in the original high-dimensional space. UMAP often uncovers intricate patterns and cluster structures that may be obscured in the original data [62].

A. Spectra Visual Inspection

We proceed by direct visual inspection of the spectra database with focus first on the high-density region

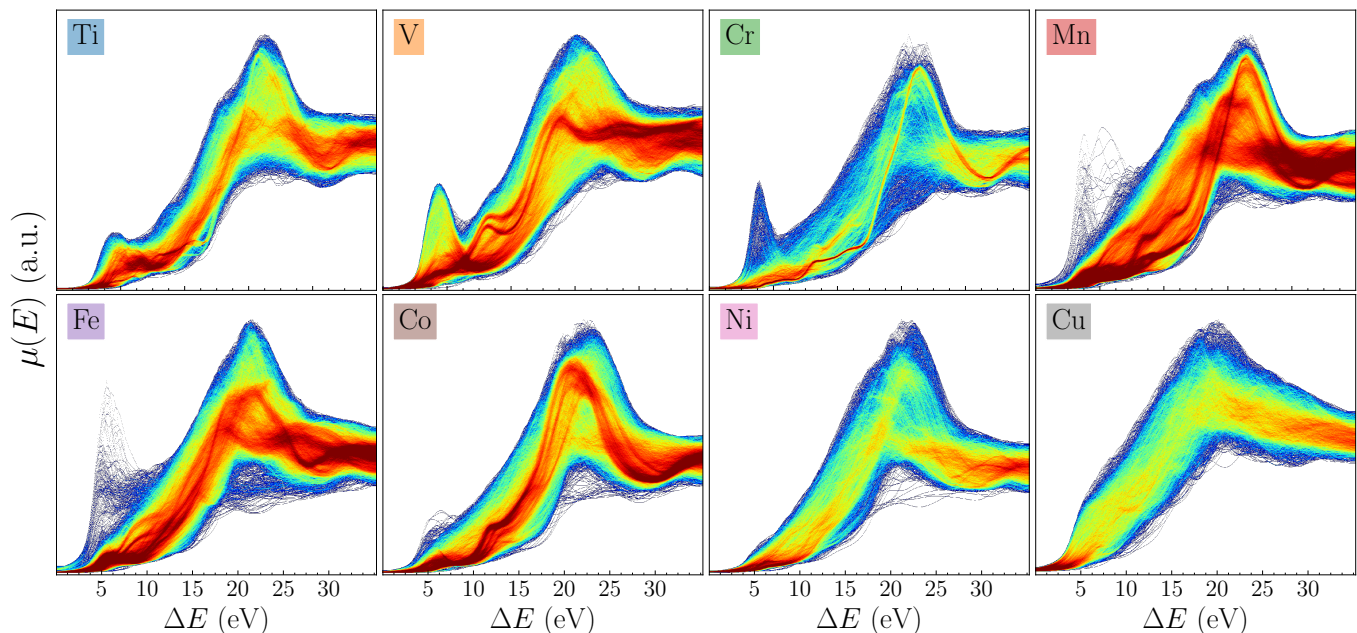


FIG. 3. Heatmaps of FEFF XANES spectra. Color represents the density of spectral features, with red indicating a higher concentration of spectra and blue indicating a lower concentration (and white indicating no data).

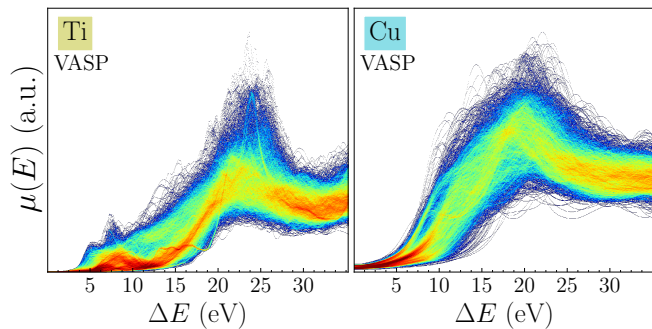


FIG. 4. Heatmaps of VASP XANES spectra. Color scheme is the same as in Fig. 3.

(dark red) of the entire FEFF database in Fig. 3. Ti and V XANES exhibit clearly visible pre-edge peaks. The pre-edge peak intensity decays when moving towards late 3d transition metals, consistent with previous observations [52]. Mn and Fe XANES show a much wider range in edge location between 10 to 20 eV than others, consistent with multiple commonly observed oxidation states. Mn XANES shows at least three edge positions as families of dark red curves and Fe XANES shows at least two with smaller spacing than Mn. Correspondingly, Mn and Fe spectra exhibit a broader spectral variation or data complexity than others. In contrast, Ni and Cu XANES show a relatively simple spectral shape with one dominant peak. Regarding Ti and Cu VASP database in Fig. 4, overall the spectral shape in VASP are similar to that in FEFF. However, the low density region (dark blue) in VASP has larger variations than FEFF in both

elements.

B. Transfer-Features Clustering

We select the node features of the absorbing site within the M3GNet latent space for transfer-feature extraction, as detailed in Sec. IID 2. While this choice is intuitively appealing, it necessitates careful examination due to the *feature smoothing* phenomenon. Feature smoothing occurs when node features gradually become similar with each other in deeper layers of the model [63, 64]. This happens because the deeper convolutional layers increasingly expand the effective receptive field from which the features are being aggregated. In the case of M3GNet, feature-smoothing could potentially dilute local structural information when it passes through the M3GNet block, thereby hindering the site-specificity of predicted XAS and model performance.

We provide empirical evidence that feature dilution in the M3GNet block is sufficiently mild. We demonstrate this by visualizing the transfer-feature in low dimensions with UMAP and identifying clusters that correspond to local physical descriptors. Specifically we use the oxidation state (OS) of the absorbing cation atom, the number of nearest neighbors of the absorbing cation atom, i.e. the coordination number (CN), and the average number of nearest neighbors of oxygen atoms in the first shell of the absorbing cation, i.e. the oxygen coordination num-

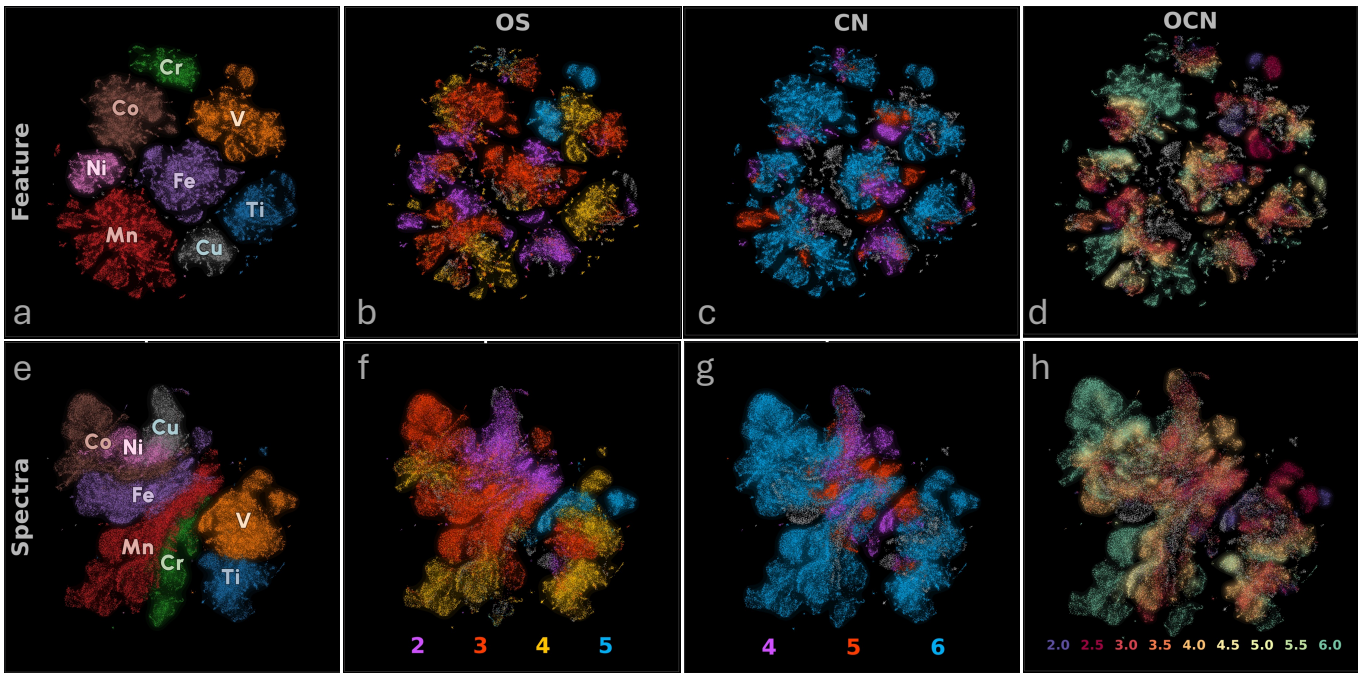


FIG. 5. UMAP plots of M3GNet featurization (a-d) and spectra (e-h) colored by absorbing element type (a and e), oxidation-state (OS) (b and f), coordination-number (CN) (c and g), and oxygen-coordination-number (OCN) (d and h). The oxygen-coordination-number is rounded to the nearest half integer for visualization purposes. The gray points represent those in which the physical descriptor can not be determined.

ber (OCN)¹ as the labels.

We plot two-dimensional UMAP of all transfer features in the FEFF ML dataset in Fig. 5a-d, colored by element type and local physical descriptors (OS, CN, and OCN). In doing so, we make two key observations as summarized below that provide compelling evidence that the latent-subspace-based transfer features can be used for XAS prediction without concern for potential feature smoothing inside the M3GNet block.

- *Element Clustering*— The transfer-feature UMAP in Fig. 5a shows distinct clusters based on the absorbing element type. The figure clearly demonstrates that these clusters map very well to the absorbing element labels from which they originally assigned (marked by different colors). This provides strong evidence that transfer features are sufficiently distinct among different absorbing elements, despite some moderate feature smoothing.
- *Local Descriptor Clustering*— The UMAP plots, when colored by local physical descriptors as shown in Fig. 5b-d, reveal consistent sub-groupings within element clusters. This observation is consistent

¹For example, consider an absorbing Ti atom with 2 nearest neighbor Zn and two nearest neighbor O atoms. If one of the O atoms has 2 nearest neighbors, and the other O atom has 4 nearest neighbors, then the OCN value for the absorbing Ti atom is 3.

across different physical descriptors, indicating that transfer features are sufficiently rich to capture local structure information, both at the level of the absorber element type and the sub-level of physical descriptors (OS, CN and OCN) within each element.

C. Spectra Clustering

Now we turn to the UMAP clustering of the FEFF spectra as shown in Fig. 5e-h. In prior studies [22, 23, 52], spectral analysis through dimensionality reduction, such as principal component analysis (PCA), is typically performed on datasets of a single absorbing element. This is because XANES is an element-specific technique, and one can trivially distinguish the K-edge XANES of 3d transition metals from their absolute edge position energy. In this study, we apply UMAP to the entire FEFF database (Ti – Cu) defined on a relative energy scale, thereby removing the information regarding absolute edge positions. This approach allows the UMAP analysis to focus exclusively on spectral shapes. By visualizing the *entire* FEFF dataset in a single UMAP plot, we can identify relationships and trends that transcend individual elements, highlighting how spectral features correlate with broader material properties.

As shown in Fig. 5e, distinct clustering patterns form for each element. In addition, elements next to each other

in the periodic table stay close together: early 3d transition metals (Ti, V, Cr and Mn) and late 3d transition metals (Fe, Co, Ni and Cu). However, there is not a clear separation between early and late 3d transition metals. The spectral similarity between different elements is supported by the observation in the raw spectra in Fig. 3, e.g., the pronounced pre-edge in Ti, V and Cr, and the single peak feature in Ni and Cu.

Further spectra clustering analysis was performed by labeling the UMAP pattern with physical descriptors (OS, CN and OCN) as shown in Fig. 5f-h. Within each element, one can see sub-level clustering with respect to the physical descriptors. For example, there is a dominant Ti^{4+} cluster (yellow) accompanied by a much smaller Ti^{2+} cluster (purple), while the Ti^{3+} cluster (red) is less significant and more scattered. This is consistent with the fact that the Ti database is dominated by Ti^{4+} cations. Furthermore, there is a clear distinction in Mn and Fe clusters with respect to OS (2+, 3+ and 4+), resonating the strong trend in edge positions in the raw spectra in Fig. 3, because edge position is a well established proxy of ion OS. Similarly, sub-regions of CN can be seen within each cluster, e.g. 4-, 5- and 6-coordinated Ti and V with 6-coordinated motifs being most prevalent, consistent with previous work based on PCA analysis [52]. OCN is a more subtle physical descriptor, which is correlated to main peak intensity [53]. However, such trend is not obvious in the raw spectra, as it is entangled with spectral variations caused by other physical descriptors [53]. It is quite remarkable that UMAP can in fact form distinct clusters of OCN, although its pattern is more complex than OS and CN.

As shown by the above analysis, 3d transition metal K-edge XANES encodes key information of the local chemical environment (e.g., OS, CN and OCN). To train an ExpertXAS model, the structure features are required to capture the variation of these physical descriptors. The success in M3GNet feature transfer in identifying such variations in Fig. 5 is a necessary condition of the validity of the feature transfer hypothesis, establishing the effectiveness of M3GNet feature transfer for predicting XANES spectra.

D. Evidence of Universality

The UMAP analysis in Fig. 5 also provides compelling evidence that XANES prediction can take advantage of the spectral features common to all absorbing elements (Ti – Cu), setting the stage for the UniversalXAS model.

When we color the UMAP pattern of the entire FEFF dataset based on the absorbing element type, distinct clusters are formed corresponding to each element, in both the featurized structure space and spectral space. Nevertheless, closer inspection reveals that certain subset of spectra of one element can be more similar to spectra in different element clusters than to those within the same element. For example, spectra at the top right of the

elongated Mn cluster in Fig. 5e are closer to the neighboring spectra in the Fe and Cr clusters than those at opposite ends of the Mn cluster. Similar observations can be made for other element clusters as well. From domain knowledge, we know that fingerprints of OS (e.g., edge position) and CN (e.g., pre-edge peak position and intensity) could serve as good candidates of the across-element trends. In addition, the analysis of the rank constrained adversarial autoencoder [53] also made the strong case for across-element spectral trends associated with additional physical descriptors, e.g., OCN, following a spectral disentanglement procedure.

To further investigate the cross-element spectral trends, we revisit the spectral UMAP analysis labeled by local physical descriptors (OS, CCN, and OCN) in Fig. 5f-h. We can see distinct sub-regions of physical descriptors (distinguishable by different colors) within each element cluster often coalesce into larger clusters across the boundary of elements. This coalescence is particularly pronounced in the OS UMAPs (e.g., OS=2: Cu^{2+} , Ni^{2+} and Fe^{2+} in pink color; OS=4: Ti^{4+} and V^{4+} in yellow color), underscoring the strong similarity in the spectral feature of different 3d transition metals at the same OS. Similar observations can also be made for CN and OCN, except that the across-element coalescence is not as concentrated as OS. These qualitative visualizations indicate that local chemical environments, in addition to the absorbing element type, are among the primary characteristics that influence XANES prediction and their trends can cross element boundaries.

IV. RESULTS

A. Performance Metrics

When evaluating XASModels, it is imperative to fairly evaluate and compare the performance of models trained on different subsets of the spectra dataset of varying scales associated with the level of theory (the multiple scattering method in FEFF and the core-hole pseudopotential method in VASP) and element type (absolute cross-section of eight 3d transition metals). To enable this comparison, we introduce a relative performance metric, η , with intention to provide a single, interpretable metric that is robust to outliers and allows for meaningful comparisons.

Our primary interest lies in spectral prediction accuracy, which we quantify using the MSE [c.f. Eq. 11] for each individual spectrum predicted by a model. To represent a model’s overall performance for a set of spectra, we introduce the Median Spectral Error, denoted as ξ ,

$$\xi = \text{Median}(\{J^{(i)}\}_{i=1}^N), \quad (12)$$

where $J^{(i)}$ is the MSE of the i -th spectrum and N is the number of spectra in the test set. We choose the median of error, instead of other representative statistics

like the mean, because different models have different error distributions which are often skewed and contain outliers. The median provides robustness against these factors.

To address the issue of varying spectral scales in different spectra sub-datasets, we normalize ξ to bring it to a similar scale, yielding the relative performance metric η ,

$$\eta = \frac{\xi_{\text{baseline}}}{\xi}. \quad (13)$$

Here, ξ_{baseline} is derived from a simple baseline model that always predicts the mean spectrum of the training set,

$$[\mu_{\text{baseline}}]_j = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \mu_j^{(i)}. \quad (14)$$

The η metric provides an intuitive measure of model performance, enables meaningful comparisons across sub-datasets with different spectral scales, and preserves the robustness to outliers inherent in ξ .

By definition, $\eta_{\text{baseline}} = 1$. The η metric is unbounded above 1, with higher values indicating better performance. While η effectively captures the overall performance, it does not provide a detailed description of the model performance. Therefore, we further complement this metric with examination of error distributions and visual inspection of predicted spectra, ensuring a comprehensive evaluation of our models.

B. Transfer-Features

Sec. III B presents qualitative evidence supporting our feature transfer hypothesis, which posits that transfer-features are effective for XAS prediction. To further examine this hypothesis, we now provide quantitative results. Additionally, we compare our featurization approach to commonly used methods in the field, specifically the Atom-Centered Symmetry Function (ACSF) [27] and Smooth Overlap of Atomic Positions (SOAP) [18].

ACSF and SOAP are established methodologies for generating consistent, symmetry-invariant representations of local atomic environments [65, 66]. ACSF uses tailored radial and angular functions, while SOAP expands local atomic density using spherical harmonics. Both methods generate descriptors that preserve essential local chemical information and ensure rotational and translational invariance. This allows ML models to learn structure-property relationships without rediscovering fundamental principles (e.g., symmetry), making them suitable for predicting properties governed by local atomic environments, such as XAS spectra [23].

Even though both ACSF and SOAP aim to provide descriptors of fixed length, the resulting representations

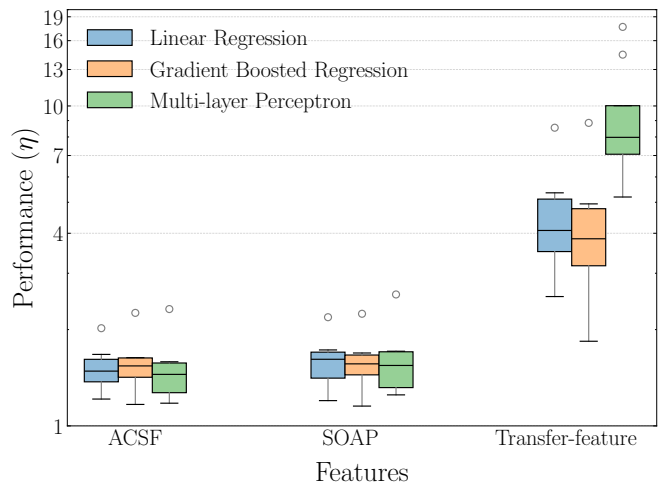


FIG. 6. Performance comparison of ML models using different feature sets for XAS prediction. The box plots show the performance (η) of three ML models of varied complexity—Linear Regression, Gradient Boosted Regression, and Multi-layer Perceptron—trained on three types of features: ACSF, SOAP, and transfer-features derived from M3GNet’s latent subspace (as used in XASModels). The results of the multi-layer perceptron correspond to the ExpertXAS model.

can get very large and redundant. To overcome this, we applied PCA to reduce the dimension of the feature vector while capturing 99% of the explained variance. This process significantly reduced the high-dimensional ACSF and SOAP representations (often several thousand dimensions) to below 100 dimensions.

Here we compare the effectiveness of three different feature sets: ACSF, SOAP, and M3GNet transfer-features. To avoid potential bias from a particular choice of the model, we trained three classes of ML models of varied complexity: a Linear Regression model, Gradient Boosted Regressor, and Multi-layer Perceptron. For the MLP-based models, we employed the same XAS block architecture, as described in Sec. IID 3, thereby making the MLP model trained on transfer-features correspond to the ExpertXAS model.

The results in Fig. 6 show that ML models trained on transfer-features to predict XAS consistently yield η values much greater than one, across model complexities thereby supporting our hypothesis that transfer-features are effective for XAS prediction. Furthermore, the performance of the transfer-feature-based models is significantly better than those trained on ACSF and SOAP across all model complexities. We will further discuss the implications of these results in Sec. V A.

C. Performance Overview

We now present quantitative results to showcase the general trends in performance of 21 instances of XAS-Models. We employ the η metric, visualized in Fig. 7, to

evaluate model efficacy. Table I delineates specific performance metrics and offers revealing comparative metrics of ExpertXAS and Tuned-UniversalXAS models. Our investigation expands upon earlier qualitative findings regarding transfer learning and the potential universality of XAS prediction models, as discussed in Secs. III B and III D.

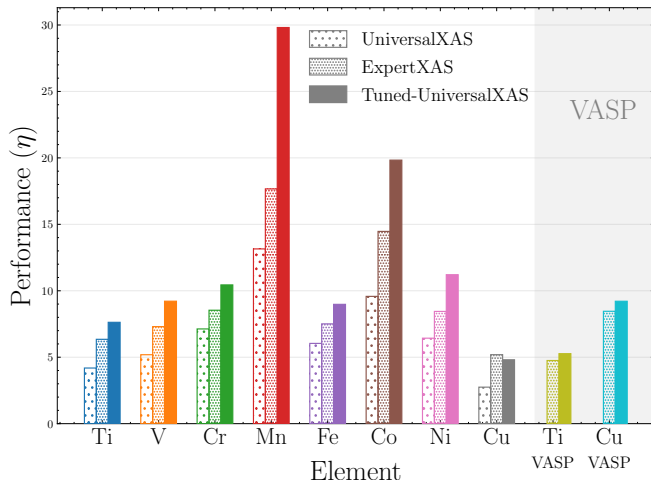


FIG. 7. XASModels performances across different elements. The bar chart shows the performance (η) of XASModels: 10 ExpertXAS models, 1 UniversalXAS model, and 10 Tuned-UniversalXAS models, evaluated across various absorbing elements using both FEFF and VASP spectra. The VASP models are highlighted with shaded regions. The performance of the Universal model is partitioned for each element, as described in Sec. IV C 2, enabling direct comparison with other models. See Table I for performance metrics.

1. ExpertXAS Performance

ExpertXAS models consistently outperform the baseline across all elements studied, demonstrating their effectiveness in predicting XAS spectra. As shown in Table I and Fig. 7, the model’s predictive power varies widely by element, with η ranging from 4.75 to 17.66, indicating the accuracy’s strong dependence on element type.

For FEFF spectra, Mn and Co show the best performance, with highest η reaching up to 17.66, while Ti and Cu exhibit lower performance, with lowest η at 5.19. Notably, even for Ti and Cu, the performance is more than five times of the baseline model. The performance trends differ for VASP spectra. Similar to FEFF models, Ti has an η value of 4.75, being the lowest among all models. However, the performance of Cu VASP is much higher than its FEFF counterpart, reaching the η value of 8.46.

2. UniversalXAS Performance

To facilitate the direct comparison of the UniversalXAS model with other models, we performed an element-wise breakdown of the UniversalXAS model performance. We first partitioned the test set into element-specific subsets and calculated the median MSE for each element separately. These element-specific values were then normalized by the same ξ_{baseline} used for other model variants, derived from the mean spectrum of the entire training set. This approach yields element-specific η values for the UniversalXAS model, enabling direct comparison with other models despite the UniversalXAS model being trained on multi-element data with varying spectral scales.

The UniversalXAS model demonstrates η values ranging from 2.75 to 13.15 across different elements as shown in Fig. 7 and Table I. While these values are generally lower than those of the corresponding ExpertXAS models, they consistently exceed the baseline performance for all elements. Notably, the element-wise performance trend of the UniversalXAS model closely mirrors that of the ExpertXAS models, maintaining a similar ranking order but with lower magnitudes.

The MSE distribution of ExpertXAS models shown in Fig. 8 demonstrates consistent patterns across different elements when examined on a logarithmic scale ($\log_{10}(\text{MSE})$). These errors manifest as bell-shaped distributions centered between -9 and -7, where smaller errors are common and larger MSE become exponentially rare. While spanning approximately three orders of magnitude (-10 to -7), these distributions consistently show lower errors than their respective baselines. The similarity in distribution shapes and their consistent positioning relative to the baselines across the transition metal series suggests robust model behavior.

3. Tuned-UniversalXAS Performance

Tuned-UniversalXAS models outperform both ExpertXAS and UniversalXAS models across all but one case (Cu FEFF), with varying degrees of improvement. The smallest improvement is observed for Cu VASP, with a 8.92% increase in η , while the largest improvement is seen for Mn FEFF, with a substantial 68.77% increase. For the exceptional case of Cu FEFF, the Tuned-UniversalXAS model performs slightly worse than the ExpertXAS model, with a decrease of 7.42% in η .

D. Example Predictions

We now provide specific examples to demonstrate the predictive power of XASModels. Fig. 9 shows spectrum predictions for Mn, Fe, and Cu, corresponding to the highest, median, and lowest performing models, respectively, within the Tuned-UniversalXAS variant. Predic-

TABLE I. Performance metric of ExpertXAS (η_E), UniversalXAS (η_U), and Tuned-UniversalXAS (η_T) models. The best performance for each element is highlighted in bold. The improvement column shows the percentage increase in performance of the Tuned-UniversalXAS model over the ExpertXAS model. The Win Rate shows improvement across spectra defined by Eq. 17 and energy grid defined by Eq. 18.

Element	η			Improvement (%) $(\eta_T - \eta_E)/\eta_E$	Win Rate (%)	
	ExpertXAS η_E	UniversalXAS η_U	Tuned-UniversalXAS η_T		Energy w_e	Spectra w_s
Ti	6.35	4.19	7.63	20.15	94.33	65.21
V	7.30	5.19	9.22	26.17	100.00	63.98
Cr	8.54	7.13	10.44	22.32	97.87	69.18
Mn	17.66	13.15	29.81	68.77	100.00	71.19
Fe	7.51	6.04	8.98	19.56	97.87	62.66
Co	14.47	9.58	19.83	37.07	100.00	67.78
Ni	8.45	6.43	11.21	32.71	99.29	65.74
Cu	5.19	2.75	4.81	-7.42	75.89	53.37
Ti VASP	4.75	N/A	5.27	10.82	74.47	63.13
Cu VASP	8.46	N/A	9.21	8.92	95.74	66.04

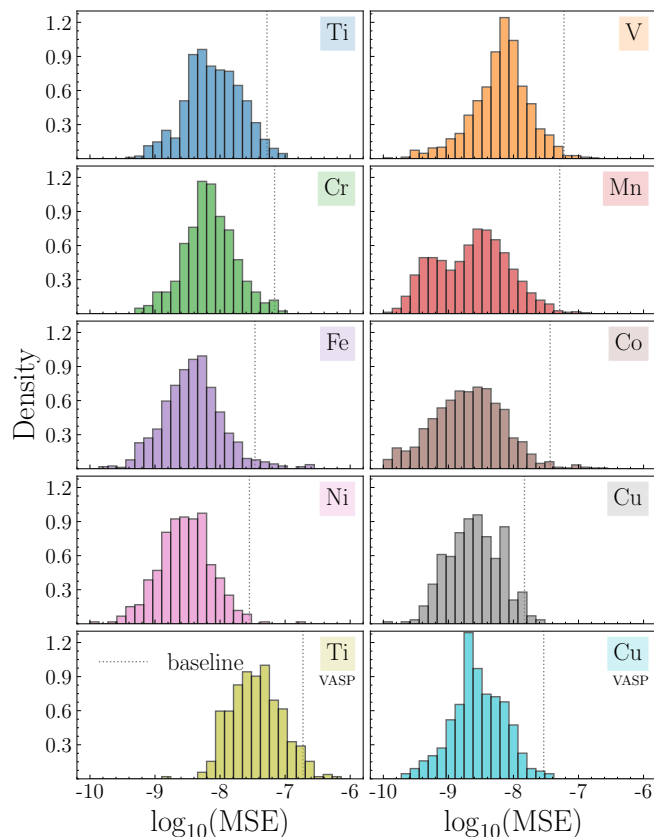


FIG. 8. The density distribution of spectral prediction errors (MSE) for ExpertXAS models across different elements. The gray dashed line represents the baseline MSE (ξ_{baseline}) defined in Eq 13.

tions are categorized into deciles based on MSE per spectrum, with D_i representing the i -th percentile of the MSE distribution. We also provide comprehensive view of example predictions from nine deciles for each of the ten

element/simulation pairs, totaling 252 spectra in Supplementary Figs. S1-S3.

As shown in Fig. 9, in the best performing XASModel of Mn, the predicted spectra closely align with the simulated spectra across all deciles. The shaded areas, representing differences between predicted and simulated spectra, are minimal even in the lower deciles. For the representative median performance of Fe, predictions show strong agreement in the higher deciles (D_{10} - D_{50}), with slightly larger deviations in the lower deciles, particularly in the main-edge and post-edge regions. For the case of Cu, which represents the model with lowest performance, predictions still accurately capture the main edge and overall spectral shape in the higher deciles, with more noticeable differences in the fine details of the lower deciles.

Overall key spectral features are consistently reproduced across all three representative models. Even in the lowest deciles for the lowest-performing models, the essential characteristics of the XAS spectra (e.g., the overall spectral shape) are captured, demonstrating the robustness of our models across a wide range of performance levels.

E. Knowledge Transfer

Comparison of models based on the performance metric η in Sec. IV C indicates the superiority of Tuned-UniversalXAS models over ExpertXAS models. In this section, we dissect the general trend of performance improvement and delve deeper into the nature of these improvements, analyzing how and where they occur. This analysis is motivated to gain insights into the extent and impact of knowledge transfer from the UniversalXAS model to the Tuned-UniversalXAS model.

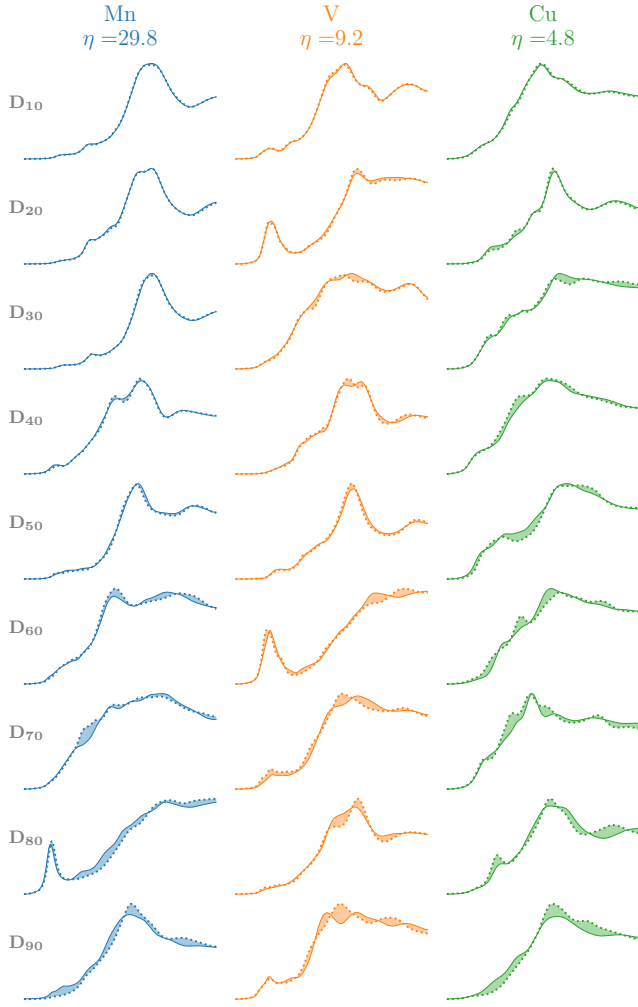


FIG. 9. Spectra from deciles D_i sorted by the MSE-per-spectra of the best model's (Tuned-Universal) prediction. Solid lines represent simulated spectra, dashed lines represent predicted spectra (ground truth), and shaded areas highlight the differences. Performance η for Mn, Fe and Cu corresponds to the maximum, median, and minimum η_i across elements.

1. Improvement Distribution

To evaluate the performance difference between Tuned-UniversalXAS and ExpertXAS models, we compare their predictions on a common set of spectra. Our analysis focuses on the residual differences at each prediction point, which vary with respect to both the individual spectrum and energy. We define our notation as follows:

$$\begin{aligned} \mu_j^{(i)} &: \text{XAS ground truth (simulation),} \\ \hat{\mu}_j^{(i,\text{expert})} &: \text{ExpertXAS prediction,} \\ \hat{\mu}_j^{(i,\text{tuned})} &: \text{Tuned-UniversalXAS prediction,} \end{aligned}$$

where i indexes the spectrum in the test set and j represents the energy grid points within each spectrum. We in-

roduce the following metric of residual difference ($\Delta_j^{(i)}$) to quantify the relative performance of the two models:

$$\Delta_j^{(i)} \equiv |\hat{\mu}_j^{(i,\text{expert})} - \mu_j^{(i)}| - |\hat{\mu}_j^{(i,\text{tuned})} - \mu_j^{(i)}|. \quad (15)$$

This metric provides a point-by-point comparison of model performance. Cases where $\Delta_j^{(i)} > 0$ indicate superior performance by the Tuned-UniversalXAS model, while $\Delta_j^{(i)} < 0$ favors the ExpertXAS model. Fig. 10 illustrates the distribution of all residual differences partitioned by its sign. We observe dominant instances of performance improvements (darker bars) compared to performance degradations (lighter bars) for most elements. Notably, for elements like Mn and Cu, where the difference in η values is small, we see a more balanced distribution between improvements and degradations. These observations aligns with earlier observations in Sec. IV C using the η metric.

2. Win Rates

To further quantify the performance differences between the Tuned-UniversalXAS and ExpertXAS models, we introduce two complementary “win rate” metrics. These metrics are based on the residual difference, $\Delta_j^{(i)}$, defined in Eq. 15. Both metrics employ the standard indicator function, denoted as $\mathbb{1}(x)$, which is defined as

$$\mathbb{1}(x) = \begin{cases} 1 & \text{if } x \text{ is true,} \\ 0 & \text{if } x \text{ is false.} \end{cases} \quad (16)$$

Our first metric, the *win rate across spectra* (w_s), provides a assessment of performance differences among model in terms of spectra predictions,

$$w_s = \frac{1}{N} \sum_{i=1}^N \mathbb{1} \left(\sum_{j=1}^M \Delta_j^{(i)} > 0 \right). \quad (17)$$

This metric calculates the proportion of spectra where the Tuned-UniversalXAS model outperforms the ExpertXAS model when considering the entire energy range.

To complement the broader view presented by w_s , we introduce the *win rate across energy* (w_e), which offers a more granular view based on energy grid,

$$w_e = \frac{1}{M} \sum_{j=1}^M \mathbb{1} \left(\frac{1}{N} \sum_{i=1}^N \Delta_j^{(i)} > 0 \right). \quad (18)$$

This metric represents the fraction of energy grids where the Tuned-UniversalXAS model outperforms the ExpertXAS model, when averaged over all spectra.

Both w_s and w_e provide valuable insights into model performance. A w_s value exceeding 0.5 indicates that the

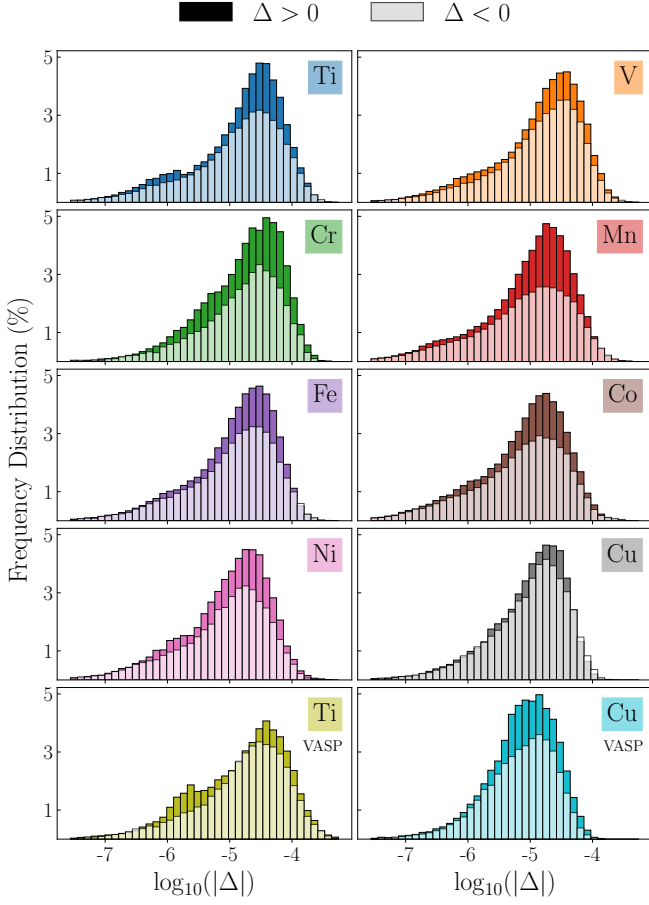


FIG. 10. Distribution of residual differences $\Delta = \{\Delta_j^{(i)} \mid i = 1, 2, \dots, N; j = 1, 2, \dots, M\}$ between Tuned-UniversalXAS and ExpertXAS predictions (see Eq. 15). For easy comparison between improvements and degradations, the data is partitioned by sign of $\Delta_j^{(i)}$ and plotted against its absolute value in logarithmic scale. Darker bars ($\Delta_j^{(i)} > 0$) indicate predictions where Tuned-UniversalXAS outperformed ExpertXAS, while lighter bars ($\Delta_j^{(i)} < 0$) indicate the opposite. Bar heights represent the percentage of the data points, with the total number of data points given by the product of the number of energy grids and the spectra count.

Tuned-UniversalXAS model performs better on the majority of spectra. Similarly, a w_e value above 0.5 suggests superior performance of the Tuned-UniversalXAS model at the majority of energy grid.

As shown in Table I, all w_s are larger than 0.5 ranging from 53.37% in Cu FEFF to 71.19% in Mn FEFF, suggesting a consistent spectrum-wise performance improvement of the Tuned-UniversalXAS model over the ExpertXAS model. The overall trend in w_s is in line with $(\eta_T - \eta_E)/\eta_E$. It is worth noting that in Cu FEFF, although $(\eta_T - \eta_E)/\eta_E$ yields a negative value of -7.42 based on the median MSE, $w_s = 53.37\%$ indicates a slight improvement based on more fine-grained spectrum-wise metric. The advantage of the Tuned-UniversalXAS model is further supported by w_e , which shows a more

significant performance improvement in the energy-wise metric with values varying from 74.74% in Ti VASP to 100% in V, Mn and Co FEFF. Similar to w_s , $w_e = 75.89\%$ in Cu FEFF, suggesting a significant energy-wise performance improvement of the Tuned-UniversalXAS model over the ExpertXAS model.

3. Energy-resolved Improvements

To further quantify the energy-wise performance metric associated with w_e , we compare the residual differences between prediction of ExpertXAS and Tuned-UniversalXAS models averaged over all spectra to examine where in the energy region those improvement occurs. For this purpose, we define an energy-resolved improvement metric $I(\Delta E_j)$ as follows:

$$I(\Delta E_j) \equiv \frac{1}{N} \sum_{i=1}^N \Delta_j^{(i)}, \quad (19)$$

where $\Delta_j^{(i)}$ is the residual difference defined in Eq. 15.

To put the significance of varying levels of improvement among spectra of different scales from different elements into perspective, we normalize $I(\Delta E)$ using the baseline model introduced in Sec. IV A as follows:

$$\bar{\mu}_{\text{baseline}} = \frac{1}{M} \sum_{j=1}^M [\mu_{\text{baseline}}]_j, \quad (20)$$

$$(21)$$

$$\tilde{I}(\Delta E) = \frac{I(\Delta E)}{\bar{\mu}_{\text{baseline}}} \times 100. \quad (22)$$

Positive values of $\tilde{I}(\Delta E)$ indicate better performance by the Tuned-UniversalXAS model, while negative values indicate better performance by the ExpertXAS model.

Fig. 11 displays $|\tilde{I}(\Delta E)|$ as bar plots for each element, partitioned by the sign of $\tilde{I}(\Delta E)$. To provide context on the practical significance of the improvement, we overlay μ_{baseline} with a dotted line corresponding to the y axis the right of each plot. The overlaid μ_{baseline} serves as a reference line to cross-relate the performance improvements to regions of practical significance. For all cases, except Cu FEFF and Ti VASP, the green bars dominate the plot, showcasing that the Tuned-UniversalXAS model consistently outperforms the ExpertXAS model across wide range of energy grid. This aligns with earlier observations based on w_e , where Cu FEFF and Ti VASP have the lowest scores. Furthermore, cross-comparison with the average spectrum μ_{baseline} reveals that the improvements are more pronounced around regions of the main edge and post-edge that are of high importance in XANES.

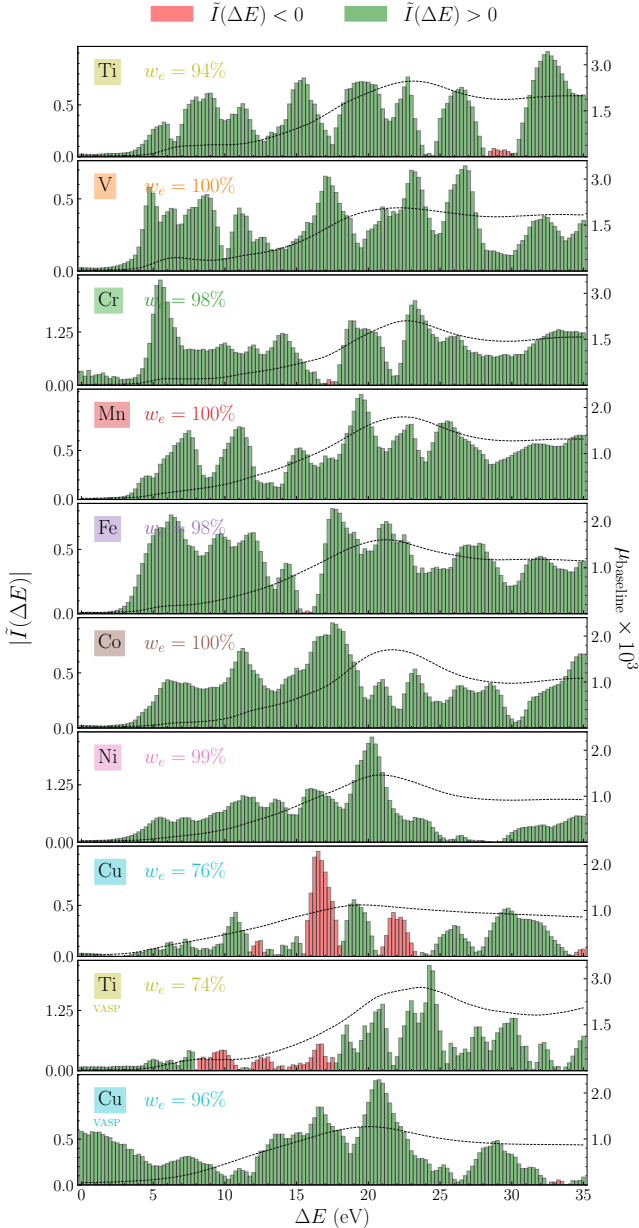


FIG. 11. Energy-resolved performance improvement $\tilde{I}(\Delta E)$ when using the Tuned-UniversalXAS model over the ExpertXAS model. The regions of improvement in Tuned-UniversalXAS predictions are highlighted in green, while the regions of degradation are highlighted in red. The spectrum used for normalization (μ_{baseline}) is overlaid as a dotted line corresponding to the right axis.

V. DISCUSSION

In this section, we reflect on the broader significance of our findings. In Sec. V A, we consider the impact of transfer-features for XAS prediction, highlighting their practical advantages and how they contribute to the improved model performance. In Sec. V B, we delve into the observed variations in model performance across ele-

ments, discussing the factors that may explain these differences and their implications for generalization. The discussion continues with an examination of how the UniversalXAS model bridges the elemental divide, offering potential for more widespread applications across different materials in Sec. V C. Furthermore, we assess how knowledge transfer from UniversalXAS enhances the performance of ExpertXAS models in Sec. V D. In Sec. V E, we address the model’s ability to cross fidelity boundaries, demonstrating the utility of this approach across different simulation environments. Finally, we highlight the huge computational speed up of XANES spectral prediction with OmniXAS as compared to first-principles simulations in Sec. V F.

A. Transfer-Features: Efficacy and Beyond

The efficacy of transfer-features for XAS prediction is remarkably evident across various model complexities, as demonstrated in Fig. 6. Training of M3GNet on interatomic potential prediction has evidently created a rich representation of materials structures that translates exceptionally well to XAS prediction tasks. While some degree of success was anticipated given the shared focus on local atomic environments, the extent of the transfer-features’ efficacy is particularly noteworthy.

Transfer-features demonstrate a marked superiority in performance compared to well-established structural descriptors, such as ACSF and SOAP. Although all these methods aim to characterize local atomic environments, the distinct edge exhibited by transfer-features suggests they capture additional, highly relevant information. The hierarchical architecture of M3GNet’s graph neural network may play a crucial role in this success, enabling a multi-scale representation that adeptly incorporates both local and intermediate-range interactions.

Beyond their predictive power, transfer-features offer several practical advantages, over ACSF and SOAP, that make them particularly suited for large-scale XAS predictions. Particularly:

- *Compactness*— Transfer features exhibit higher information density for XAS prediction compared to ACSF and SOAP. When subjected to dimensionality reduction techniques like PCA (preserving 99% of explained variance), ACSF and SOAP features undergo a significant reduction—approximately two orders of magnitude—from their initial representations in the thousands to below 100 dimensions (see Supplementary Fig. S4). In contrast, transfer features, originating from a compact 64-dimensional space, experience only minimal reduction. This indicates that transfer features contain less redundancy and encode XAS-relevant information in a more compact manner.
- *Scalability*— Transfer-features offer computational speed advantages that scale well with system size

and dataset complexity. The transfer-features can be generated for large datasets due to native GPU support. Moreover, transfer-features are compact and do not need to undergo additional dimensionality reduction post-processing before practical use, a process that typically involves computationally expensive techniques like PCA that scale cubically with feature dimension [67].

- *Efficiency*— Even simple linear models exhibit enhanced performance when trained with transfer-features as shown in Fig. 6 and Supplementary Table S7. This opens up possibilities for more accessible and computationally efficient ML modeling for XAS prediction, particularly beneficial for larger-scale studies or resource-constrained environments.

These practical advantages, combined with their strong predictive performance, position transfer-features as a promising approach for large-scale XAS predictions across materials.

B. Variations in Model Performance

While all ExpertXAS models performed above baseline, their performance varies widely across elements, with η ranging from 4.75 to 17.66, warranting a discussion of the underlying causes.

As is typical in machine learning, larger datasets generally lead to better performance, and we observe somewhat similar trend here. For instance, Mn, with the largest dataset exhibits the best performance. However, this relationship is not entirely trivial. For example, Co and V have nearly identical dataset sizes (10,753 vs. 10,813), yet their performance differs significantly (14.47 vs. 7.30). This suggests that dataset size alone does not fully account for the performance variation. We provide Supplementary Fig. S5 for illustrative purposes to showcase this empirical observation.

Beyond dataset size, other factors such as the complexity of local atomic environments and variations in spectra can also influence performance. Similarly, the transferability of local encodings from inter-atomic potential prediction to XAS prediction can vary across elements, affecting the model’s accuracy. This interplay of model performance with dataset size, spectral complexity, and feature transfer is non-trivial and intersects with the broader domain of neural scaling laws [68] that is beyond the scope of this work.

C. Crossing the Elemental Divide

The UniversalXAS model, trained on all FEFF data simultaneously, operates as a multi-task model designed to generalize across multiple elements. Although its performance is generally lower compared to the element-specific ExpertXAS models—an expected outcome given

the broad generalization goal—what is particularly noteworthy is that UniversalXAS still performs well above the baseline for all elements. This demonstrates its robustness and effectiveness as a generalizable XAS prediction framework.

The success of UniversalXAS challenges the traditional element-by-element approach for XAS analysis, showing that a model can predict spectra for a wide range of materials without sacrificing significant accuracy. Interestingly, the performance trends of UniversalXAS closely mirror those of the ExpertXAS models. The elements where ExpertXAS performed well, such as Mn and Fe, are also the ones where UniversalXAS achieved higher accuracy. This parallel behavior suggests that while UniversalXAS captures shared spectral characteristics across elements, it still retains the ability to identify and adapt to element-specific features that define each element’s XAS spectra.

D. Transfer of Universal Knowledge

The Tuned-UniversalXAS model demonstrates significant performance improvements after being fine-tuned for individual elements, which aligns with expectations. However, the most striking result is its consistent ability to surpass the performance of element-specific ExpertXAS models (with one exception for Cu FEFF). This outcome highlights a key finding of this study. Considering the novelty of this approach, we delve deeper to establish the significance of this observation, the consistency of the improvements, and the practical relevance of these gains.

The first point to note regarding the improvement in η values is that while there is a clear performance enhancement, the extent of this improvement varies across elements—a theme consistent with the broader discussion on model performance variations. While Fig. 7 and Table I clearly show the extent of improvement in η , a caveat to note is that similar improvements in η do not necessarily come with the same level of confidence due to variations in dataset sizes. To ensure that these improvements are not artifacts of random variation, we conducted a paired difference test, which confirmed the statistical significance of the observed gains (refer to Supplementary Fig. S6 for more details).

In Sec. IV E 1, we examined whether the performance differences based on the η metric alone might be concealing any systematic biases or outliers. Fig. 10 compares the differences between the Tuned-UniversalXAS and ExpertXAS models across the entire energy grid for all spectra. The distribution of the residual differences reveals a clear pattern: when considering each prediction point, the changes in prediction accuracy are predominantly positive, spread across the energy grid, and not driven by isolated outliers or noise.

The results from Sec. IV E 3 further elucidate the nature of these improvements, showing that the regions

of improvement within the spectra are consistent with the regions of importance in XAS (e.g., main edge and post edge). This alignment underscores the practical relevance of the performance gains, as they are concentrated in the critical regions needed for accurate XAS analysis. The win rates (percentage of spectra where the Tuned-UniversalXAS model outperforms the ExpertXAS model) shown in the the same Fig. 11 shed insight into how these localized gains translate into improved prediction accuracy at the spectrum level.

E. Crossing the Fidelity Divide

The performance improvement of the Tuned-UniversalXAS model over both VASP datasets is clear and definitive, showing a 10.82% improvement for Ti and 8.92% for Cu. This is further supported by spectrum-wise win rates (63.13% for Ti and 66.04% for Cu), as well as consistent improvements across energy points and key regions of significance.

This result represents a distinctive benefit in crossing the fidelity divide between FEFF and VASP simulations. While previous efforts focused on fine-tuning within a single fidelity (FEFF), the current findings demonstrate how a cascading transfer learning approach can effectively transfer knowledge from low-cost FEFF simulations to the more computationally intensive VASP simulations, offering substantial practical advantages for efficient and scalable predictions.

F. Computational Performance

We envision that the proposed OmniXAS framework provides a practical solution to tackle the computational barrier in XAS spectral analysis. In average, an XAS-Model takes only a fraction of second (about 0.16 second) to predict a material XANES spectrum, with most of the time spent on featurization (i.e., the M3GNet Block). On the other hand, performing a XANES spectral simulation using VASP requires high-performance computing facilities and an exemplary calculation on Ti K-edge XANES of anatase TiO_2 takes about 11.57 CPU hours on Intel(R) Xeon(R) Gold 6336Y processors. Therefore, deep learning XASModels can lead to a remarkable speed up of over five orders of magnitude in XAS prediction. The huge computational speed up of the OmniXAS framework should be comparable or better than typical graph neural network potential frameworks, because first-principles spectral calculations are more expensive than total energy calculations of the ground state. Such a highly efficient XAS prediction capability allows high-throughput XAS modeling applied to a broad material space and opens the door for real-time XAS analysis.

VI. CONCLUSION

In this study, we developed OmniXAS, a universal deep learning framework for predicting K-edge X-ray absorption spectra of eight 3d transition metals (Ti – Cu) based on the 3dtm_xanes_ml2024 dataset, utilizing a cascading combination of inductive transfer learning and domain adaptation techniques. The integration of M3GNet-derived latent features, referred to as transfer-features, consistently outperformed traditional methods like ACSF and SOAP, demonstrating clear advantages for large-scale applications.

By training a single universal model, we successfully addressed the elemental divide often encountered in machine learning models for materials science, establishing a robust foundation for general-purpose XAS predictions across diverse elements.

Our non-linear embedding analysis of the latent space further reinforced the efficacy of this approach. Fine-tuning the UniversalXAS model into Tuned-UniversalXAS models led to substantial performance improvements across nearly all elements, with these gains being both quantitatively significant and relevant for practical applications. Furthermore, the method demonstrated its ability to extend across fidelity boundaries, bringing higher performance efficiency from limited amount of computationally intensive dataset. This transfer learning framework is generalizable to enhance deep-learning models that target other properties in materials research.

DATA & SOFTWARE AVAILABILITY

The software implementation associated with this study is openly available on GitHub (<https://github.com/AI-multimodal/OmniXAS>) [69]. The 3dtm_xanes_ml2024 dataset used in the machine learning model development of this work is available in the Zenodo repository (<https://zenodo.org/records/14145457>). Input and output of the spectral simulation will be made available through the Materials Cloud platform [70] (<https://doi.org/10.24435/materialscloud:85-8x>).

ACKNOWLEDGMENTS

This research is supported by the U.S. Department of Energy, Office of Science, Office Basic Energy Sciences, Award Number FWP PS-030. This research used theory and computation resources of the Center for Functional Nanomaterials, which is a U.S. Department of Energy Office of Science User Facility, and the Scientific Data and Computing Center, at Brookhaven National Laboratory (BNL) under Contract No. DE-SC0012704. M. R. C. acknowledges BNL Laboratory Directed Research and Development (LDRD) grant no. 24-004. This research also

used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC Awards No.

BES-ERCAP-20690, 18006, and 14811. We thank Mark Hybertsen for helpful discussions. We would also like to extend our gratitude to Nina Cao and Pavan Ravindra for their valuable feedback on manuscript.

-
- [1] J. J. Rehr and R. C. Albers, Theoretical approaches to x-ray absorption fine structure, *Rev. Mod. Phys.* **72**, 621 (2000).
- [2] J. E. Penner-Hahn, X-ray absorption spectroscopy in coordination chemistry, *Coordination Chemistry Reviews* **190**, 1101 (1999).
- [3] F. De Groot and A. Kotani, *Core level spectroscopy of solids* (CRC press, 2008).
- [4] F. Farges, G. E. Brown, J. Rehr, *et al.*, Ti k-edge xanes studies of ti coordination and disorder in oxide compounds: Comparison between theory and experiment, *Physical Review B* **56**, 1809 (1997).
- [5] O. Müller, M. Nachtgeal, J. Just, D. Lützenkirchen-Hecht, and R. Frahm, Quick-exafs setup at the superxas beamline for in situ x-ray absorption spectroscopy with 10 ms time resolution, *Journal of synchrotron radiation* **23**, 260 (2016).
- [6] F. De Groot, M. Krisch, and J. Vogel, Spectral sharpening of the pt l edges by high-resolution x-ray emission, *Physical Review B* **66**, 195112 (2002).
- [7] K. G. Yager, P. W. Majewski, M. M. Noack, and M. Fukuto, Autonomous x-ray scattering, *Nanotechnology* **34**, 322001 (2023).
- [8] P. M. Maffettone, D. B. Allan, S. I. Campbell, M. R. Carbone, T. A. Caswell, B. L. DeCost, D. Gavrilo, M. D. Hanwell, H. Joress, J. Lynch, *et al.*, Self-driving multimodal studies at user facilities, arXiv preprint arXiv:2301.09177 (2023).
- [9] M. R. Carbone, H. J. Kim, C. Fernando, S. Yoo, D. Olds, H. Joress, B. DeCost, B. Ravel, Y. Zhang, and P. M. Maffettone, Flexible formulation of value for experiment interpretation and design, *Matter* **7**, 685 (2024).
- [10] F. Meng, B. Maurer, F. Peschel, S. Selcuk, M. Hybertsen, X. Qu, C. Vorwerk, C. Draxl, J. Vinson, and D. Lu, Multicode benchmark on simulated ti k-edge x-ray absorption spectra of ti-o compounds, *Phys. Rev. Mater.* **8**, 013801 (2024).
- [11] M. Newville, Larch: an analysis package for xafs and related spectroscopies, in *Journal of Physics: Conference Series*, Vol. 430 (IOP Publishing, 2013) p. 012007.
- [12] J. Kas, F. Vila, C. Pemmaraju, T. Tan, and J. Rehr, Advanced calculations of x-ray spectroscopies with feff10 and corvus, *Journal of Synchrotron Radiation* **28**, 1801 (2021).
- [13] M. R. Carbone, F. Meng, C. Vorwerk, B. Maurer, F. Peschel, X. Qu, E. Stavitski, C. Draxl, J. Vinson, and D. Lu, Lightshow: a python package for generating computational x-ray absorption spectroscopy input files, *Journal of Open Source Software* **8**, 5182 (2023).
- [14] K. Mathew, C. Zheng, D. Winston, C. Chen, A. Dozier, J. J. Rehr, S. P. Ong, and K. A. Persson, High-throughput computational x-ray absorption spectroscopy, *Scientific data* **5**, 1 (2018).
- [15] S. B. Torrisi, M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram, and L. Hung, Random forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships, *npj Comput. Mater.* **6**, 1 (2020).
- [16] Y. Chen, C. Chen, C. Zheng, S. Dwaraknath, M. K. Horton, J. Cabana, J. Rehr, J. Vinson, A. Dozier, J. J. Kas, *et al.*, Database of ab initio l-edge x-ray absorption near edge structure, *Scientific data* **8**, 153 (2021).
- [17] A. A. Guda, S. A. Guda, A. Martini, A. Kravtsova, A. Algasov, A. Bugaev, S. P. Kubrin, L. Guda, P. Šot, J. A. van Bokhoven, *et al.*, Understanding x-ray absorption spectra by means of descriptors and machine learning algorithms, *npj Computational Materials* **7**, 203 (2021).
- [18] C. D. Rankine and T. Penfold, Accurate, affordable, and generalizable machine learning simulations of transition metal x-ray absorption spectra using the xanesnet deep neural network, *J. Chem. Phys.* **156**, 164102 (2022).
- [19] H. Guo, M. R. Carbone, C. Cao, J. Qu, Y. Du, S.-M. Bak, C. Weiland, F. Wang, S. Yoo, N. Artrith, *et al.*, Simulated sulfur k-edge x-ray absorption spectroscopy database of lithium thiophosphate solid electrolytes, *Scientific data* **10**, 349 (2023).
- [20] C. Cao, M. R. Carbone, C. Komurcuoglu, J. S. Shekhawat, K. Sun, H. Guo, S. Liu, K. Chen, S.-M. Bak, Y. Du, *et al.*, Atomic insights into the oxidative degradation mechanisms of sulfide solid electrolytes, *Cell Reports Physical Science* **5** (2024).
- [21] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [22] M. R. Carbone, M. Topsakal, D. Lu, and S. Yoo, Machine-learning x-ray absorption spectra to quantitative accuracy, *Phys. Rev. Lett.* **124**, 156401 (2020).
- [23] A. Ghose, M. Segal, F. Meng, Z. Liang, M. S. Hybertsen, X. Qu, E. Stavitski, S. Yoo, D. Lu, and M. R. Carbone, Uncertainty-aware predictions of molecular x-ray absorption spectra using neural network ensembles, *Physical Review Research* **5**, 013180 (2023).
- [24] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, *Sci. Data* **1**, 1 (2014).
- [25] H. Kwon, T. Hsu, W. Sun, W. Jeong, F. Aydin, J. Chapman, X. Chen, M. R. Carbone, D. Lu, F. Zhou, *et al.*, Spectroscopy-guided discovery of three-dimensional structures of disordered materials with diffusion models, arXiv preprint arXiv:2312.05472 (2023).
- [26] H. Kwon, W. Sun, T. Hsu, W. Jeong, F. Aydin, S. Sharma, F. Meng, M. R. Carbone, X. Chen, D. Lu, *et al.*, Harnessing neural networks for elucidating x-ray absorption structure–spectrum relationships in amorphous carbon, *The Journal of Physical Chemistry C* **127**, 16473 (2023).

- [27] C. D. Rankine, M. M. Madkhali, and T. J. Penfold, A deep neural network for the rapid prediction of x-ray absorption spectra, *The Journal of Physical Chemistry A* **124**, 4263 (2020).
- [28] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nature Computational Science* **2**, 718 (2022).
- [29] J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen, Parameter-free calculations of x-ray spectra with feff9, *Phys. Chem. Chem. Phys.* **12**, 5503 (2010).
- [30] F. Karsai, M. Humer, E. Flage-Larsen, P. Blaha, and G. Kresse, Effects of electron-phonon coupling on absorption spectrum: K edge of hexagonal boron nitride, *Physical Review B* **98**, 235205 (2018).
- [31] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python materials genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.* **68**, 314 (2013).
- [32] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, The numpy array: a structure for efficient numerical computation, *Computing in science & engineering* **13**, 22 (2011).
- [33] [Numpy.std v2.1 manual](#).
- [34] A. H. England, A. M. Duffin, C. P. Schwartz, J. S. Uejio, D. Prendergast, and R. J. Saykally, On the hydration and hydrolysis of carbon dioxide, *Chemical Physics Letters* **514**, 187 (2011).
- [35] D. Carta, G. Mountjoy, A. Regoutz, A. Khiat, A. Serb, and T. Prodromakis, X-ray absorption spectroscopy study of tio₂-x thin films for memory applications, *The Journal of Physical Chemistry C* **119**, 4362 (2015).
- [36] A. Gaur, B. Shrivastava, and S. Joshi, Copper k-edge xanes of cu (i) and cu (ii) oxide mixtures, in *Journal of Physics: Conference Series*, Vol. 190 (IOP Publishing, 2009) p. 012084.
- [37] L. Torrey and J. Shavlik, Transfer learning, in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (IGI global, 2010) pp. 242–264.
- [38] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, A comprehensive survey on transfer learning, *Proceedings of the IEEE* **109**, 43 (2020).
- [39] L. Duan, D. Xu, and I. Tsang, Learning with augmented features for heterogeneous domain adaptation, *arXiv preprint arXiv:1206.4660* (2012).
- [40] B. Kulis, K. Saenko, and T. Darrell, What you saw is not what you get: Domain adaptation using asymmetric kernel transforms, in *CVPR 2011* (IEEE, 2011) pp. 1785–1792.
- [41] Y. Zhu, Y. Chen, Z. Lu, S. Pan, G.-R. Xue, Y. Yu, and Q. Yang, Heterogeneous transfer learning for image classification, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 25 (2011) pp. 1304–1309.
- [42] J. Zhou, S. Pan, I. Tsang, and Y. Yan, Hybrid heterogeneous transfer learning through deep learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 28 (2014).
- [43] P. Prettenhofer and B. Stein, Cross-language text classification using structural correspondence learning, in *Proceedings of the 48th annual meeting of the association for computational linguistics* (2010) pp. 1118–1127.
- [44] J. T. Zhou, I. W. Tsang, S. J. Pan, and M. Tan, Heterogeneous domain adaptation for multiple classes, in *Artificial intelligence and statistics* (PMLR, 2014) pp. 1095–1103.
- [45] R. Xia and C. Zong, A pos-based ensemble model for cross-domain sentiment classification, in *Proceedings of 5th international joint conference on natural language processing* (2011) pp. 614–622.
- [46] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345 (2010).
- [47] J. Bergengren, Über die röntgenabsorption des phosphors, *Zeitschrift für Physik* **3**, 247 (1920).
- [48] V. Kunzl, A linear dependence of energy levels on the valency of elements, *Collection of Czechoslovak Chemical Communications* **4**, 213 (1932).
- [49] L. Pauling, *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*, Vol. 18 (Cornell university press, 1960).
- [50] J. Suchet, Physical chemistry of semiconductors, *Chemical Physics of Semiconductors* , 224 (1965).
- [51] T. Yamamoto, Assignment of pre-edge peaks in k-edge x-ray absorption spectra of 3d transition metal compounds: electric dipole or quadrupole?, *X-Ray Spectrometry: An International Journal* **37**, 572 (2008).
- [52] M. R. Carbone, S. Yoo, M. Topsakal, and D. Lu, Classification of local chemical environments from x-ray absorption spectra using supervised machine learning, *Phys. Rev. Mater.* **3**, 033604 (2019).
- [53] Z. Liang, M. R. Carbone, W. Chen, F. Meng, E. Stavitski, D. Lu, M. S. Hybertsen, and X. Qu, Decoding structure-spectrum relationships with physically organized latent spaces, *Physical Review Materials* **7**, 053802 (2023).
- [54] J. Krause, A. Perer, and K. Ng, Interacting with predictions: Visual inspection of black-box machine learning models, in *Proceedings of the 2016 CHI conference on human factors in computing systems* (2016) pp. 5686–5697.
- [55] D. Castelvechi, Can we open the black box of ai?, *Nature News* **538**, 20 (2016).
- [56] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *International conference on machine learning* (pmlr, 2015) pp. 448–456.
- [57] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions, *arXiv preprint arXiv:1710.05941* (2017).
- [58] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, *arXiv preprint arXiv:1207.0580* (2012).
- [59] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* <https://doi.org/10.1145/3292500.3330701> (2019).
- [60] D. P. Kingma, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [61] [Learningratefinder — pytorch lightning 2.2.5 documentation](#) (2024).
- [62] L. McInnes, J. Healy, and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, *arXiv preprint arXiv:1802.03426* (2018).
- [63] C. Cai and Y. Wang, [A note on over-smoothing for graph neural networks](#) (2020), [arXiv:2006.13318 \[cs.LG\]](#).

- [64] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, Measuring and relieving the over-smoothing problem for graph neural networks from the topological view, in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34 (2020) pp. 3438–3445.
- [65] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.* **134**, 074106 (2011).
- [66] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).
- [67] G. Abraham and M. Inouye, Fast principal component analysis of large-scale genome-wide data, *PloS one* **9**, e93766 (2014).
- [68] Y. Bahri, E. Dyer, J. Kaplan, J. Lee, and U. Sharma, Explaining neural scaling laws, arXiv preprint arXiv:2102.06701 (2021).
- [69] AI-multimodal, [Ai-multimodal/omnixas](#) (2024).
- [70] L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, *et al.*, Materials cloud, a platform for open computational science, *Scientific data* **7**, 299 (2020).