

# Determining electronic properties from *L*-edge x-ray absorption spectra of transition metal compounds with artificial neural networks

Johann Lüder \*

*Department of Materials and Optoelectronic Science, National Sun Yat-sen University, 70 Lien-Hai Rd.,  
Kaohsiung 80424, Taiwan, Republic of China*  
*and Center for Crystal Research, National Sun Yat-sen University, 70 Lien-Hai Rd.,  
Kaohsiung 80424, Taiwan, Republic of China*



(Received 19 November 2019; accepted 10 December 2020; published 29 January 2021)

X-ray absorption spectroscopy at the *L*-edge probes transitions of  $2p$  electrons into unoccupied *d* states. Applied to transition metal atoms, this experimental technique can provide valuable information about the electronic structure of *d* states. However, multiplet effects, spin-orbit coupling, and a large number of possible transitions can cause a rather involved nature of  $2p$  XAS spectra, which can often complicate extracting of information directly from them. Here, artificial neural networks trained on simulated spectra of a  $2p$  XAS model Hamiltonian are presented that can directly determine information about atomic properties and the electronic configuration of *d* states from *L*-edge x-ray absorption spectra. Moreover, the adaptable nature of artificial neural networks (ANNs) allows extending their capability to obtain information about the electronic ground state and core hole lifetimes from  $2p$  XAS spectra as well as to incorporate external factors, such as temperature and experimental convolution that can affect details in spectral features. The effects of noise and background contributions in spectra on the accuracy of ANNs are discussed and the method is validated on experimental spectra of transition metal compounds, including metal-organic molecules and metal oxides.

DOI: [10.1103/PhysRevB.103.045140](https://doi.org/10.1103/PhysRevB.103.045140)

## I. INTRODUCTION

X-ray spectroscopy (XS) is an important tool for many branches of science and technology [1–3]. Since the discovery of x rays, a large variety of specialized XS techniques have been developed, and many of them are routinely used as scientific tools, providing detailed information about properties of materials that lead to accelerated technological and scientific advancements [4–11]. Many technological applications, such as environmental sensors, catalysts, and energy storage and harvesting devices, can benefit from these insights, which allow us to understand and optimize processes in materials, e.g., involving details of nanostructures [12], charge transfer effects [13], and chemical stability [14,15], that can be employed to enhance their design and functionalization. One specialized technique is soft x-ray absorption spectroscopy (XAS) at the *L* edge that employs tunable photon energies probing energy-dependent transitions of electrons from occupied to unoccupied states obeying dipole selection rules [16]. This technique is element, state, and orientation selective. In light transition metal (TM) atoms, this technique can probe electron transitions from occupied  $2p$  to unoccupied  $3d$  states induced by the absorption of photons and help to reveal the nature of their  $3d$  states.

For extensive analysis, experimental XAS results are frequently combined with theoretical simulations [17–20]. Then, experimental and computational insights can explain possible

causes leading to changes in spectra. Detailed pictures can be drawn from the underlying mechanisms, explaining the role of the electronic and chemical structure and relate them to changes in properties of materials [21–26]. However, the independent particle methods, that are often used in other branches of theoretical spectroscopy [27–29], cannot describe  $2p$  XAS processes in TM ions, due to strong electron-hole interactions, multiplet effects, and spin-orbit coupling (SOC) [30,31]. These effects govern the electronic transitions in  $2p$  XAS and strongly affecting spectral shapes, including the nontrivial branching ratio between peaks corresponding to transitions from  $p_{3/2}$  and  $p_{1/2}$  semicore states. Several theoretical many-body methods have been developed to model  $2p$  XAS spectra at the *L* edge, such as first-principle methods employing, e.g., the Bethe-Salpeter equation [32–35] and configuration interaction related methods [36,37]. In contrast to first-principle methods, XAS model Hamiltonians are computationally efficient and can yield very good agreement with experiments [38,39], although their parametrization can be challenging and time consuming. Recently, model Hamiltonian approaches (MHA) were combined with DFT results through extracting local properties using maximally localized Wannier functions [40] or extracting the parametrization from the hybridization function, which is known from the dynamical mean-field theory [41]. Still, none of these methods provides a direct way to determine electronic properties from spectra, but rather require computing the electronic structure, followed by comparing an obtained theoretical XAS spectrum with the experimental spectrum. Additionally, the possibly large number of transitions can lead to feature-rich spectra that

\*johann.lueder@mail.nsysu.edu.tw

are difficult to analyze. To directly analyze spectra, a function is needed that reverses the conventional computation of  $L$ -edge XAS spectra performed with, e.g., a MHA. Then, the model Hamiltonian can be used to determine physical properties such as the electronic configuration of the ground state and, in principle, also the theoretical  $L$ -edge XAS spectrum with the results given by the inverse function. This function is, however, unknown.

Artificial neural networks (ANNs) have attracted much attention in material science [42–44] because they can fit any real-valued function with arbitrary accuracy [45] and allow unbiased learning of fundamental relations from data. Employing ANNs may lead to solutions to complex problems, as ANNs can accelerate computational methods and accelerate materials discovery [46–51]. For instance, ANNs are frequently combined with density functional theory (DFT) computations [52–54] to efficiently predict structures and properties of compounds, which can be computationally expensive by DFT means [55–57]. With regard to  $2p$  XAS spectroscopy, one question at hand is: Is it possible to construct useful ANNs that represent this inversion and can obtain information about the electronic structure from experimental  $2p$  XAS spectra in the form of parameters used in, e.g., a MHA?

Here, an ANN method is presented that obtains information on the TM ion's local electronic structure from experimental  $L$ -edge XAS spectra in the form of parameters used in a MHA. The MHA is the basis of the ANN method. The ANN can obtain relative energy levels of the  $d$  states in a TM ion and screening factors of Slater-Condon integrals. Furthermore, extensions of the method provide estimations of core-hole lifetimes and can consider external factors such as temperature and experimental broadening, as well as predictions of expectation values of the MHA, including spin state and orbital occupations. Several ANN architectures were tested to demonstrate limitations, capabilities, and accuracy dependencies. The method is applied to and tested with previously published experimental  $2p$  XAS spectra of well-studied systems of transition metal monoxides and transition metal-organic complexes.

## II. METHODS

In the following, an ANN approach is described that can extract information about the local electronic structure of transition metal ions such as relative  $d$ -level positions and  $d$ -state occupations from  $2p$  XAS spectra with high accuracy. Then, the influence of external factors such as temperature and experimental convolution as well as expectation values prediction (EP) is included through extensions of the method. The ANNs are constructed and trained on simulated  $2p$  XAS spectra, the so-called reference spectra, computed for light TM ions. The employed  $2p$  XAS MHA is based on crystal-field multiplet theory (CMT) while charge transfer (CT) effects are neglected. Since  $2p$  XAS spectra of light transition metal have large energy separations (i.e., spectroscopic fingerprint), the approach works with specialized ANNs for each element. The trained ANNs can be applied to experimental spectra to extract sets of parameters (that are atomic and ligand field parameters in the language of CMT) describing

characteristic electronic properties. Then, the determined parameters allow recomputing any spectrum with the same theoretical framework as the reference spectra to estimate the accuracy of the method directly. In the following, CMT and the ANN implementations are briefly explained. Then, the performance of different ANN types is compared and the influence of noise and background in spectra is discussed before the method is applied to experimental spectra.

### A. A $2p$ XAS model Hamiltonian

For TM complexes with partly filled  $3d$  shell, a CMT based approach is used to build a database of computed  $L$ -edge XAS spectra for  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  ions. CMT considers a central metal ion described by valence  $3d$  states and  $2p$  semicore states. The TM ion is placed in a crystal field (CF) representing the chemical surrounding, leading to the lifting of the degeneracy of  $3d$  states. The strength of the CF is described by optical parameters, for instance, in Ballhausen notation [58]. The  $d$ -level splitting is then given for  $D_{4h}$  symmetry by

$$\begin{aligned}\epsilon_{d_{a_{1g}}} &= 6D_q - 2D_s - 6D_t \\ \epsilon_{d_{b_{1g}}} &= 6D_q + 2D_s - D_t \\ \epsilon_{d_{b_{2g}}} &= -4D_q + 2D_s - D_t \\ \epsilon_{d_{e_g}} &= -4D_q - D_s + 4D_t,\end{aligned}\quad (1)$$

where  $D_q$ ,  $D_s$ , and  $D_t$  are Ballhausen parameters for low symmetry CFs. For  $O_h$  symmetries, the  $d$ -level split reads as

$$\begin{aligned}\epsilon_{d_{e_g}} &= 6 \cdot D_q \\ \epsilon_{d_{t_{2g}}} &= -4 \cdot D_q.\end{aligned}\quad (2)$$

Negative values of  $D_q$  in Eq. (2) yield a  $C_4$  symmetric CF [58]. The  $L$ -edge XAS model Hamiltonian [40] is

$$\begin{aligned}\hat{H} &= \sum_{i,j} \epsilon_{d_{i,j}} \hat{d}_i^\dagger \hat{d}_j + \sum_{i,j} \epsilon_p \hat{p}_i^\dagger \hat{p}_j \\ &+ \zeta_d \sum_{i,j} \langle d_i | \vec{l} \cdot \vec{s} | d_j \rangle \hat{d}_i^\dagger \hat{d}_j + \zeta_p \sum_{i,j} \langle p_i | \vec{l} \cdot \vec{s} | p_j \rangle \hat{p}_i^\dagger \hat{p}_j \\ &+ \sum_{i,j,k,l} U_{ijkl}^{dd} \hat{d}_i^\dagger \hat{d}_j^\dagger \hat{d}_l \hat{d}_k + \sum_{i,j,k,l} U_{ijkl}^{dp} \hat{d}_i^\dagger \hat{p}_j^\dagger \hat{p}_l \hat{d}_k.\end{aligned}\quad (3)$$

In Eq. (3), the nonrelativistic single-electron  $d$ - and  $p$ -state energies are given by  $\epsilon_d$  and  $\epsilon_p$ , respectively. The corresponding  $\hat{d}_j$  ( $\hat{d}_j^\dagger$ ), and  $\hat{p}_j$  ( $\hat{p}_j^\dagger$ ) operators annihilate (create) states. Spin-orbit coupling constants are  $\zeta_d$  for  $d$  states and  $\zeta_p$  for  $p$  states. They are linked by the vector product  $\vec{l} \cdot \vec{s}$  with the angular-momentum operator ( $\vec{l}$ ) and the spin operator ( $\vec{s}$ ). The onsite Coulomb interaction between  $d$  electrons can be described by Slater-Condon integrals  $F_{dd}^0$ ,  $F_{dd}^2$ , and  $F_{dd}^4$ . Likewise, the interaction between  $d$  electrons and  $p$ -core hole are described by  $F_{pd}^0$ ,  $F_{pd}^2$ ,  $F_{pd}^4$ ,  $G_{pd}^1$ , and  $G_{pd}^3$  [31]. The monopole terms between  $3d$  electrons, i.e.,  $F_{dd}^0$ , and between  $3d$  and  $2p$  electrons, i.e.,  $F_{pd}^0$ , are expressed through  $U_{ijkl}^{dd}$ ,  $F_{dd}^2$ , and  $F_{dd}^4$  and  $U_{ijkl}^{dp}$ ,  $G_{pd}^1$ , and  $G_{pd}^3$ , respectively [59], where  $U$ 's are

Hubbard- $U$  parameters; they read

$$U_{dd} = F_{dd}^0 - \frac{2}{63}(F_{dd}^2 + F_{dd}^4) \quad (4)$$

and

$$U_{pd} = F_{pd}^0 - \frac{1}{15}G_{pd}^1 - \frac{3}{70}G_{pd}^3. \quad (5)$$

The values of the Slater-Condon integrals other than  $F_{dd}^0$  ( $F_{pd}^0$ ) are taken from atomic calculations [59]. The  $F^k$  ( $G^k$ ) integral values representing direct (exchange) Coulomb interactions are scaled by a factor  $S_1$  ( $S_2$ ) that is usually about 0.8 to account for correlation effects and electronic screening in solids [31]. In CMT, the effects of  $U_{pd}$  and  $U_{dd}$  are canceled out between ground and excited states since the total number of electrons does not change [59].

The XAS spectra can be obtained with the solutions of Eq. (3) yielding eigenstates  $|\psi_i\rangle$  as a sum of Slater determinants with eigenenergy  $E_i$

$$\hat{H} |\psi_i\rangle = E_i |\psi_i\rangle. \quad (6)$$

Then, the XAS spectrum can be obtained from

$$I(\omega) = -\frac{1}{Z} \sum_i \Im \left( \langle \psi_i | \hat{D}^\dagger \frac{1}{\omega - \hat{H} + E_i + i\Gamma/2} \hat{D} | \psi_i \rangle \right) \times \exp(-\beta E_i), \quad (7)$$

where  $\omega$  represents the excitation energy, and  $\hat{D}$  is the dipole operator describing a  $p$ -electron excitation into an unoccupied  $d$  state. Equation (7) results in a Lorentzian broadening of the transition intensities through the imaginary shift  $\Gamma$  from the real axis, with  $\beta$  being the inverse temperature representing 10 K (if not stated otherwise) and  $Z$  is the partition function. The broadening of computed transition intensities, frequently performed to facilitate comparison with experimental spectra [22–24,60,61], employs a Gaussian broadening (GB) for experimental convolution factors additionally to the Lorentzian broadening (LB) for lifetime broadening. GB is characterized by a constant factor, i.e., a full width at half maximum (FWHM). The LB based convolution depends on the core-hole lifetime. Generally, the lifetimes of holes in  $p_{3/2}$  and  $p_{1/2}$  states differ, leading to differences in peak widths in the  $L_3$  and  $L_2$  features, respectively. It is common to employ an energy-dependent stepwise function that uses two broadening factors  $\Gamma_1$  for  $L_3$  and  $\Gamma_2$  for  $L_2$  features before and after an energy window defined by  $E_1$  and  $E_2$ , and a linear function within the energy window that starts at  $\Gamma_1$  and ends at  $\Gamma_2$  [61,62]. If not stated otherwise, the core hole lifetimes in the tested TM ions were estimated to correspond to values of  $\Gamma$  between 0.3 to 0.5 eV with increasing values from Mn to Ni. Gaussian convolution of 0.1 eV accounted for other effects leading to broadening of the experimental spectra. The solutions of Eq. (6) and the consecutive convolution of the transition intensities were obtained with the Quany code [40] which performs a Lanczos algorithm on a random  $d^n$  configuration generated by the tridiagonal Krylov bases to determine the electronic ground state.

## B. Artificial neural networks

In the presented method, an ANN replaces the unknown inverse function, i.e., it can reconstruct the parametrization of a  $2p$  XAS model Hamiltonian from a spectrum. That is to say; it allows reading the parametrization of Eq. (3) from a spectrum, including the consecutively applied convolution of transition intensities and external factors.

For accurate results, the ANN needs a suitable network architecture and effective training, indicated by a positive evaluation of test data, before it can provide reliable results. The smallest building block of an ANN is an artificial neuron or node, which is a weighted processing unit with a simple mathematical representation. The activation function  $f$  describes how a neuron evaluates given information. Besides widely used sigmoidal functions [63,64], the rectified linear unit (RELU) function is a common choice in ANN supported computational materials science [65–67]. Individual neurons are collected in layers. In a fully connected (so-called dense) ANN, each neuron in a layer is connected to all neurons in neighboring layers. This allows passing information in one direction from layer to layer by so-called forward propagation. In an ANN, a single layer with  $D_N$  neurons taking input from  $D_i$  neurons in a previous layer is represented as

$$y_j = \sum_{p=0}^{D_N} w_{p,j} f \left( \sum_{i=0}^{D_i} w_{i,p} x_i + b_p \right), \quad (8)$$

where neurons process the input  $x_i$  through the activation function  $f$ , weights  $w_{j,p}$ , and bias  $b_p$ . This process starts at the input layer, i.e., the descriptor vector, and continues to subsequent hidden layers until it reaches the last layer, the so-called output layer, representing the output vector or feature vector. Then, complex relations can be learned from reference data constructing complex functions through a series of layers and neurons with adaptable weights. For multilayer ANNs with  $N$  hidden layers, a general form is given by

$$y_j = f_o \left( \sum_{k_N=0}^{D_N} w_{j,k_N}^{(N)} f_{N-1,k_{N-1}} \left( \sum_{k_{N-1}=0}^{D_{N-1}} w_{k_N,k_{N-1}}^{(N-1)} f_{N-2,k_{N-2}} \right. \right. \\ \left. \left. \times \dots \left( \sum_{k_1=0}^{D_1} w_{k_2,k_1}^{(2)} f_{1,k_1} \left( \sum_{i=0}^{D_i} w_{k_1,i}^{(1)} x_i \right) \right) \right) \right) \quad (9)$$

omitting any bias term. The last activation function  $f_o$ , which yields the output  $y$ , employs, in general, a linear function for regression tasks.

The ANN architecture describes the connection patterns between neurons within a network. This affects how neurons pass information through the network. The width is the number of neurons in a layer and the depth refers to the number of layers. Both width and depths are hyperparameters of the ANN in addition to learning rates and, e.g., momentum used in some optimizers.

During training, a randomly selected subset of reference data, i.e., training data, is used to optimize the weights in the ANN through back propagation employing the gradient of a cost function. The cost function compares and penalizes differences between predicted value ( $y_p$ ), i.e., values computed by the ANN through passing the input vector through

the network, and the actual value ( $y_a$ ). Common choices of cost functions for regression tasks are, for instance, the mean absolute error (MAE) or the root mean square error (RMSE).

Each ANN can be optimized with respect to hyperparameters, which are different from parameters, e.g., the weights, in the ANN. In contrast to trainable weights of ANNs, hyperparameters must be optimized outside the training of ANNs. A validation dataset containing different data from the training dataset can be used to evaluate the ANN performance concerning hyperparameters besides detecting overfitting. With adjusted weights and hyperparameters, the ANN is tested on an independent dataset, i.e., test dataset, that was used for neither training nor validation.

### C. Artificial neural network design

The input for an ANN is a reference (for training, test, and evaluation of the ANN) or an experimental (for application) spectrum. The output is a set of features that describe the local electronic properties of the central TM ion. Here, three types of ANNs are tested. The first ANN type, labeled A2pX, can determine  $d$ -level positions described by Ballhausen parameters and screening factors in the model Hamiltonian Eq. (3), while lifetime broadening factors are taken from literature [68], the experimental broadening is estimated with an FWHM of 0.1 eV and a temperature of  $T = 10$  K.

Temperature ( $T$ ) can influence the appearance of  $2p$  XAS spectra due to the thermal occupation of electronic states at higher energies. The effect is known to lead to noticeable  $2p$  XAS spectral changes in, for instance,  $\text{Co}^{2+}$  ions, but also  $\text{Ni}^{2+}$  ions [69,70]. This is expressed in Eq. (7) by the  $\beta$  factor leading to a weighted contribution to an XAS spectrum from electronic states different from the ground state. In the second ANN type (labeled as A2pX-T2), temperature and the experimental convolution by a Gaussian function with a specific FWHM are taken as an external factor and therefore included in the input vector. In addition, lifetime broadening factors are part of the output vector, allowing estimations of  $2p$  core hole lifetimes.

The third ANN type (A2pX-T2EP) can additionally predict expectation values of the  $2p$  XAS MHA, such as the spin state and the occupation of  $d$  states in the ground state configuration. This could allow avoiding computing solutions of Eq. (6). While the latter is not computationally expensive, including the EP in the ANN has to be explored for automated analysis and characterization of  $2p$  XAS spectra. It should be noted, A2pX-T2EP serves two purposes. One is to analyze a spectrum by reconstructing a representative  $2p$  XAS model Hamiltonian. Here, the ANN is said to *analyze* a spectrum in this case. The other purpose is obtaining the same ground-state properties (or expectation values) as a solution to the former Hamiltonian. Then, the ANN *predicts* parts of the solutions of the model Hamiltonian, such as the spin state and occupation numbers.

Here, *recomputed* spectra refer to simulated spectra employing the  $2p$  XAS MAH with the parameters determined by an ANN. Any difference between recomputed and reference spectrum (and other properties) measures the impact of errors in parameter estimations on the spectrum.

For the datasets of A2pX, reference spectra are computed at 1000 data points in an energy range of 30 eV, including a 5 eV pre-edge window. Datasets containing 60 000 spectra were created with randomly generated parameters for each of the tested TM ions. The intervals of the optical parameters were  $D_q = [0.0, 0.2]$ ,  $D_q = [-0.2, 0.0]$  eV and for  $O_h$  and  $C_4$  CFs, respectively, and  $D_q = [0.0, 0.2]$ ,  $D_t = [-0.2, 0.2]$ , and  $D_s = [-0.2, 0.2]$  eV for  $D_{4h}$  CFs. The screening parameters were between 0.6 and 1.0. The features of the descriptor vector are reference spectra. The optical and screening parameters are the features of the target vector to be determined by the ANNs.

For A2pX-T2(EP), separate databases are required that contain additionally information needed for the input and output vectors. The extended databases include  $T$  from 1 to 1000 K, the FWHM of the Gaussian broadening in a range from 0.05 to 0.25 eV. The lifetime broadening factors intervals are  $[0.3, 0.7]$  eV for  $\Gamma_1$  and  $[0.0, 0.2]$  eV with  $\Gamma_2$  with  $\Gamma_2' = \Gamma_1 + \Gamma_2$ . The input vector size increases by two features, which are  $T$  and FWHM, and the output vector size increases from 5 to 15 features for A2pX-T2EP. The output vector includes lifetime broadening factors for  $L_2$  and  $L_3$ , expectation values for  $J^2$ ,  $J_z$ ,  $L^2$ , and  $S^2$ , the occupations of  $a_{1g}$ ,  $b_{1g}$ ,  $b_{2g}$ , and  $e_g$  states, besides the already introduced parameters used in the simpler ANN (A2pX) that are the screening factors  $S_1$  and  $S_2$ , and the description of the  $d$ -level splitting through  $D_q$ ,  $D_s$ , and  $D_t$ . The increasing number of features in the output vector can decrease the accuracy of the ANN due to the increased complexity of the learning problem [71]. Hence, the number of reference spectra was also increased. Databases containing 360 000 spectra were created for each of the tested TM ions, from which 10 000 spectra were randomly selected for testing. Details are given in Table I.

### D. Dataset split and training details

Before ANN training, a dataset of a TM in a CF symmetry was split into three different subsets for training, validation, and testing. Besides, all features of the input and output vectors, i.e., parameters and spectra, were normalized in  $[0, 1]$  intervals (from here on referred to as box normalization). For A2pX, for instance, the first 2000 spectra were randomly selected as test datasets. Then, the remaining data (not including the test dataset) was used to construct several subsets with sizes of  $N$  between 250 and 35 000 spectra. This allows probing the required dataset size (DSS) for ANN training. The latter subset were randomly split into training and validation datasets with a ratio of 80% to 20%, respectively. For A2pX-T2(EP),  $N$  ( $N \leq 350\,000$ ) of the remaining spectra were randomly selected and also divided into training and validation datasets with a ratio of 80:20. This follows the same approach as for the A2pX dataset splitting. An overview of the computational setup and used parameters for these cases is given in the first data column of Table I.

The validation data were not used to train the model but to monitor the training progress, including the detection of over- and underfitting by comparing the evolution of the cost function values of training and validation datasets. Furthermore, the validation datasets were employed to optimize and

TABLE I. Overview of parameters and parameter ranges used to generate the reference dataset, employed in descriptor vector and in output vector of the ANNs, as well as the sizes of the used datasets.

	A2pX	A2pX-T2	A2pX-T2EP
$T$ (K)	10	10	[1,1000]
FWHM (eV)	0.1	0.1	[0.05,0.25]
Descriptor vector	[[spectrum]]	[[T,FWHM,spectrum]]	[[T,FWHM,spectrum]]
$D_q$ (eV)	[0.0,0.2] <sup>a</sup>	[0.0,0.2]	[0.0,0.2]
$D_s$ (eV)	[-0.2,0.2]	[-0.1,1.4]	[-0.1,1.4]
$D_t$ (eV)	[-0.2,0.2]	[-0.6,0.6]	[-0.6,0.6]
$S_1$	[0.60,1.00]	[0.60,1.00]	[0.60,1.00]
$S_2$	[0.60,1.00]	[0.60,1.00]	[0.60,1.00]
$\Gamma_1$ (eV)	element specific <sup>b</sup>	[0.3,0.7]	[0.3,0.7]
$\Gamma_2 = \Gamma_2' - \Gamma_1$ (eV)	element specific <sup>c</sup>	[0.0,0.2]	[0.0,0.2]
Output vector	[[ $D_q, D_s, D_t, S_1, S_2$ ]]	[[ $D_q, D_s, D_t, S_1, S_2, \Gamma_1, \Gamma_2$ ]]	[[ $D_q, D_s, D_t, S_1, S_2, \Gamma_1, \Gamma_2, J^2, J_z, S^2, L^2, n(a_{1g}), n(b_{1g}), n(b_{2g}), n(e_g)$ ]]
Reference DSS	60 k	360 k	360 k
Tested subset sizes (DSS) $N$	[250,2.5 k,12.5 k,25 k,37.5 k]	[50 k,100 k,200 k,300 k]	[50 k,100 k,200 k,300 k]
Training dataset size	80% of DSS	80% of DSS	80% of DSS
Validation dataset size	20% of DSS	20% of DSS	20% of DSS
Test dataset size	2 k	10 k	10 k

<sup>a</sup>Except for  $C_4$  and  $O_h$  symmetry cases in which the  $D_q$  ranges were [-0.2,0.2] eV and for the CoPc case with the parameter ranges being increased (see text).

<sup>b</sup>Ni:0.48; Fe:0.36; Co:0.43; Mn:0.32 eV.

<sup>c</sup>Ni:0.52; Fe:0.37; Co:0.43; Mn:0.34 eV.

validate hyperparameters. The tested hyperparameters include the initial learning rates, choice of the neuron activation functions in each layer, as well as layer width and ANN depth. A grid search was undertaken to find optimal learning rates and a Monte Carlo search for different combinations of activation functions in the HLs that included sigmoidal functions, the exponential linear unit, and the RELU. As a result of the grid search, the RELU activation function was used in the input and each hidden layer because no other combination of activation functions yielded equally accurate or significantly improved results. The tested ANNs consisted of fully connected layers of neurons. A manual search was performed with nine selected cases with different depths and widths and ANN architectures included one, three, and six hidden layers. The number of neurons in the hidden layers is given in Table II for narrow, medium, and wide ANNs. Details on input and output layer sizes and features are given in Table I. All ANNs were implemented with the high-level neural network API Keras [72]. A stochastic gradient descent method [73] iteratively optimized the weights of the ANNs in 250 epochs in the training phase. The learning rates were reduced by a factor of 0.2 to fine tune the optimization when ten consecutive iterations

(so-called epochs) did not result in improvement of the cost function value taken as MAE between  $y_a$  and  $y_p$ . In addition, the mean squared error (MSE) and the RMSE are reported.

### III. RESULTS AND DISCUSSION

In the following, the capabilities of ANNs applied to spectra of  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  central ions are discussed. Details are presented for  $\text{Co}^{2+}$  (if not stated otherwise), while further information about other TM ions is in the Appendix.

#### A. Training, validation, and testing

Figure 1 (Fig. S1) shows the MAE (MSE) for the trained ANNs with one, three, and six hidden layers for different training dataset sizes. The figure shows the results for  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  ions in a CF with  $D_{4h}$  symmetry. MAE and MSE decrease with increasing dataset size. There is little to no improvement for ANNs with only a few neurons or for narrow ANNs. Wide ANN architectures with one hidden layer can already achieve some agreement between predicted and actual data. Comparatively, the MSE drastically increased for  $\text{Ni}^{2+}$  with only one HL at large dataset sizes. For most cases of

TABLE II. Number of neurons in the hidden layers (HL) for narrow, medium, and wide ANN architectures. Layer widths decrease with proximity to the output layer.

	Narrow	Medium	Wide
1HL	16	256	2048
3HL	16;16;8	256;256;64	2048;2048;1024
6HL	16;16;8;8;8;8	256;256;64;64;32;32	2048;2048;1024;1024;512;512

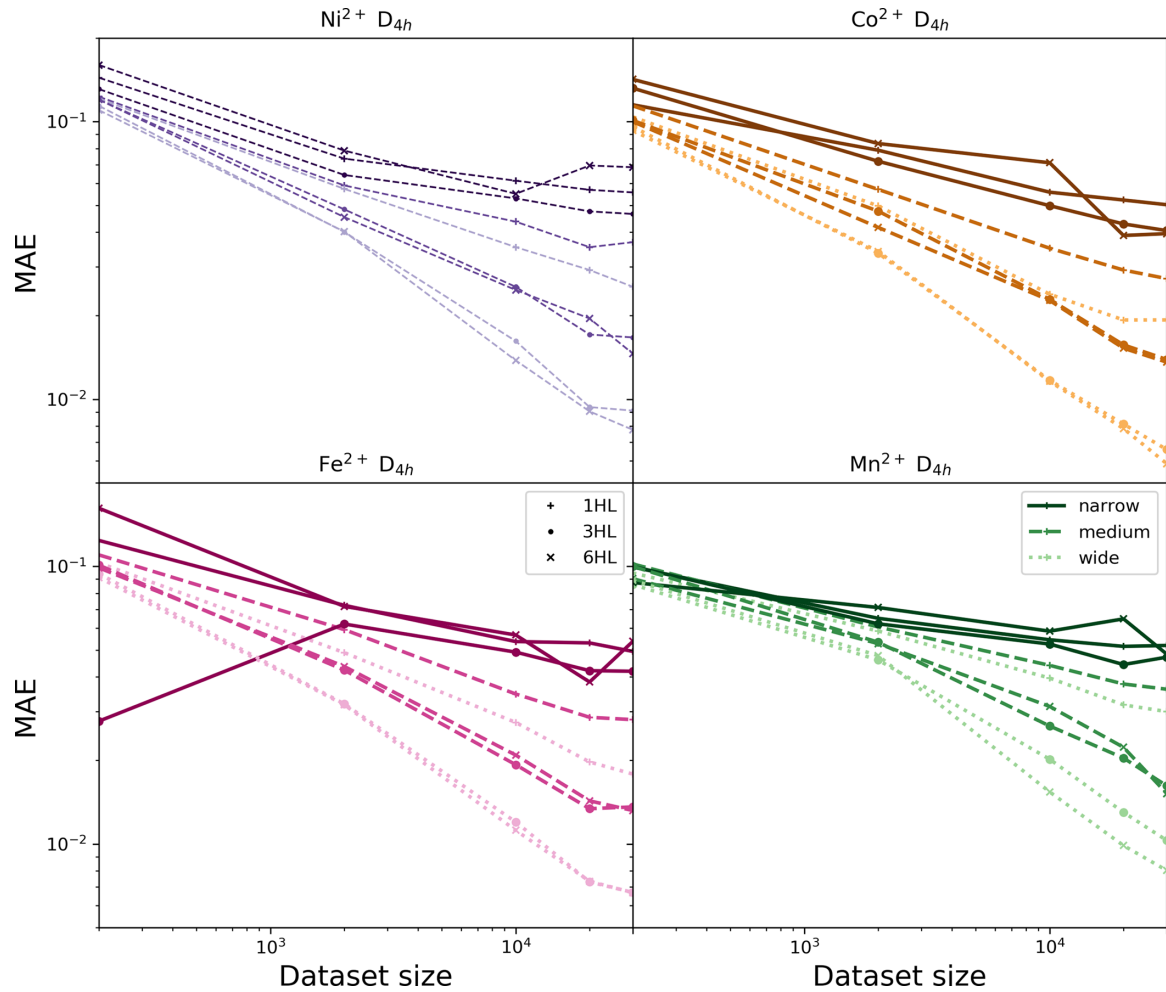


FIG. 1. The MAE of validation data on trained ANNs with different architectures with one, three, and six HLs for (from left to right)  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  in  $D_{4h}$  CF as a function of dataset size.

ANNs consisting of several hidden layers, training on datasets larger than 20 000 spectra yielded very accurate predictions of optical and screening parameters. MSEs of less than  $10^{-3}$  (MAE of less than  $10^{-2}$ ) were achieved for the medium and wide ANNs. The analysis of the trained ANNs for  $O_h$  and  $C_4$  CFs (shown in Fig. S2) results in similar trends that are: (i) often accuracy increases with increasing training dataset size and (ii) deeper ANNs with more neurons (i.e., wider) obtain smaller MSE/MAE. For  $O_h$  symmetry, the reduced parameter space causes MSE and MAE to be approximately one order of magnitude smaller at large dataset sizes. In general, the results suggest to use wide ANN with three to six HLs—the latter will be the basis for the remaining discussion.

Figure 2 shows a comparison of the actual and the predicted orbital energies (evaluated with the predicted Ballhausen parameters) as well as the screening parameters of the Slater-Condon integrals. The results are given for  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  in a  $D_{4h}$  CF. The orbital energies have a range from  $-1.8$  eV to  $2.8$  eV. For all tested elements, the ANNs accurately detect the screening parameters. However, there are noticeable differences in accuracy of orbital energies. Energy predictions of  $d_{b_{1g}}$  and  $d_{b_{2g}}$  are less accurate than for  $d_{a_{1g}}$  and  $d_{e_g}$  states. This is particularly obvious for  $d_{b_{2g}}$  in  $\text{Ni}^{2+}$  and for  $d_{b_{1g}}$  in  $\text{Mn}^{2+}$ . The effect is strongest for orbital energies close

to 0 eV, suggesting that those computed spectra are similar to each other and that the contributions from  $d_{b_{1g}}$  and  $d_{b_{2g}}$  states to these XAS spectra are not significantly different from being detected with high accuracy. On the other hand,  $d_{a_{1g}}$  and  $d_{e_g}$  are more accurately determined for all tested elements.

Table III provides MAEs and RMSEs obtained with A2pX (wide and six HL) on the test data for the TM ions in  $D_{4h}$  symmetric CFs. The MAE (less than 2%) and RMSE (less than 5%) are of similar magnitude but slightly larger than those given in Fig. 1 for the validation data demonstrating effective training and potentially useful ANNs. Large errors are in  $D_q$  and  $D_s$  (RMSEs of 1.0 to 4.5%), while the screening parameters and  $D_t$  results are more accurate (0.8 to 2.4%). A corresponding table (Table S1) for  $O_h$  symmetry is in SI. The errors for this case are smaller (less than 0.5%), reflecting a decrease in complexity by a reduced parameter space, which also benefited the training results.

For A2pX-T2 and A2pX-T2EP, the same generic hidden layer architecture, as used in A2pX containing up to six dense HL, can obtain comparable accuracy. Initial tests indicated that modifications in training, architecture design, and optimization approaches lead to similar outcomes as those seen by A2pX. The common trends are: (i) wide and deep (six HL) ANNs perform better than narrow and flat (one HL and three

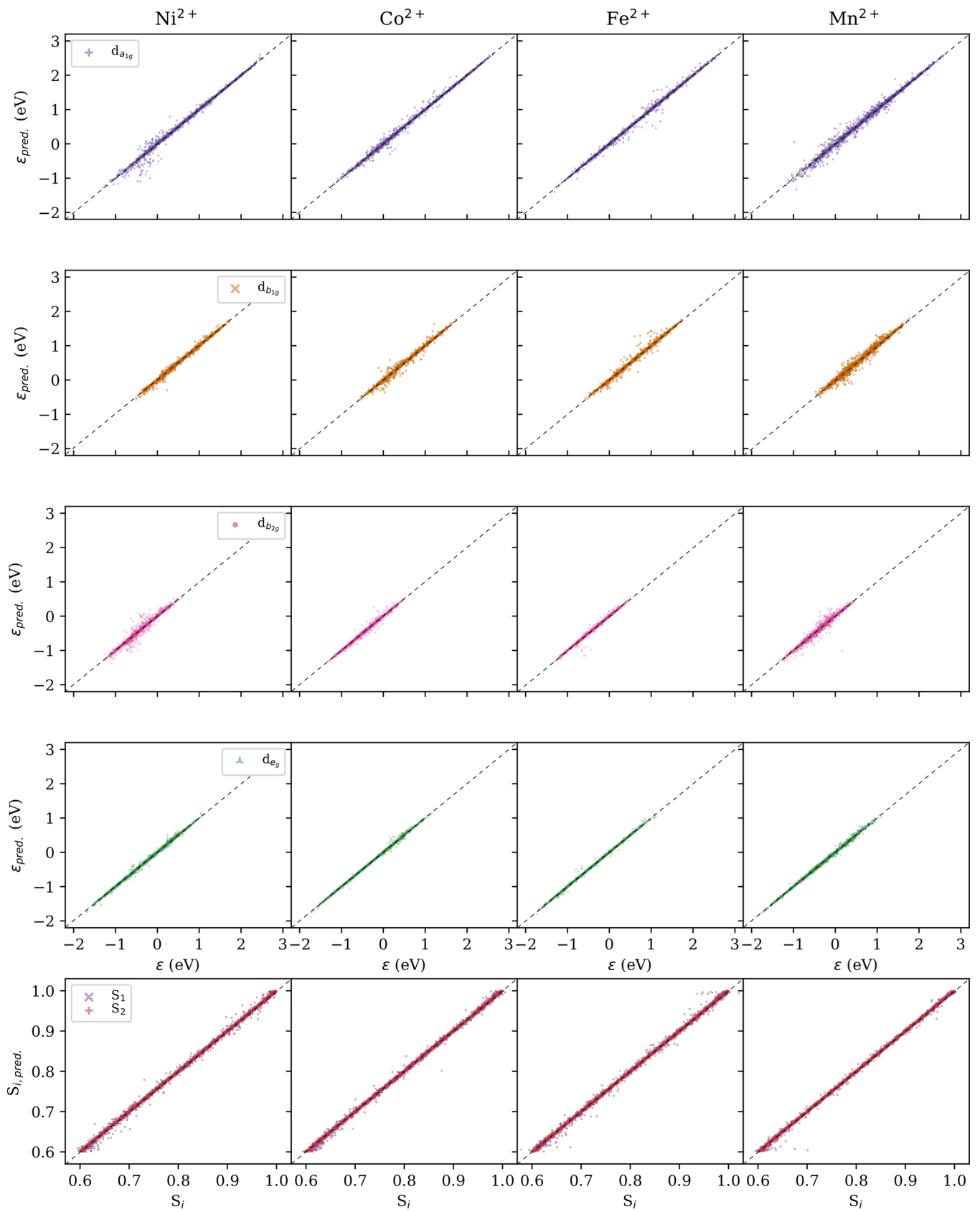


FIG. 2. From top to bottom: Actual vs predicted (pred.) relative positions of  $d$ -orbital energies  $\epsilon$  for  $d_{a1g}$ ,  $d_{b1g}$ ,  $d_{b2g}$ ,  $d_{eg}$  and screening parameters ( $S_1$  and  $S_2$ ), and from left to right:  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  in wide ANNs with six HLs. Data are computed with the validation dataset on the trained ANNs (A2pX).

TABLE III. RMSE (MAE) of normalized  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ , and  $S_2$  determined on the normalized test datasets (2000 spectra) for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX (wide six HL). Values given in percent.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	4.28 (1.96)	2.52 (1.25)	3.77 (1.86)	2.54 (1.29)
$D_s$	4.48 (1.72)	3.39 (1.44)	0.95 (0.57)	3.68 (1.53)
$D_t$	2.44 (1.20)	1.77 (0.89)	1.19 (0.75)	1.80 (0.94)
$S_1$	1.30 (0.76)	1.19 (0.75)	2.43 (1.26)	1.18 (0.77)
$S_2$	0.84 (0.61)	0.86 (0.59)	1.35 (0.87)	0.68 (0.48)

HL) architectures, (ii) RELU as the activation function for all layers provides similar or better accuracy than different combinations of other activation functions of the HL, and (iii) a small number of training data can limit the accuracy. However, an increased number of training data is required to achieve similar accuracy between A2pX and A2pX-T2(EP). The number of training data increased from ca. 5000 spectra per feature in the output vector in A2pX to more than 15 000 spectra per feature. This reflects the increased complexity the method faces when  $T$  and the FWHM are part of the descriptor vector.

Table IV gives a summary of RMSEs and MAEs for A2pX-T2. Performance details of A2pX-T2EP are in Table V, for which Fig. 3 presents corresponding details for  $\text{Co}^{2+}$  and Figs. S3, S4, and S5 in the Appendix show the data for Ni, Fe, and Mn for  $D_{4h}$  cases, respectively. For the  $\text{Co}^{2+}$  ion, the RMSE values are less than 5% for all parameters but for  $\Gamma_2$  (6.0%). The average RMSEs for parameter determination (2–3%) are comparable for A2pXT2 and A2pX-T2EP. The ANN determines  $D_t$  and  $D_s$  more accurately than  $D_q$ . Besides  $\Gamma_2$ ,  $S_1$  also has a relatively large error with ca. 2%. Regarding EP, the ANN yields RMSE of less than 2% for all expectation values but  $\langle J_z \rangle$  (2.4%). For the other ions, the trends are similar to those seen for  $\text{Co}^{2+}$ . Concerning the determination of the  $2p$  XAS parameters, the RMSEs and MAEs among the tested elements are comparable but, in general, larger for Ni and Fe than for Co. In systems with  $O_h$  symmetry, this trend is similar, but errors are generally smaller, due to reduced complexity. Details are given in Tables IV and V for  $D_{4h}$  CF and in Tables S2 and S3 for  $O_h/C_v$  CF for A2pX-T2 and Ap2X-T2EP, respectively. However, individual output features can show significantly larger deviations. Most significant RMSEs are seen in  $\Gamma_2$ , that is 5 to 9% in A2pX-T2(EP) followed by either  $\Gamma_1$  (1 to 3%) or  $S_1$  (1.7 to 2.4%) and depend on element and CF symmetry. Among Ballhausen parameters in  $D_{4h}$  systems, the largest errors are in  $D_q$ . For A2pX, the

RMSEs are between 3 and 5%, while they are 2 to 3% for A2pX-T2(EP). The RMSEs of other parameters are smaller, which confirms a relative insensitivity of computed spectra regarding small changes of  $D_q$  as pointed out by previous studies [69].

For a quantitative analysis of ANN performance with more features in the output vector, i.e., A2pX-T2 and A2pX-T2EP, an additional measure for quantifying errors in the output vector associated with spectral features is introduced. Here, the mean absolute percentage error (MAPE) of recomputed spectra is taken with respect to the normalized superposition of the recomputed and reference spectrum. The MAPE, here called the spectrum error (SE), is employed to relate the influence of feature errors in normalized output vector ( $y_p - y_a$ ), labeled as  $\Delta$ , in the corresponding spectra. SEs are evaluated for all spectra in test datasets and their corresponding recomputed spectra. Examples comparing recomputed and test spectra for A2pX-T2EP covering the whole range from smallest to largest SE are given in Figs. S9, S10, S11, and S12 for  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Mn}^{2+}$  ( $D_{4h}$ ), respectively. The figures compare test spectra (black curve) with recomputed spectra (dashed red) and are ordered according to their SE.

By visual inspection, a SE of less than 3% gives excellent agreement, while SE of 4 to 7% seems still acceptable to some degree. At larger values, differences between recomputed and reference spectra can become significant. This can be observed in, e.g., additional/missing features that appear more frequently. At SE of 10% and larger values, mismatches in the spectra and overall different spectral shapes are common. In some cases, a small shift, for instance, can cause SEs between 10 and 16%. For all tested ions, ca. 90% of the recomputed test spectra have a SE of less than 3%. The 7% threshold is reached for 97.6, 98.4, 98.4, and 97.6% of all test spectra for  $\text{Ni}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Co}^{2+}$ , respectively. The largest SE is noticeable larger for  $\text{Ni}^{2+}$  (34%) than for other elements (ca. 30%), while  $\text{Mn}^{2+}$  has a max. SE of only 24%. This

TABLE IV. RMSE (MAE) of normalized  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  determined on the normalized test datasets (10 000 spectra) for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX-T2. Values given in percent.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	2.86 (1.71)	2.76 (1.53)	1.93 (1.14)	2.69 (1.37)
$D_s$	1.11 (0.73)	1.37 (0.84)	1.59 (1.01)	3.01 (1.55)
$D_t$	0.97 (0.63)	1.06 (0.69)	1.35 (0.83)	2.37 (1.36)
$S_1$	1.83 (1.20)	2.00 (1.33)	2.43 (1.60)	2.21 (1.44)
$S_2$	1.46 (1.00)	1.48 (1.02)	1.48 (1.01)	1.19 (0.83)
$\Gamma_1$	2.83 (1.94)	2.84 (1.91)	1.95 (1.39)	1.39 (0.98)
$\Gamma_2$	8.63 (5.64)	7.96 (5.13)	6.88 (4.18)	5.77 (3.53)



TABLE V. RMSE (MAE) of normalized  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$ , and predicted the electronic configuration determined on the test datasets for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX-T2EP. Values given in percent.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	2.95 (1.71)	2.10 (1.27)	1.70 (1.00)	2.75 (1.29)
$D_s$	0.99 (0.67)	1.08 (0.70)	1.36 (0.84)	3.07 (1.52)
$D_t$	0.94 (0.60)	1.00 (0.59)	1.11 (0.67)	2.46 (1.34)
$S_1$	1.73 (1.16)	1.80 (1.19)	2.12 (1.39)	2.27 (1.40)
$S_2$	1.40 (0.93)	1.40 (0.93)	1.33 (0.89)	1.21 (0.80)
$\Gamma_1$	2.60 (1.76)	2.49 (1.65)	1.70 (1.21)	1.33 (0.92)
$\Gamma_2$	7.94 (5.09)	7.14 (4.48)	6.03 (3.61)	5.34 (3.19)
$\langle J^2 \rangle$	1.49 (0.35)	1.46 (0.39)	0.93 (0.39)	1.35 (0.34)
$\langle J_z \rangle$	4.93 (1.11)	3.14 (0.44)	2.35 (0.52)	3.05 (0.31)
$\langle S^2 \rangle$	1.49 (0.23)	1.22 (0.25)	1.97 (0.29)	1.65 (0.21)
$\langle L^2 \rangle$	1.42 (0.23)	1.17 (0.24)	1.26 (0.28)	1.57 (0.27)
$n(a_{1g})$	1.36 (0.23)	1.74 (0.33)	1.39 (0.26)	1.68 (0.19)
$n(b_{1g})$	1.83 (0.38)	1.18 (0.24)	1.30 (0.21)	1.04 (0.28)
$n(b_{2g})$	1.25 (0.26)	1.69 (0.29)	1.51 (0.35)	2.11 (0.44)
$n(e_g)$	0.80 (0.16)	0.96 (0.18)	1.25 (0.22)	1.30 (0.25)

demonstrates that for simulated  $2p$  XAS signals, the ANNs can reliably reproduce parametrizations.

The main purpose of extending A2pX-T2 to A2pX-T2EP is to estimate expectation values of electronic states such as momentum and orbital occupations. As seen above, A2pX-T2EP and A2pX-T2 obtain similarly accurate outcomes for the test spectra. Hence, both are equally capable of determining the parametrization of the  $2p$  XAS model Hamiltonian. Hence, the following discussion will focus on A2pX-T2EP.

For A2pX-T2EP, Fig. 3 shows the predicted ( $y_p$ ) and actual values ( $y_a$ ) of Ballhausen parameters, as well as screening and broadening factors at variable temperature and FWHM evaluated for the test dataset in normalized form for  $\text{Co}^{2+}$ . The distribution of SEs is given in histograms in Fig. 3 for  $\text{Co}^{2+}$  and in Figs. S3, S4, and S5 for  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$ , respectively. The latter three figures contain also the distribution of  $y_p$  and  $y_a$ . The 21 bins in each histogram are the basis for sizes and shading of data points in the other subplots. In addition, the Pearson correlation between SE and the residuals ( $\Delta$ ), given as  $c_\Delta$ , or the relative values of  $y_a$ , given as  $c_a$ , are indicated for each parameter. The correlation factors are small and reach only in a few cases, values being marginally larger than 2%: that is  $S_1$  for Fe and  $D_t$  for Mn ions. Overall, the results show acceptable accuracy for the majority of test cases. Comparing  $y_p$  and  $y_a$ , only a few SE outliers (i.e., large SE values) are present and most of them are very close to the diagonal. The largest SE in the  $\text{Co}^{2+}$  test dataset is partly due to a miscalculation of  $\Gamma_2$ . Given also these cases, it supports that an output vector with several small feature errors can cause large SEs.

Among the tested ions, the  $D_q$  distributions of actual vs predicted data points have a slightly larger number of significant outliers among Ballhausen parameters (also given by larger RMSE/MAE values). In contrast, there is a somewhat wider spread of data points near the diagonal for  $D_s$ .  $\Gamma_2$  is difficult to determine for all ANNs and shows therefore the widest distributions. The plots reveal that, in some cases, data points with large SEs are forming clusters in certain  $y_a$  ranges. Clustering in positive/negative regions of  $y_a$  correlates with

negative/positive values of  $c_a$ . At small values of  $D_q$ ,  $S_1$ , and  $\Gamma_1$  or large values of  $D_s$ , differences between recomputed and reference spectra increase for  $\text{Co}^{2+}$ . For the  $\text{Mn}^{2+}$  and  $\text{Fe}^{2+}$  ions, there are clusters of large SE in  $D_t$ . Other parameters and the results for Ni, in general, show weaker indications of SE clusters.

Note that in some cases, atomic parameters can have large  $\Delta s$ . Despite this disagreement, those data points show a reasonable resemblance between the recomputed and reference spectrum (SE of less than 7%) in many instances. This is seen, for instance, in  $D_q$  and  $D_t$  for  $\text{Ni}^{2+}$  in Fig. S3. Hence, not every single-parameter estimation error results in large SEs. In contrast, spectra with large SEs can be very close to the diagonal. This implies that reasonably accurate estimates of most parameters still can result in significant SEs for a few instances. This intrinsically limits the accuracy of the ANN-based method. In contrast, large mismatches in single or few features do not necessarily yield in large SEs. Consequently, small changes in spectra could result in noticeable  $\Delta s$  for single or a few parameters in a limited number of cases.

Figure 4 compares  $y_p$  and  $y_a$  of selected EP, including  $J^2$ ,  $J_z$ ,  $L^2$ ,  $S^2$ , and  $d$ -state occupations in a  $\text{Co}^{2+}$  ion in  $D_{4h}$  CFs. The same datapoint scale as in Fig. 3 applies. In addition, histograms of the residuals on a logarithmic count scale are overlaid. Figures S6, S7, and S8 in SI show the results for  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$ , respectively. The corresponding RMSEs and MAEs are summarized in Table V for  $D_{4h}$  and in Table S3 for  $O_h/C_{4v}$  systems. Overall, the errors are comparable but slightly larger (with the exception of the broadening factors) for EP than for analyzed parameters. This is partly caused by a few data points with very large  $\Delta s$ , which can be 50 to 100%, for a few EP features. RMSEs of less than 2% were yielded for  $\langle J^2 \rangle$ ,  $\langle L^2 \rangle$ , and  $\langle S^2 \rangle$ . Prediction of  $J_z$  expectation values shows the largest uncertainties in all cases; this includes elements as well as ligand-field symmetry. For this parameter, the RMSEs are between 2 to 5% for  $D_{4h}$  and 2 to 11% for  $O_h$  ligand fields. Values of orbital occupations, in general, show smaller RMSEs, whereas the occupation of  $t_{2g}$  and  $e_g$  states in  $\text{Ni}^{2+}$  ( $O_h$ ) reaches RMSEs of almost

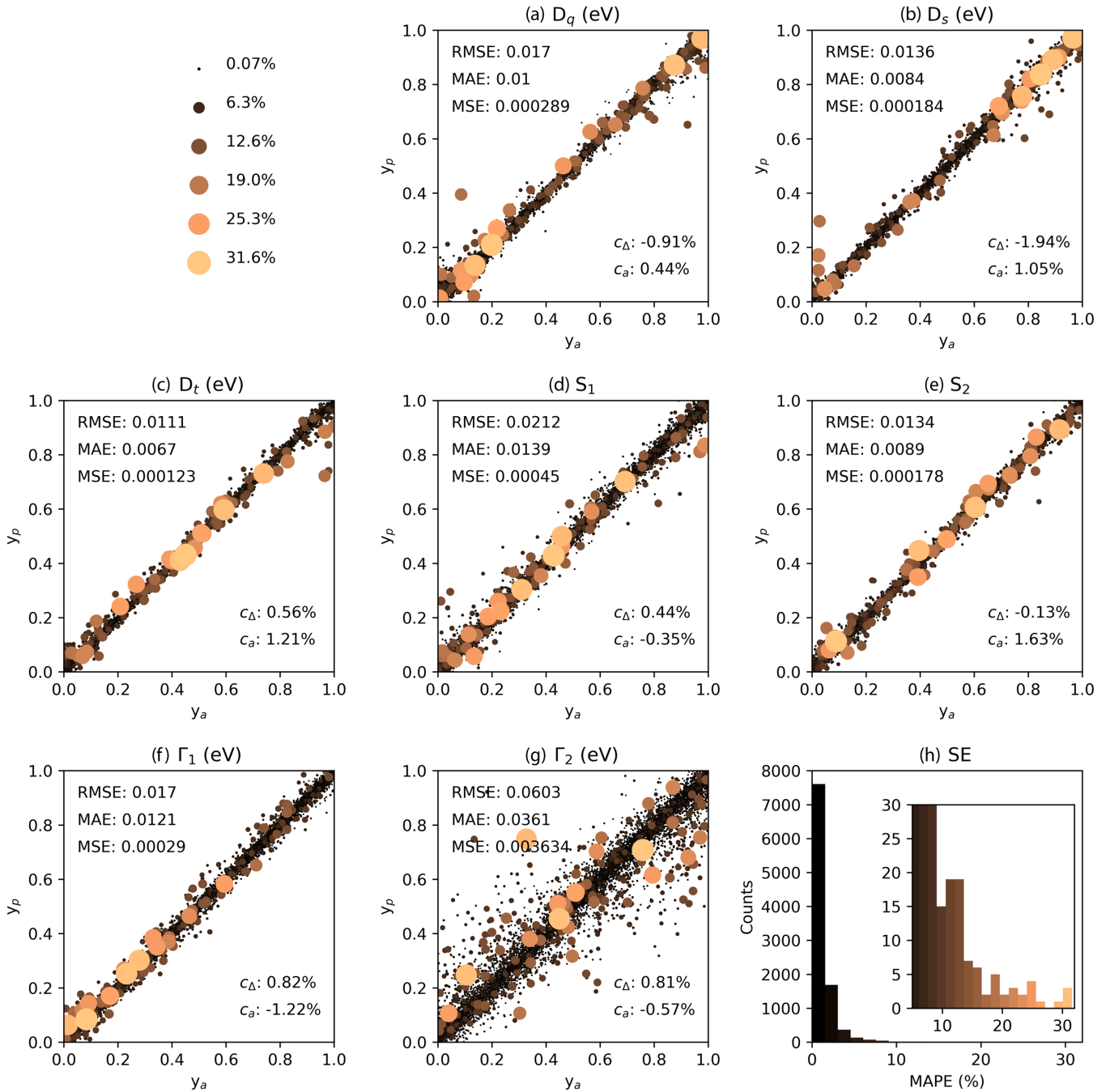


FIG. 3. From (a) to (g): Comparison of normalized  $y_p$  and  $y_a$  in  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$  and  $\Gamma_2$  for  $\text{Co}^{2+}$  with variable  $T$  and FWHM in the description vector evaluated on the test dataset (10 000 spectra). Data point sizes and shades represent the SE between recomputed and reference spectra. (h) shows the distribution of SE in a histogram in which the color shades are defined for (a) to (g). The inset in (h) shows a part of the histogram with larger resolution in the counts axis.

10%. This error is less for other elements, e.g., ca. 4% in  $\text{Co}^{2+} O_h$ , and less than 3% in  $D_{4h}$  systems. Clustering of data points with large SE is also pronounced for EP. In  $\text{Co}^{2+} D_{4h}$ , for instance,  $\langle J^2 \rangle$ ,  $\langle S^2 \rangle$ , and  $n(b_{1g})$  contain clusters for small values while  $\langle J_z \rangle$  and  $n(e_g)$  have them for large values. Large SE values have a wider spread in the expectation value prediction than for parametrization. Still, distributions in histograms show that expectation value estimation is accurate in most cases, i.e., the majority of  $\Delta s$  is in an acceptable range. However, the observed large uncertainties (close to

100%) in a few instances might limit applications for A2pX-T2EP, because there is no absolute certainty that the presented approach will yield correct electronic configurations. Interestingly, prediction of expectation values benefits from passing the spectra in the descriptor vector. In contrast, another ANN architecture (dense six HL in the wide setup) did not obtain reliable results (i.e., large RMSE and MAE) when the descriptor vector contained only the parameters from the model Hamiltonian. This included two cases; one for which the parameters were determined by A2pX-T2(EP) and the other

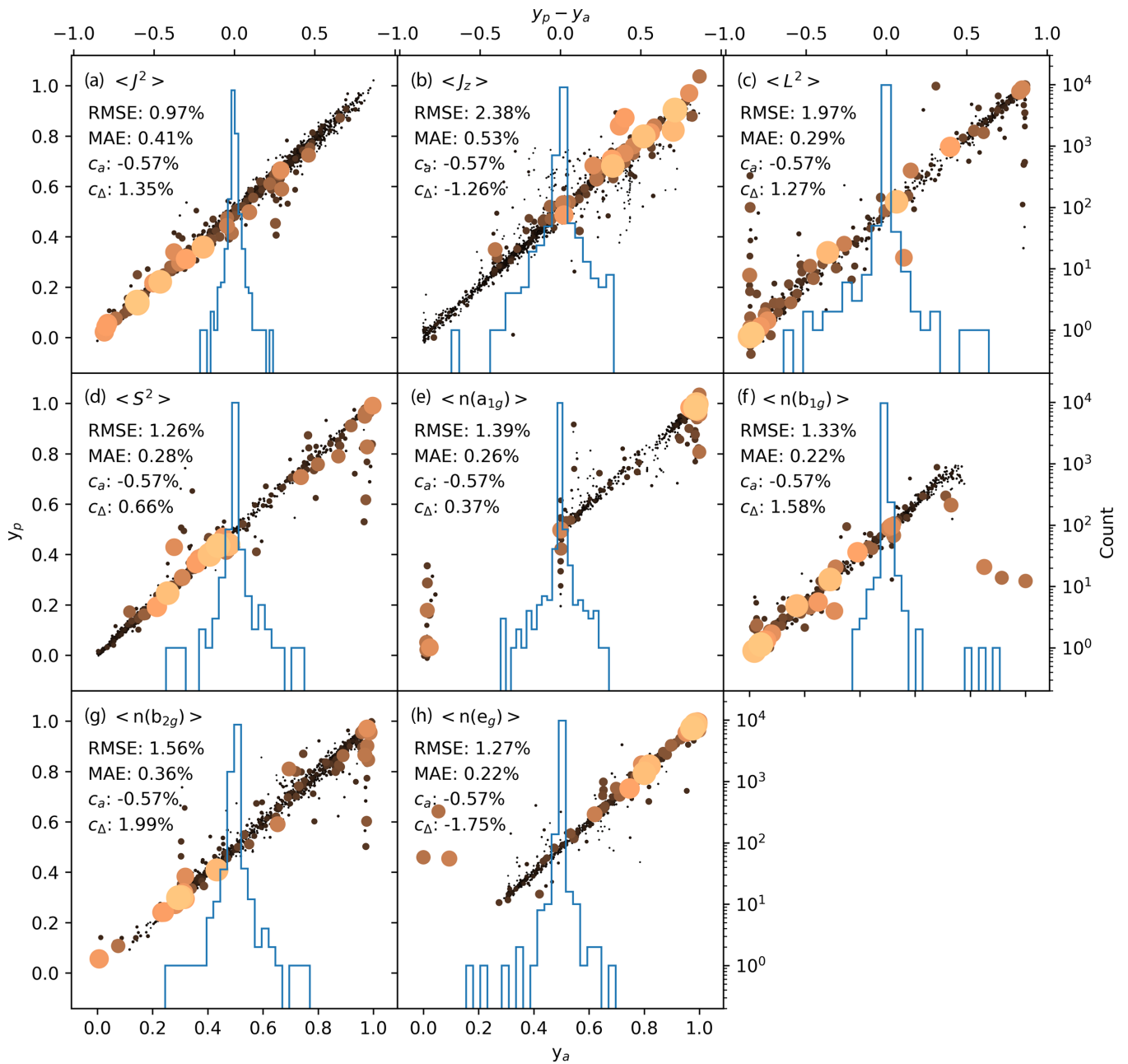


FIG. 4. Comparison of normalized  $y_p$  vs  $y_a$  for predicted properties by A2pX-T2EP for  $\text{Co}^{2+}$  determined on test dataset (10 000 spectra). The overlaid histograms show the distribution of difference between  $y_p$  vs  $y_a$  on a logarithmic count scale. Color code and data point sizes are as in Fig. 3. The scales of the histograms are on the top right; the scales of predicted ( $y_p$ ) vs actual data ( $y_a$ ) are given on the bottom left. Pearson correlation to SE for  $\Delta_y = y_p - y_a$  ( $c_{\Delta}$ ) and  $y_a$  ( $c_a$ ) are given.

one for which the parameters were taken directly from the test dataset.

In the end, the ANN-based analysis of  $2p$  XAS spectra and a simultaneous prediction of electronic configurations are obtainable to high accuracy for most spectra, while errors in electronic property prediction can be substantial in a limited number of cases. The most pronounced increase in RMSE/MAE among the different types of ANN (A2pX, A2pX-T2 and A2pX-T2EP) was observed in the screening factors and  $\Gamma_2$ . Besides, shifts in spectra can also cause large SEs. In addition, ANNs can include effects of temperature and experimental broadening as an additional input feature

to accurately estimate the parameters of the  $2p$  XAS model Hamiltonian.

### B. Impact of noise and background signal

At this point, the effect of small changes in spectra, which could also arise from noise and might lead to noticeable changes in the output vector, needs to be analyzed. In addition, background contributions are common in x-ray spectra and various mechanisms, such as solvents or secondary excitation processes and sometimes excitations to other edges, can contribute to them [69,74]. For instance, in the case of  $\text{Ni}^{2+}$ ,

the intensity around 867 eV for  $2p$  XAS of NiO is ascribed to excitation of electrons from  $2p_{3/2}$  to free states [40], see Fig. 7. Besides, intensities beyond 872 eV show an approximately constant background signal. Hence, random noise (or white noise) and background contributions in spectra must be considered on how they affect the capabilities of the ANN methods for practical applications.

The influence of white noise and background contributions in spectra on the ANN's performance is tested with different background models and varying signal-to-noise ratios (SNR) and signal-to-background ratios (SBR) in normalized test spectra. When adding white noise, the ANN output vectors obtained from a spectrum with noise ( $y_p^{\text{noise}}$ ) are compared to the output vectors obtained for the same spectra without noise. The resulting RMSEs are reported. To distinguish these RMSEs from the performance errors of the ANNs discussed above, noise-based errors, that evaluate residuals  $y_p^{\text{noise}} - y_p$ , are labeled  $\Delta\text{RMSE}$ . The latter errors were averaged over 100 trials, each having 10 000 test spectra. The SNR ratio is reported as  $20 \log [(RMS(I_{\text{XAS}})/RMS(I_{\text{noise}}))]^2$  where RMS is the root mean square of the  $2p$  XAS spectrum's intensity  $I_{\text{XAS}}$  or the noise contribution  $I_{\text{noise}}$  to it. Background intensities were modeled with a linear function. In addition, step backgrounds were modeled by two hyperbolic tangent functions, i.e.,  $0.5\beta(\tanh(\alpha(E - E_{L_i})) + 1)$ , a so-called soft-step function, centered at the  $L_3$  or  $L_2$  peaks. The factors  $\alpha$  (taken as  $2 \text{ eV}^{-1}$ ),  $\beta$ , and  $E_{L_i}$  influence the width, the max. height, and the center of the soft-step function, respectively. The background is added to the spectra of the test datasets. Then spectra with background are passed to the trained ANNs and their performance is evaluated with respect to the test parameters. The difference between parameters obtained with background and original parameters is evaluated through SBR that is computed like the SNR but using the background instead of the noise intensity. The analysis is also based on  $\Delta\text{RMSE}$ . A max. noise range/background contributions between  $1 \times 10^{-4}\%$  and 2.0% of the max. intensity of the reference spectra was considered.

It is noteworthy mentioning that the impact of noise on the accuracy of the ANNs becomes significantly stronger with a feature-wise (i.e., channel-wise) normalization instead of spectrum-wise normalization, which applies equally to Z-score scaling [75] and the here used box normalization. This is caused by large noise-induced fluctuations in low-intensity regions in spectra. Still, the feature-wise normalization can generally result in better performance of the ANNs; however, their noise sensitivity may hinder or challenge real-world application.

Figure 5 shows how random noise in test spectra affects A2pX-T2EP's capability and accuracy to determine features of the output vector for  $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$ . Data points show the mean values obtained through evaluating 100 trials each having 10 000 test spectra and error bars give their ranges (i.e., min. to max. error). The error bars are negligible for parameter estimation while they are noticeable for expectation value prediction, especially for  $\langle J^2 \rangle$ .

The effect of noise is most noticeable in  $\text{Mn}^{2+}$  while its impact decreases slightly for parameter estimation and significantly for expectation value estimation with increasing Z number. Substantial uncertainties are in  $\Gamma_2$  followed by  $\Gamma_1$  for

all tested ions. For SNR larger than ca. 4 dB, this correlates to some degree with the overall feature resolved accuracy of the ANNs, given in Table V. Estimating convolution factors,  $D_q$  and screening factors is also sensitive to random noise. For instance, for the Ni ion, the  $\Delta\text{RMSEs}$  of the Ballhausen parameters and  $S_1$  increase faster with noise than  $\Gamma_1$ . In contrast, expectation value prediction is less affected by random noise. At low SNR, the  $\Delta\text{RMSEs}$  are by a factor of ca. 4 to 10 smaller compared to the ones of the parameter determination. The largest  $\Delta\text{RMSEs}$  are seen in  $\langle J^2 \rangle$  for all ions, while  $\langle S^2 \rangle$ ,  $\langle L^2 \rangle$ , and orbital occupations are less affected among expectation values. The observed  $\Delta\text{RMSEs}$  are smaller than the RMSEs at large SNR, which should not result in severe limitation for practical applications as long as the SNR is reasonable.

Figure 6 compares the effect of SBR probed by linear background contributions on the  $\Delta\text{RMSEs}$  for all tested ions in  $D_{4h}$  CFs. The max.  $\Delta\text{RMSEs}$  are much larger (by a factor of 15 for parameter and 75 for expectation value estimation at ca. 3 dB) than for random noise. Therefore the magnitude of the  $\Delta\text{RMSEs}$  indicates severe limitations at small SBR. At large SBR, the  $\Delta\text{RMSEs}$  decrease to an acceptable size.  $\Gamma_2$  shows the largest dependence, followed by  $\Gamma_1$  and the screening factors. Only for  $\text{Ni}^{2+}$ , the  $\Delta\text{RMSE}$  of  $\Gamma_1$  is smaller than some RMSE values of the ligand-field and atomic parameters. The results for the step function backgrounds are given in the Appendix in Fig. S13. The trends follow roughly the same as for the linear background. Two points are noticeable that are: (i) the magnitude of  $\Delta\text{RMSE}$  caused by background signals is much larger than for white noise, and (ii) there is a weak correlation between feature accuracy in the test data and an increase of  $\Delta\text{RMSE}$  with an increasing background. Moreover, step backgrounds centered either at  $L_3$  or  $L_2$  lead to comparable values of  $\Delta\text{RMSE}$ . Hence, accurate background reduction can be of equal importance for  $L_3$  and  $L_2$  features with small SBR.

It can be concluded that removal of background contributions from experimental spectra is essential to increase the usefulness of the ANN-based method in analyzing XAS spectra at the  $L$  edge. In contrast, noise removal becomes only necessary for very noisy signals and for high accuracy predictions for some expectation values, while spin and orbital momenta as well as orbital occupations are not significantly affected by noise.

### C. Revisiting the $2p$ XAS of TM compounds with ANN

Furthermore, the ANNs are tested on experimental spectra. Recomputed spectra served to validate the results. The theoretical and experimental spectra were shifted in energy to facilitate comparison. For A2pX, the results for TM oxides are presented in Fig. 7. Experimental spectra are taken from Refs. [76–79] for NiO, FeO, CoO, and MnO. A2pX determined  $10D_q$  values for  $\text{Ni}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Mn}^{2+}$  ions are 1.47, 0.99, 0.74, and 0.78 eV, respectively. The values are in reasonable agreement with previous results [40,41]. The determined screening factors are between 0.63 and 0.8 (details are given in Fig. 7). The ANN yielded the exchange screening ( $S_2$ ) being less than the Coulomb screening (i.e.,  $S_2 > S_1$ ). The agreement of the computed XAS spectra with the

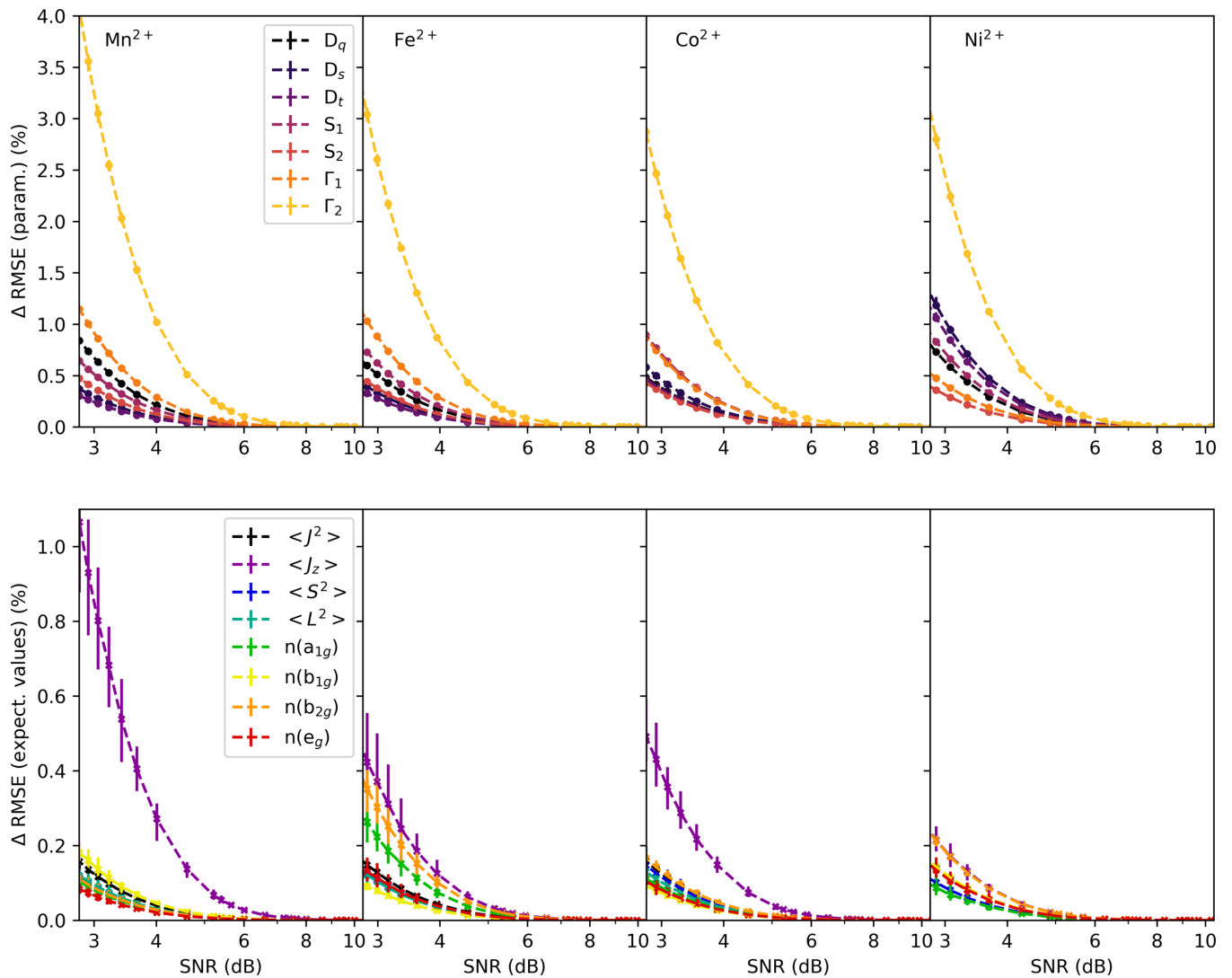


FIG. 5. Influence of noise in test spectra on the ANN's analytic (top: parameter estimation) and prediction (bottom: expectation value estimation) performance (A2pX-T2EP) evaluated as RMSE between  $y_p^{\text{noise}}$  and  $y_a$ .

experimental spectra shown in Fig. 7 is remarkable, considering the simplicity of the underlying methods. Even small features, for instance, at 780 eV in  $\text{Co}^{2+}$ , 710 eV in  $\text{Fe}^{2+}$ , and 645 eV in  $\text{Mn}^{2+}$  are qualitatively reproduced. In addition, the  $L_3$ - $L_2$  branching ratios are comparable, which confirms the ANNs accuracy. Limitations are present mostly in the peak widths caused by underestimated broadening and neglecting effects of temperature.

More challenging is the  $L$ -edge XAS of TM ions with  $D_{4h}$  symmetric CF like the ones in Co- and Fe-phthalocyanine (Pc), for which experimental spectra are taken from Refs. [29] and [80], respectively. A2pX determined parameters of the model Hamiltonian are given in Fig. 8 for the  $L$ -edge XAS of FePc and CoPc. The figure compares the theoretical (re-computed) with the experimental spectra.

For FePc, the agreement between experimental and computed spectrum is acceptable, but the same limitations (fixed temperature and underestimated broadening) as in the case of simple TM oxides seem to apply. The computed spectrum is based on the ANN analysis of the experimental spectrum. The similarity between the spectra is seen in, e.g., the tail

of the  $L_3$  peak, a small feature at 706.5 eV and the shape of the main peak. Also, the  $L_2$  peaks have an overall similar spectral shape. The electronic configuration of the  $d$  subspace determined by A2pX,  $a_{1g}^1 b_{1g}^1 b_{2g}^2 e_g^2$ , is a singlet state. It should be noted that this result is based on CMT and presents the best fit to the experimental spectrum that the ANN can produce. Previous works determined similar but also different electronic configurations in FePc [80–82], for which several authors have discussed the true ground state of FePc in recent years. At this point, the ANNs cannot exclude other electronic configurations of FePc since the results are limited to CMT.

To accurately reproduce the spectrum of CoPc, the parameter ranges and the size of the dataset needed to be increased. A dataset size of more than 100 000 spectra with ranges for  $D_q = [0.0, 0.3]$  eV,  $D_s = [-2.0, 2.0]$  eV, and  $D_t = [-1.0, 1.0]$  eV were used. This gives a similar per-feature dataset size of ca. 15 000 as the A2pX-T2EP case and results in equivalent performance of the ANN as discussed above. Then, the computed CoPc spectrum belonging to a doublet ground state ( $a_{1g}^1 b_{1g}^1 b_{2g}^1 e_g^4$ ) is reproduced and agrees with previously reported results [29].

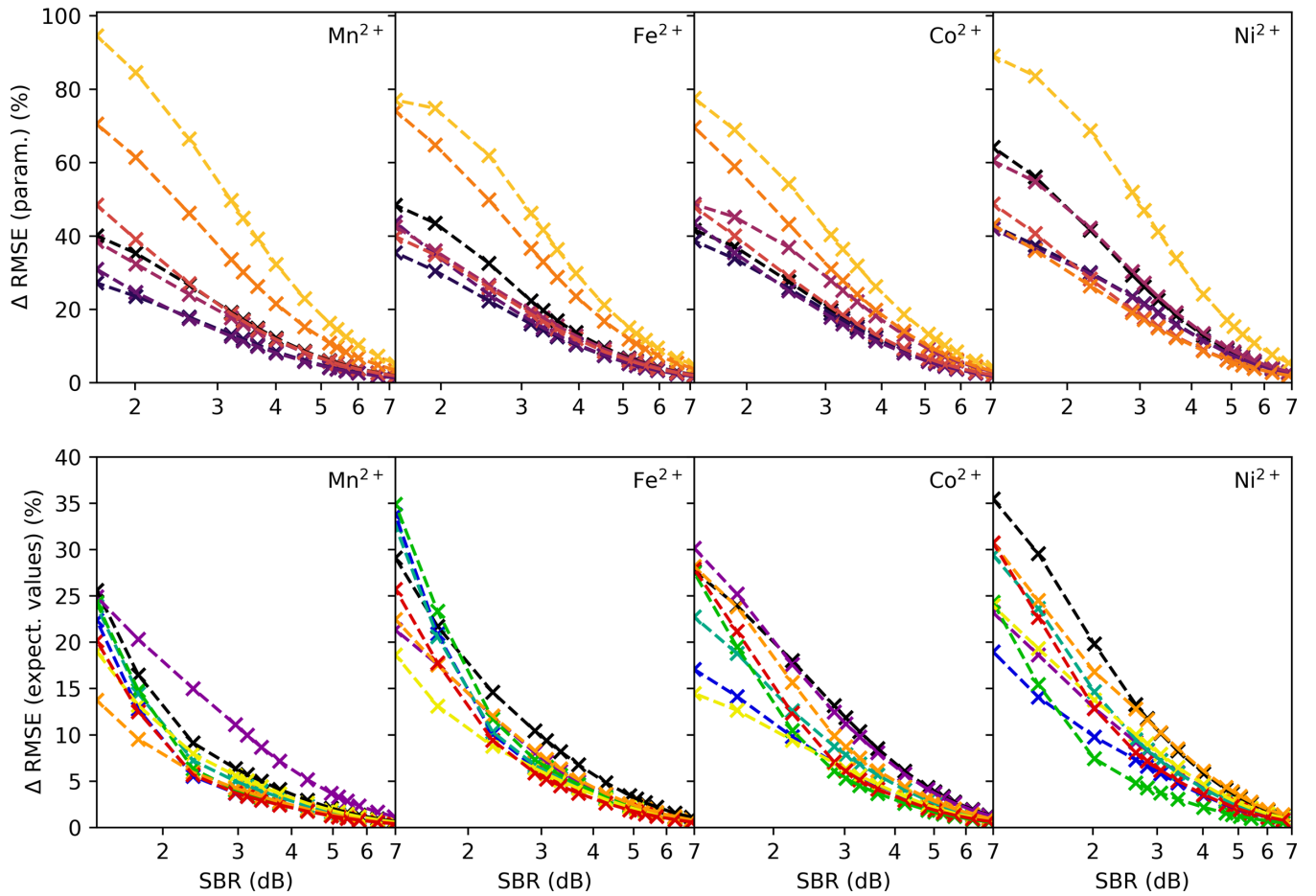


FIG. 6. Influence of linear background in test spectra on the A2pX-T2EP's analytic (parameter estimation) and predictive (expectation value estimation) performance measured in  $\Delta$ RMSE. Same curve labeling as in Fig. 5 applies.

The recomputed spectra based on the analysis given by A2pX display a few features that can be improved, such as sharp peaks of high intensity in  $L_2$  and an overall rougher appearance due to underestimated broadening effects. A2pX-T2 is tested on the already introduced experimental  $2p$  XAS spectra of  $\text{Co}^{2+}$  and  $\text{Fe}^{2+}$  ions. As indicated by the above results, A2pX-T2 will yield similarly accurate outcomes as A2pX-T2EP. Moreover,  $2p$  XAS spectra of  $\text{Co}^{2+}$  at the  $L$  edge are known to display noticeable temperature depen-

dence [69,70]. Applying A2pX-T2 should improve on these issues. Still, it must be understood that background contributions can severely affect the outcome. The discussion is based on resemblance between recomputed spectra and experimental spectra. Only  $2p$ - $3d$  transitions that contribute to the experimental spectrum, as described by Eq. (7), must be present for optimal performance of the ANNs. Therefore, other contributions such as background intensities have to be removed. Still, the description itself is incomplete given,

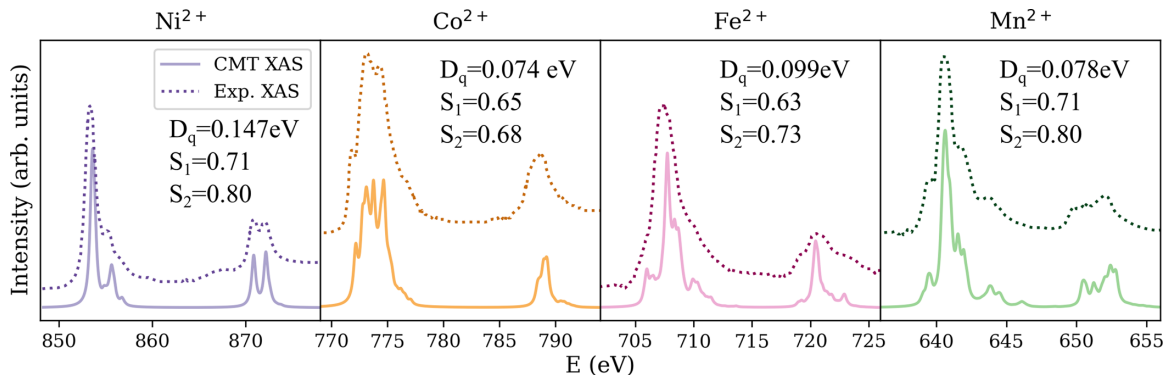


FIG. 7. Comparison of experimental  $L$ -edge XAS spectra (dotted) of NiO [76], CoO [77], FeO [78], and MnO [79] and the simulated spectra (solid) based on the analysis of the trained six HLs medium width ANN.

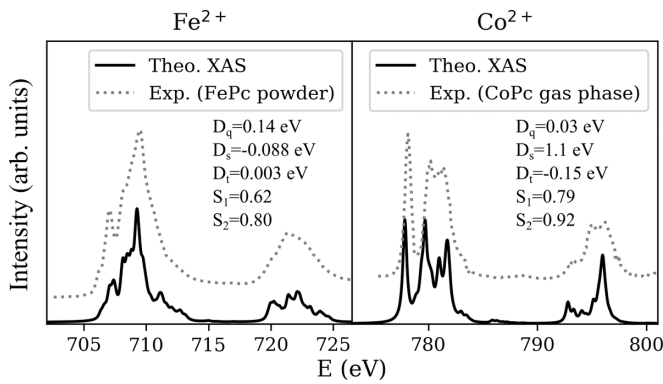


FIG. 8. Comparison of experimental  $L$ -edge XAS spectra (dotted) of FePc (powder) [80] and CoPc (gas phase) [29] with the simulated spectra (solid) that are based on the evaluation using the six HLLs medium width ANN.

for instance, by missing charge transfer (CT) effects (see above discussion). Overcompensation of the ANNs for missing physical effects (e.g., the earlier mentioned CT) may occur.

For background subtraction in the  $2p$  XAS spectrum of FePc, one step function was centered at  $L_3$  at ca. 710 eV and the height ( $\beta$ ) was chosen to approx. level the intensity before  $L_2$  at 0 and the other function was center at  $L_2$  at around 722.3 eV with a height that levels the intensity beyond 728 eV at 0. No noise reduction was applied. The temperature was set to 8 K, at which the experiment was performed [80], and the experimental FWHM equaled 0.1 eV.

A2pX-T2 determined  $D_q = 0.13$  eV,  $D_s = 0.35$  eV,  $D_t = 0.037$  eV,  $S_1 = 0.62$ , and  $S_2 = 0.86$ . The lifetime broadening factors are  $\Gamma_1 = 0.59$  eV and  $\Gamma_2 = 0.76$  eV. Compared to the A2pX results,  $D_s$  experiences a large change, while the other atomic and ligand field parameters are similar to the ones determined by A2pX. In an earlier study, Miedema *et al.* performed the analysis of  $2p$  XAS of FePc using planar symmetry that relates values of  $D_s$  and  $D_t$  to the value of  $D_q$ . Concerning the formal rule for planar symmetry and  $D_q$ , the A2pX-T2 determined values are mismatched,  $D_s$  by a factor of 2 and  $D_t$  by a factor of 0.5. Note that chemical environments can lead to a reduction of symmetry in FePc. Previously, Snezhkova *et al.* detected symmetry breaking in FePc through interaction of the molecule with its chemical surrounding, i.e., when adsorbed on Cu(111) surface, by XPS measurements and first-principle calculations [83]. While a mismatch of parameters, especially for small values, cannot be excluded, a reduced symmetry may at least partly contribute to the yielded A2pX-T2 outcome.

Figure 9 shows the recomputed  $2p$  XAS spectrum and the experimental spectrum of FePc. The agreement between recomputed and experimental spectra is satisfactory, which demonstrates that A2pX-T2 performed well for this case. Compared to the recomputed spectrum employing the output vector of A2pX, the agreement in the spectral shape of  $L_2$  improved. The A2pX-T2-based recomputed spectrum reproduces the steep increase between 720 to 722 eV followed by a gradual decrease between 722 and 725 eV. The similarity in  $L_3$  did somewhat increase. The pre-peak at around 707 eV, the

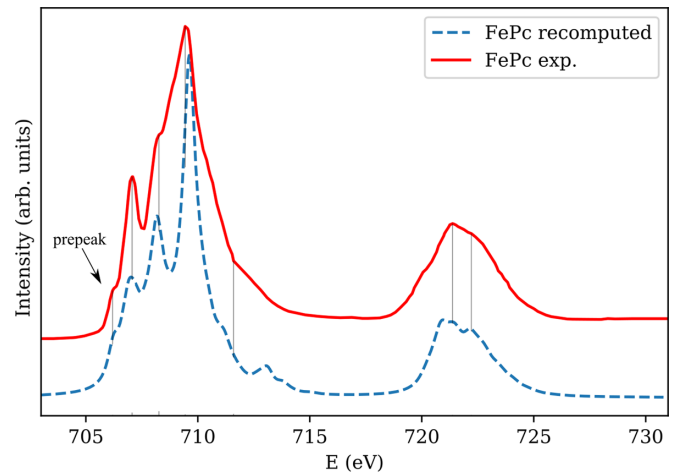


FIG. 9. Comparison of experimental (red) [80] and recomputed spectra (dashed blue) based on the results of A2pX-T2 for FePc. Vertical lines are given for reference of spectral feature positions.

slight shoulder at ca. 712.5 eV, and the main peak at 710 eV (although the latter peak is slightly shifted to higher energies and overestimated in intensity) are still present. A peak at 711 eV is overestimated by a bit. Moreover, a small peak at 713 eV appeared.

For FeO, the temperature was set to 848 K, a temperature at which FeO is a stable phase, and an experimental FWHM of 0.1 eV is assumed. Note that there was little difference in the recomputed spectrum for input vectors with temperatures at 400 K and 848 K passed to the ANN. In contrast, the results show larger differences if low temperatures, e.g. 10 K, are used. A subtraction of the stepped background was applied with two soft-step functions. One step function was centered at  $L_3$  at ca. 713 eV and the height ( $\beta$ ) was chosen to slightly lower intensity before  $L_2$  and the other function was center at  $L_2$  at around 725.3 eV with a height that levels the intensity beyond 730 eV closer to 0. Also, the signal was truncated (i.e., set to 0) before 706 eV. No noise reduction was applied.

$D_q$  was determined as 0.022 eV, the screening parameters as  $S_1 = 0.90$  and  $S_2 = 0.92$ , and the lifetime broadening factors are 0.77 ( $\Gamma_1$ ) and 1.03 eV ( $\Gamma_2$ ). The determined value of  $D_q$  for TMO is underestimated with regard to previous reports [69] and the A2pX results, which usually give values around 0.1 eV (or 1 eV for  $10D_q$ ). It is well known that there can be a certain insensitivity towards  $D_q$  values in  $2p$  XAS spectra, especially for TMOs, allowing a range of values to reproduce experimental results. Here, the experimental and recomputed spectra are similar. This potentially points towards a more delicate relation between ligand field parameters, screening factors, and lifetime broadening (potentially also for  $D_{4h}$  systems) in the underlying MHA as commonly assumed by fixed values of screening factors of 0.8 or restrictions such as  $S_1 = S_2$ .

Figure 10 shows the comparison between experimental and recomputed (based on A2pX-T2 results)  $2p$  XAS spectra of FeO. There is a noticeable improvement for both  $L_3$  and  $L_2$  peaks in the recomputed  $2p$  XAS spectrum compared to the results in Fig. 7. The overestimation of the main peaks in  $L_3$  and  $L_2$  was reduced. Similarly, the pre-slope starting at

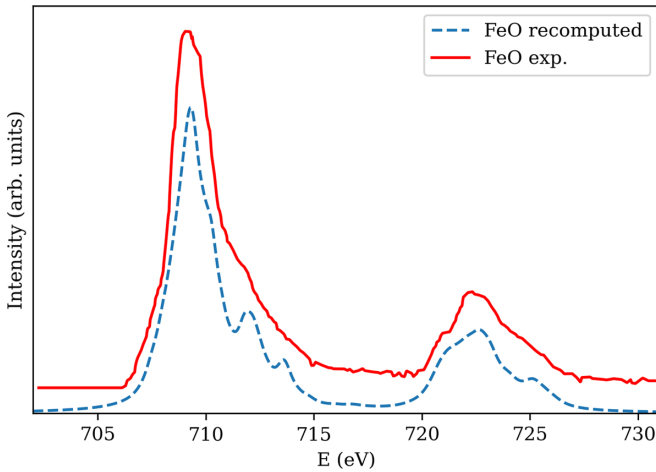


FIG. 10. Comparison of experimental (red) [78] and recomputed spectra (dashed blue) based on the results of A2pX-T2 for FeO.

706 eV and the width of the main peak are well matched in  $L_3$ . The shoulders at ca. 711 and 714 eV are somewhat over pronounced but follow the tail of  $L_3$  in their intensity ratios. Especially the latter peak at 714 eV is sensitive to the value of  $D_q$ . At larger  $D_q$  values, the peak would shift to higher energies which would diminish the agreement for the given parameter set. At ca. 721 eV, the close-by main peak and the tail are well represented in  $L_2$ .

Background subtraction in the  $2p$  XAS spectra of CoPc and CoO was performed with soft step functions to align the onsets of the  $L_3$  and  $L_2$  peaks at 0 and to remove background contributions beyond  $L_2$  transition intensities. For CoPc, A2pX-T2 determined the parameters  $D_q = 0.0043$  eV,  $D_s = 1.00$  eV,  $D_t = -0.13$  eV,  $S_1 = 0.91$ ,  $S_2 = 0.79$ ,  $\Gamma_1 = 0.44$  eV, and  $\Gamma_2 = 0.75$  eV at a FWHM of 0.2 eV and the temperature set to 708 K at which the experiment was performed. For CoO, A2pX-T2 determined the parameters as  $D_q = 0.0083$  eV,  $S_1 = 0.82$ ,  $S_2 = 0.81$ ,  $\Gamma_1 = 0.68$  eV, and  $\Gamma_2 = 0.73$  eV at a set FWHM of 0.2 eV and a set temperature of 303 K.

The recomputed spectrum of CoPc based on the A2pX-T2 analysis is given in Fig. 11. The enhanced agreement to the experimental gas-phase spectrum, compared to the one given in Fig. 8, appears to be mostly due to an increased broadening of transition intensities and increased temperature. Other minor changes in intensity distribution result from changes in other parameters. The first peak at 778.2 eV and the following two peaks still match well in their max. intensity and energy spacing. Even a shoulder at 783.5 eV is reproduced, while  $L_2$  features have less agreement, in particular, at 793.5 eV and the peak structure between 795 and 797 eV. Figure 12 shows the recomputed spectrum of CoO, which also displays a strong resemblance to the experimental spectrum. The main features of  $L_3$  and  $L_2$  are captured, including smaller shoulders at ca. 772.5, 777, and 792 eV.

Although the agreement between the experimental and simulated spectra of the selected test cases is encouraging, the limits of the applied method should be communicated. There is a dependence on the database and its parameter ranges, e.g., the A2pX case of CoPc, reflecting the well-known extrapo-

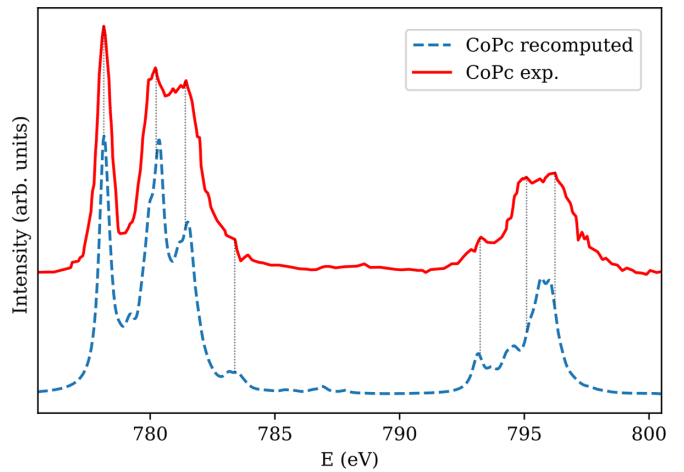


FIG. 11. Comparison of experimental (red) [29] and recomputed spectra (dashed blue) based on the results of A2pX-T2 for CoPc. Vertical lines are given for reference of spectral feature positions.

lation weakness of ANNs. Another factor is the underlying theory used to construct the database. The presented ANN approach is based on CMT and the ANNs can only gain insights that any CMT-based XAS method would have produced, in particular, charge transfer effects are neglected, which may cause the ANN to compensate for this effect.

While the extensions of A2pX, including temperature and FWHM, increase the similarity between experimental and computed spectra, agreement of the MHA parametrization with those of previously published results is somewhat limited through potential influence of, for instance, the lack of charge transfer effects in the approach or the alignment between the experimental and computed spectra. At the same time, ANNs still demonstrate tremendous capabilities and appear to be promising tools to determine the parametrization of a  $2p$  XAS model Hamiltonian and potentially its expectation values of angular and spin momenta as well as  $d$ -orbital occupation from spectra without or little noise and background contributions. The here presented analysis and application

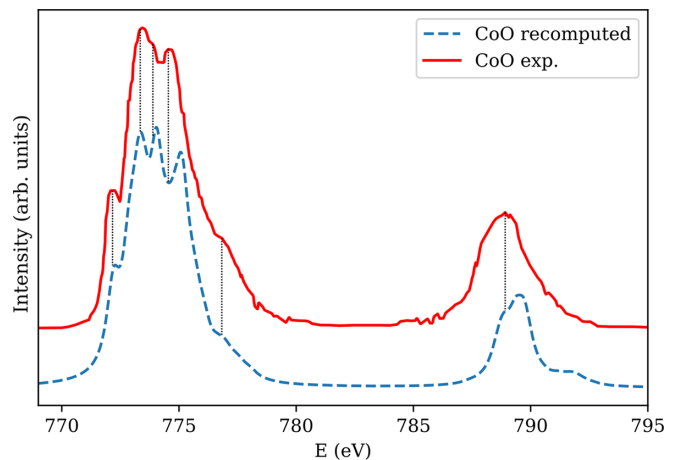


FIG. 12. Comparison of experimental (red) [77] and recomputed spectra (dashed blue) based on the results of A2pX-T2 for CoO. Vertical lines are given for reference of spectral feature positions.



to experimental  $2p$  XAS spectra showed that practical applications, however, can be limited by noise and especially background intensities in spectra, at least for dense ANN and the chosen data processing strategy. This is neither surprising nor unexpected as post-processing (e.g., background and noise reduction) is still a routine part of the evaluation process for experimental  $2p$  XAS.

The proposed ANN-based approach points out the need for improvements and continued developments that are required for routine application of (dense) ANN in  $2p$  XAS analysis; that can be an automated background subtraction that, in principle, can also be ML supported and/or a more robust approach for expectation value prediction. For these points, the normalization process and data preprocessing can be, to some extent, essential as the Z-score and box normalization indicated for noise-related errors (see above). Here, noise reduction seems to be less significant for box normalization compared to background intensity removal. Nonetheless, the ANN-based evaluation of experimental spectra and the consecutively performed recomputation of spectra were successfully performed. The obtained results are particularly encouraging for the metal-organic compounds with reduced symmetries, which demonstrates the capability to accelerate analysis for a wide variety of materials in the near future. Concerning  $L$ -edge XAS of metal-organic compounds containing TM ions, the method could provide an unbiased tool to obtain experimental electronic ground state configurations and gain insights on how different chemical environments influence them. Hence, the question if ANNs can represent the inversion function of conventional computations of  $2p$  XAS spectra of light transition metal ions is affirmed.

#### IV. CONCLUSIONS

In summary, artificial neural networks are presented that can estimate relative energies of  $3d$  levels, screening factors of the Coulomb and exchange interaction, and core hole lifetimes directly from experimental  $L$ -edge spectra of transition metal compounds. The method is based on dense neural network architectures with six hidden layers, which work sufficiently well for various transition metal ions. The artificial neural networks were trained and optimized on theoretical  $2p$  XAS spectra of transition metal ion employing a model Hamiltonian approach that is based on crystal-field multiplet theory. In general, obtained results become less accurate when more information (i.e., number of parameters) is to be extracted from a spectrum. This can partly be compensated by larger training dataset sizes and larger neural networks. In addition, temperature and an experimental convolution factor can be included as input features of the neural network to account for these effects in spectra. Then, simulated spectra employing the ANN results can obtain a very high degree of similarity to experimental spectra if proper background signal removal is applied. In combination with a  $2p$  XAS model Hamiltonian, the gained results can be used to determine physical properties such as the ground state electronic configuration and  $d$ -level occupations from experimental spectra in one-shot evaluations without the need for prior first-principles electronic structure calculations.

#### ACKNOWLEDGMENTS

J.L. is thankful for financial support by the Ministry of Science and Technology (MOST) Taiwan under Grant No. MOST109-2112-M-110-009 and acknowledges the National Center for High-Performance Computing for computer time and facilities.

#### APPENDIX: ADDITIONAL DETAILS ON ANN PERFORMANCE

In the following we present further details about the performance of the developed ANNs used to evaluate  $2p$  XAS spectra and predict results of a  $2p$  XAS model Hamiltonian as presented in the main paper.

##### 1. Optimizing ANNs

The dependence on training performance on dataset size is shown in Fig. S1. The impact of training dataset size on the MSE and MAE in validation data for  $O_h$  systems is shown in Fig. S2. Tables S1, S2, and S3 summarize the RMSE and MAE evaluated on the test datasets and obtained with A2pX, A2pX-T2, A2pX-T2EP, respectively, for  $O_h$  systems.

##### 2. Analysis of spectra and prediction of electronic states

Figure S3 shows the actual ( $y_a$ ) vs the analyzed ( $y_p$ ) values for  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  for  $Ni^{2+}$  obtained from A2pX-T2EP, i.e., the results are based on the dataset including variable  $T$  and FWHM as additional features in the input vector. Likewise, Figs. S4 and S5 show the results for  $Fe^{2+}$  and  $Mn^{2+}$ , respectively. Moreover, A2pX-T2EP simultaneously predicts expectation values of  $J^2$ ,  $J_z$ ,  $L^2$ ,  $S^2$  and the occupations of  $d$  states that are shown in Fig. S6 for  $Ni^{2+}$ , Fig. S7 for  $Fe^{2+}$ , and Fig. S8 for  $Mn^{2+}$ . The corresponding figures for  $Co^{2+}$  are given in the main paper. Table IV in the main paper summarizes the accuracy of A2pX-T2EP for these elements.

##### 3. Comparison between Reference Spectra and Recomputed Spectra

Figures S9 to S12 show comparisons between reference spectra (solid black) used to test the performance of the A2pX-T2EP and the corresponding recomputed spectra (red dashed) for  $Ni^{2+}$ ,  $Co^{2+}$ ,  $Fe^{2+}$ , and  $Mn^{2+}$ , respectively. First, the ANNs determined the parameters from the reference spectra. Then, the parameters were used in Eqs. (3) and (7) to yield the recomputed spectrum. This was done for all spectra in the test datasets. The spectrum errors (SE—defined in the main text) were computed and a selection of them spanning the whole range of SEs is given in the figures. SE (%) and the ranking position from best ( $/0$ ) to worst ( $/9999$ ) are indicated as well.

##### 4. Influence of Noise

Figure S13 shows the impact of different step backgrounds on the ANN results. It displays the impact of a step background on the  $\Delta$ RMSE in parameter and expectation value estimation.

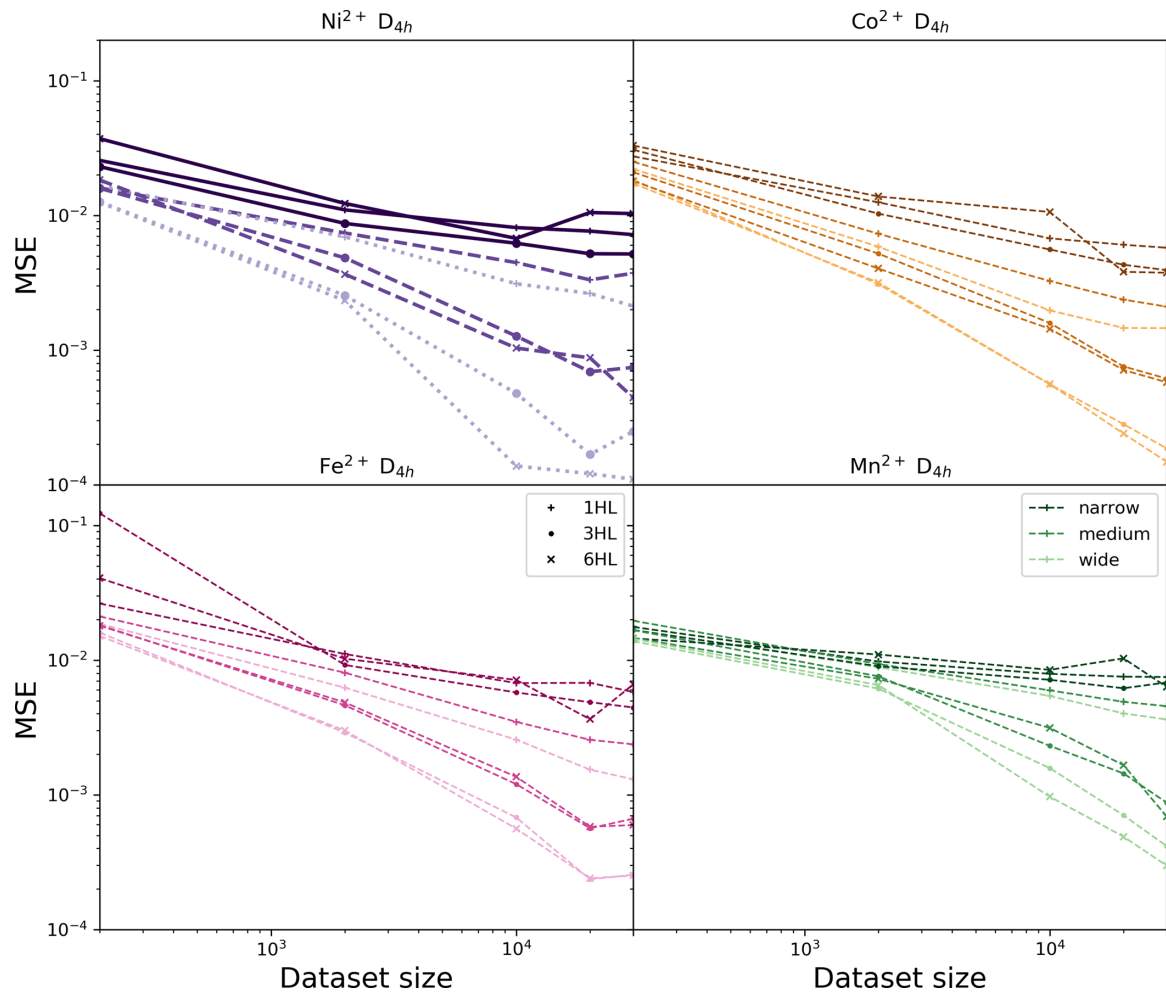


FIG. S1. The MSE of validation data on trained ANNs with different architectures with one, three, and six HLs for (from left to right)  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  in  $D_{4h}$  CF as a function of dataset size.

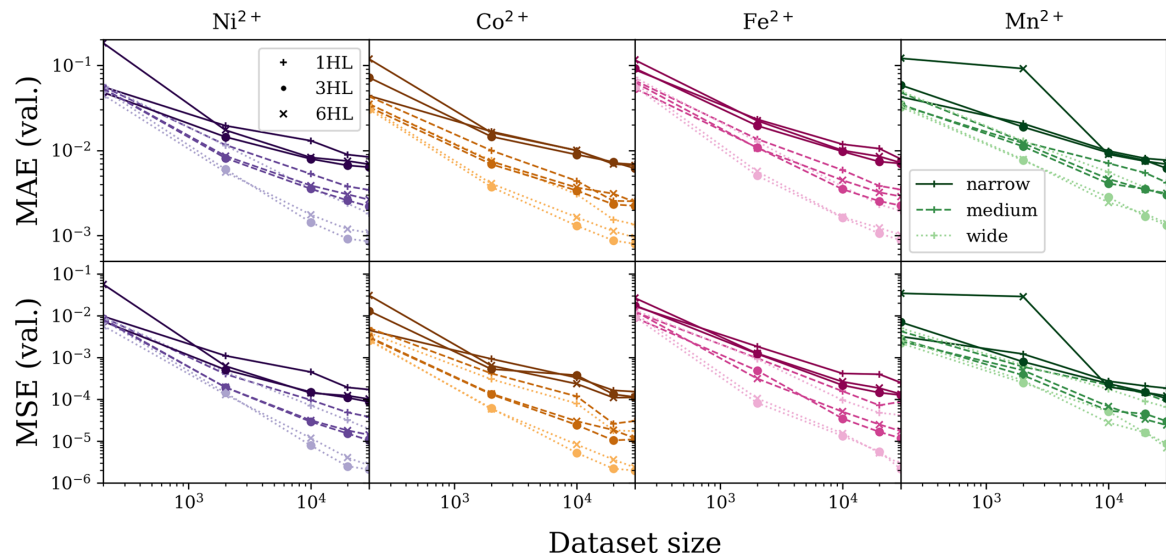


FIG. S2. The MSE and MAE of validation data on trained ANNs with different architectures with one, three, and six HLs (A2pX) for (from left to right)  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Fe}^{2+}$ , and  $\text{Mn}^{2+}$  in  $O_h$  and  $C_{4v}$  CFs as a function of dataset size.

TABLE S1. RMSE (MAE) of normalized  $D_q$ ,  $S_1$ , and  $S_2$ , determined on the normalized test datasets (2000 spectra) for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX for  $O_h/C_{4v}$  systems. Values are given in %.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	0.44 (0.24)	0.16 (0.12)	0.16 (0.12)	0.16 (0.12)
$S_1$	0.36 (0.24)	0.19 (0.13)	0.18 (0.13)	0.20 (0.13)
$S_2$	0.30 (0.21)	0.17 (0.12)	0.16 (0.12)	0.15 (0.11)

TABLE S2. RMSE (MAE) of normalized  $D_q$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  determined on the normalized test datasets (10 000 spectra) for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX-T2 for  $O_h$  and  $C_{4v}$  systems. Values are given in %.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	1.20 (0.52)	0.88 (0.66)	0.76 (0.47)	0.91 (0.54)
$S_1$	0.43 (0.33)	0.47 (0.36)	0.65 (0.49)	0.77 (0.55)
$S_2$	0.40 (0.30)	0.41 (0.32)	0.52 (0.40)	0.60 (0.44)
$\Gamma_1$	0.60 (0.43)	0.64 (0.49)	0.82 (0.62)	0.72 (0.53)
$\Gamma_2$	2.15 (1.38)	1.21 (0.88)	1.76 (1.21)	3.00 (1.87)

TABLE S3. RMSE (MAE) of normalized  $D_q$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ ,  $\Gamma_2$  predicted the electronic configuration determined on the normalized test datasets (10000 spectra) for  $\text{Mn}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Ni}^{2+}$  with A2pX-T2EP for  $O_h$  and  $C_{4v}$  systems. Values are given in %.

	$\text{Mn}^{2+}$	$\text{Fe}^{2+}$	$\text{Co}^{2+}$	$\text{Ni}^{2+}$
$D_q$	1.17 (0.57)	0.52 (0.30)	0.40 (0.25)	0.71 (0.35)
$S_1$	0.42 (0.31)	0.76 (0.50)	0.70 (0.52)	0.75 (0.54)
$S_2$	0.54 (0.41)	0.54 (0.39)	0.54 (0.39)	0.63 (0.45)
$\Gamma_1$	0.88 (0.64)	0.90 (1.53)	0.90 (0.62)	0.76 (0.56)
$\Gamma_2$	2.60 (1.56)	2.63 (1.53)	2.63 (1.53)	3.64 (2.21)
$\langle J^2 \rangle$	0.09 (0.01)	1.13 (0.29)	1.21 (0.34)	1.26 (0.42)
$\langle J_z \rangle$	2.53 (0.41)	10.55 (2.04)	5.76 (0.98)	7.81 (1.27)
$\langle S^2 \rangle$	0.09 (0.01)	1.81 (0.13)	0.81 (0.04)	1.29 (1.06)
$\langle L^2 \rangle$	0.09 (0.01)	1.79 (0.13)	0.79 (0.07)	0.60 (0.28)
$n(t_{2g})$	0.09 (0.02)	2.63 (0.55)	3.81 (0.68)	8.45 (1.37)
$n(e_g)$	0.09 (0.02)	2.40 (0.51)	3.64 (0.67)	8.08 (1.31)

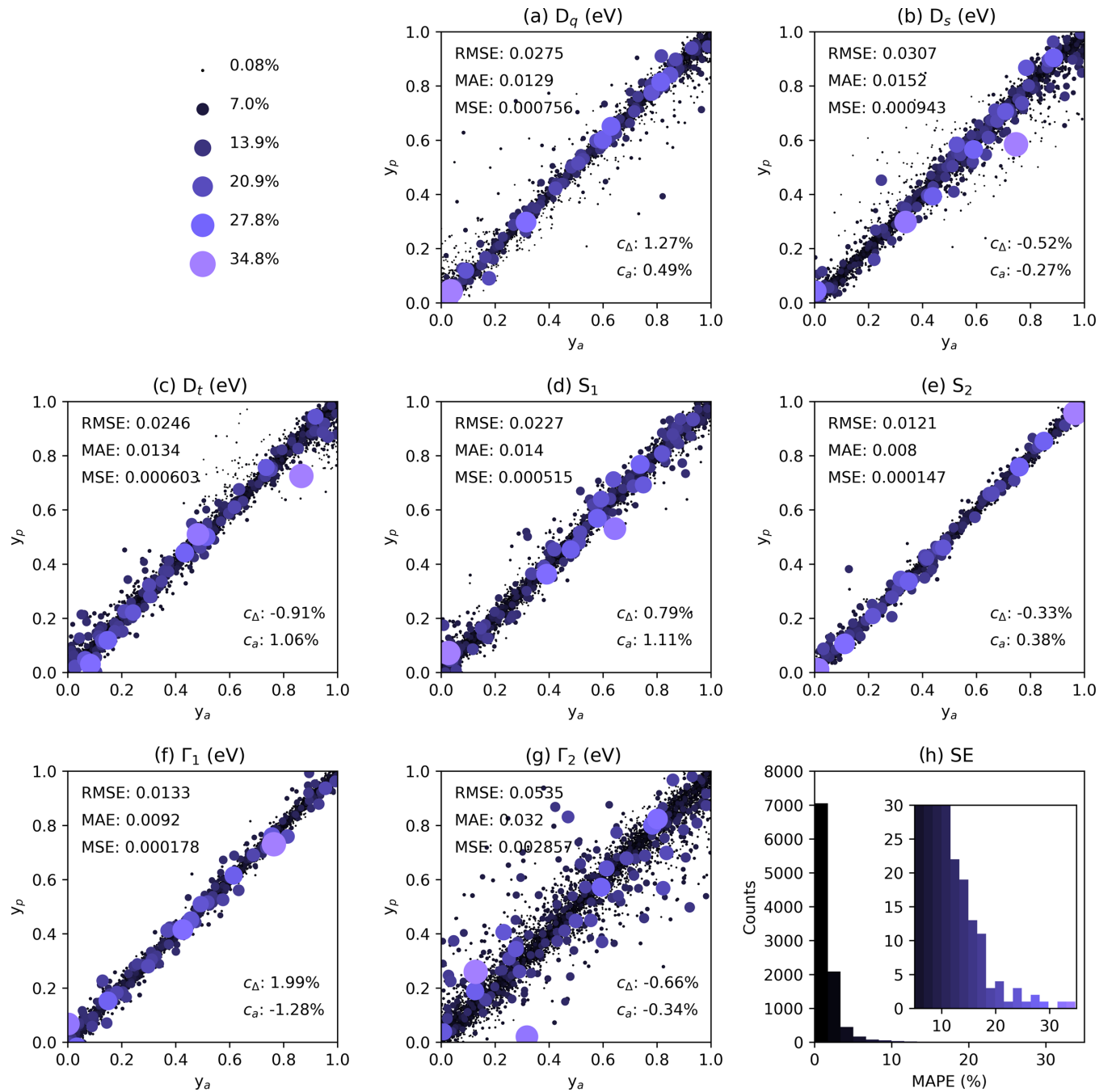


FIG. S3. From (a) to (g): Comparison of normalized  $y_p$  and  $y_a$  in  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  for  $\text{Ni}^{2+}$  with variable  $T$  and FWHM in the description vector evaluated on the test dataset (10 000 spectra). Data point sizes and shades represent the SE between recomputed and reference spectra. (h) The color shades are defined in the histogram of the spectrum error (SE) in 21 bins. The inset shows a part of the histogram with larger resolution in the counts axis.

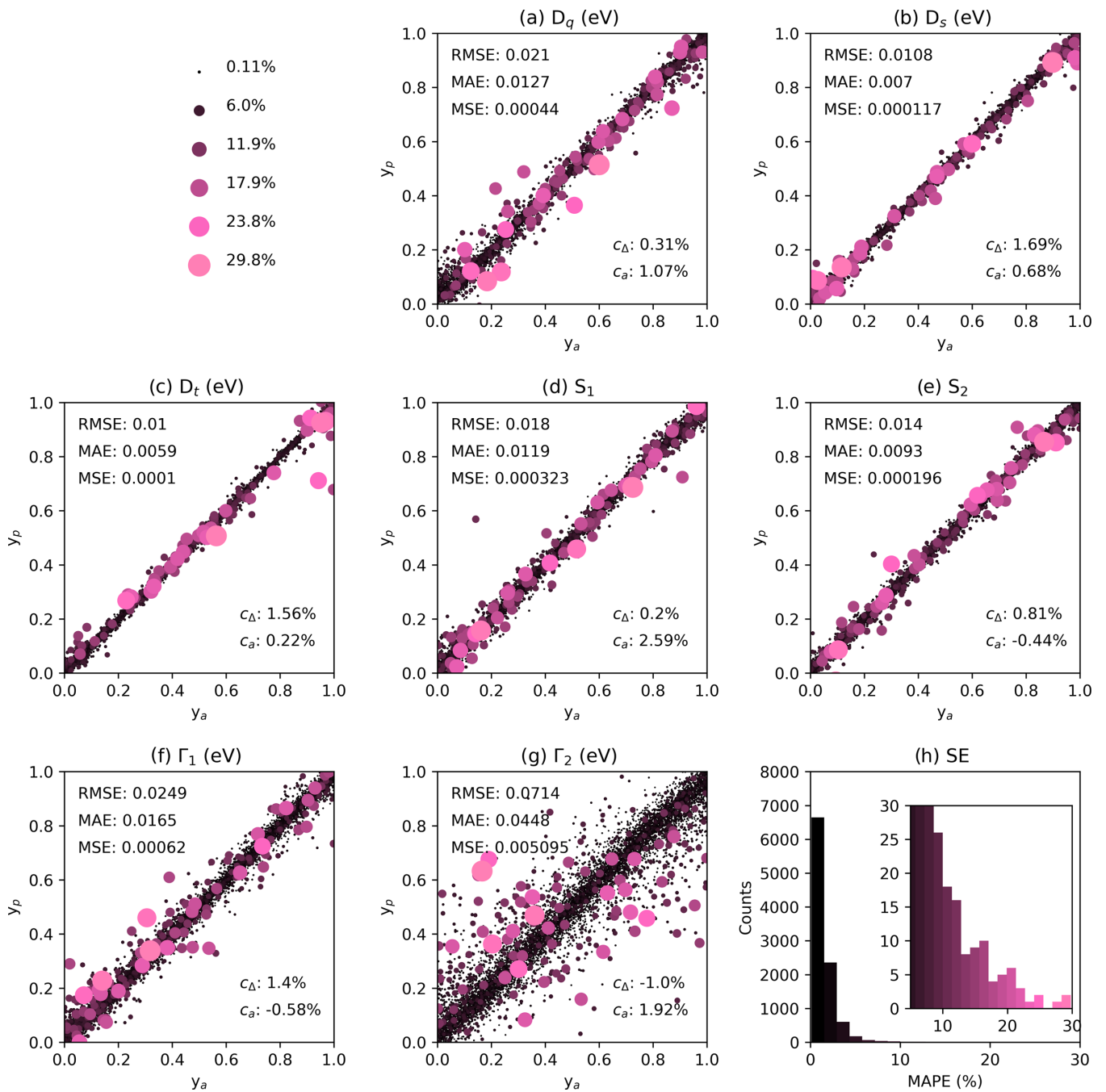


FIG. S4. From (a) to (g): Comparison of normalized  $y_p$  and  $y_a$  in  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  for  $\text{Fe}^{2+}$  with variable  $T$  and FWHM in the description vector evaluated on the test dataset (10 000 spectra). Data point sizes and shades represent the SE between recomputed and reference spectra. (h) The color shades are defined in the histogram of the spectrum error (SE) in 21 bins. The inset shows a part of the histogram with larger resolution in the counts axis.

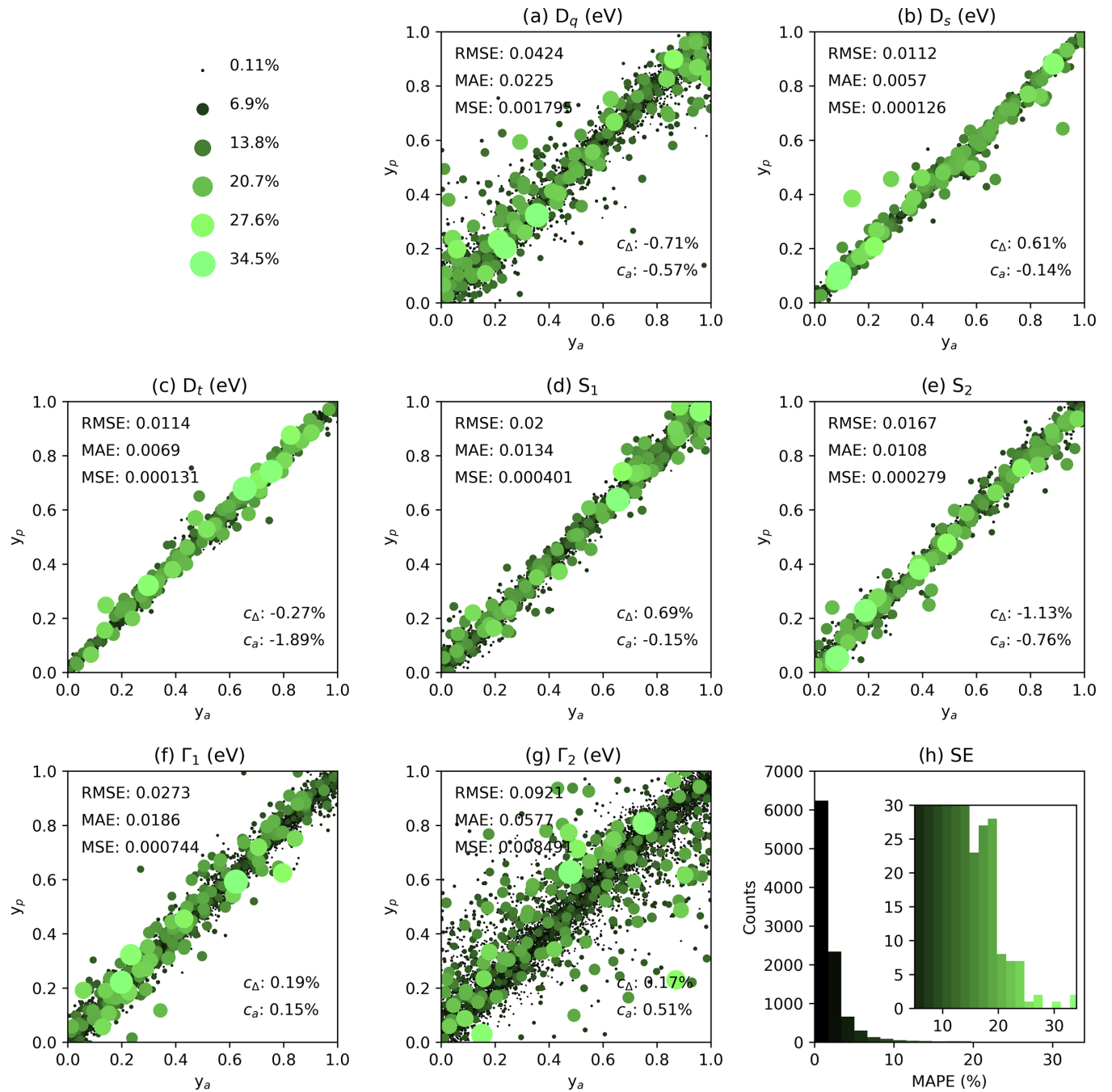


FIG. S5. From (a) to (g): Comparison of normalized  $y_p$  and  $y_a$  in  $D_q$ ,  $D_s$ ,  $D_t$ ,  $S_1$ ,  $S_2$ ,  $\Gamma_1$ , and  $\Gamma_2$  for  $Mn^{2+}$  with variable  $T$  and FWHM in the description vector evaluated on the test dataset (10 000 spectra). Data point sizes and shades represent the SE between recomputed and reference spectra. (h) The color shades are defined in the histogram of the spectrum error (SE) in 21 bins. The inset shows a part of the histogram with larger resolution in the counts axis.

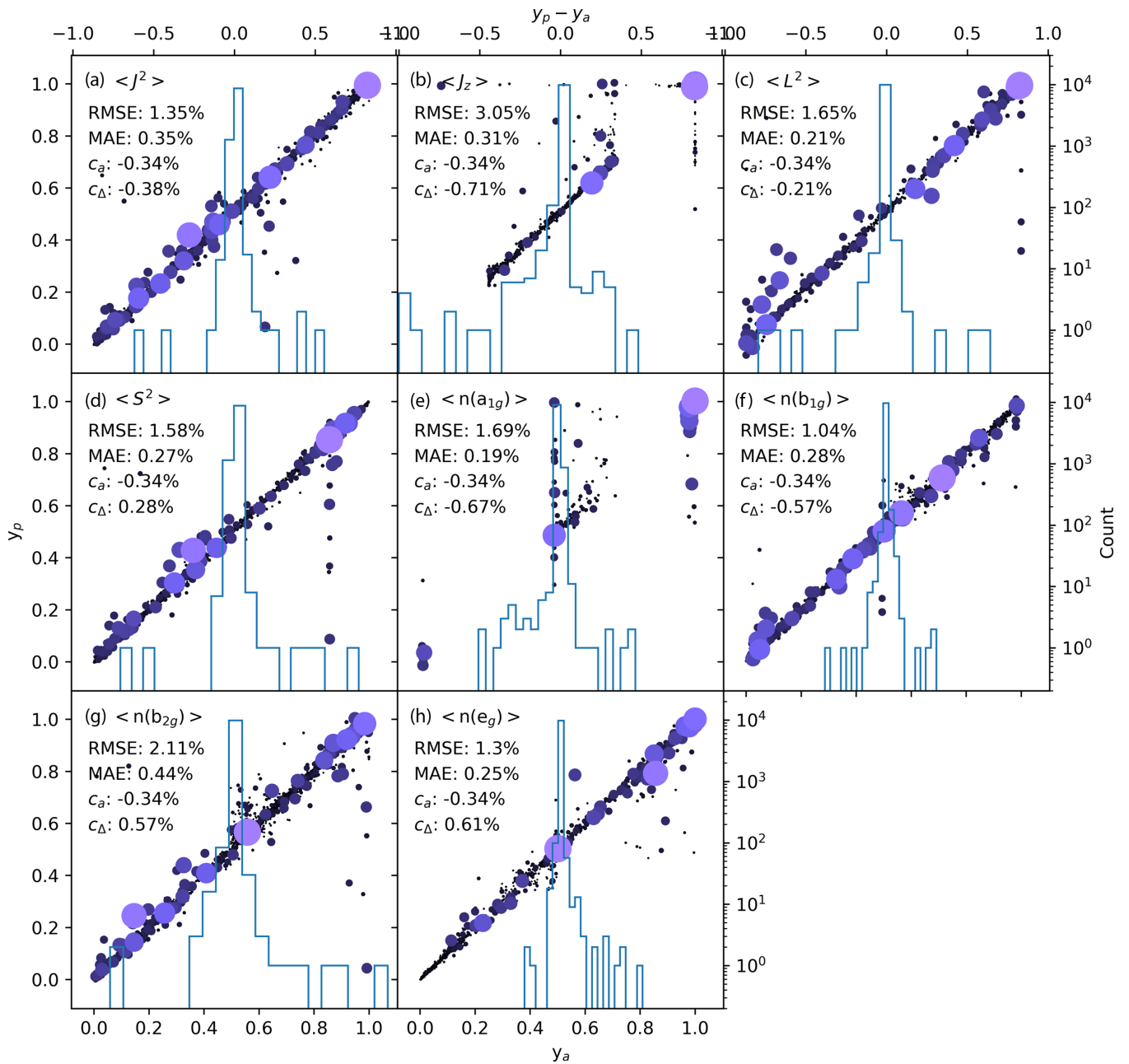


FIG. S6. Comparison of normalized  $y_p$  vs  $y_a$  for predicted properties by A2pX-T2EP for  $Ni^{2+}$  determined on the test dataset (10 000 spectra). The overlaid histograms show the distribution of difference between  $y_p$  vs  $y_a$  on a logarithmic count scale. Pearson correlation to SE for  $\Delta_y = y_p - y_a$  ( $c_\Delta$ ) and  $y_a$  ( $c_a$ ). Color code and data point sizes are given in Fig. S3. The scales of the histograms are on the top right; the scales of predicted ( $y_p$ ) vs actual data ( $y_a$ ) are given on the bottom left.

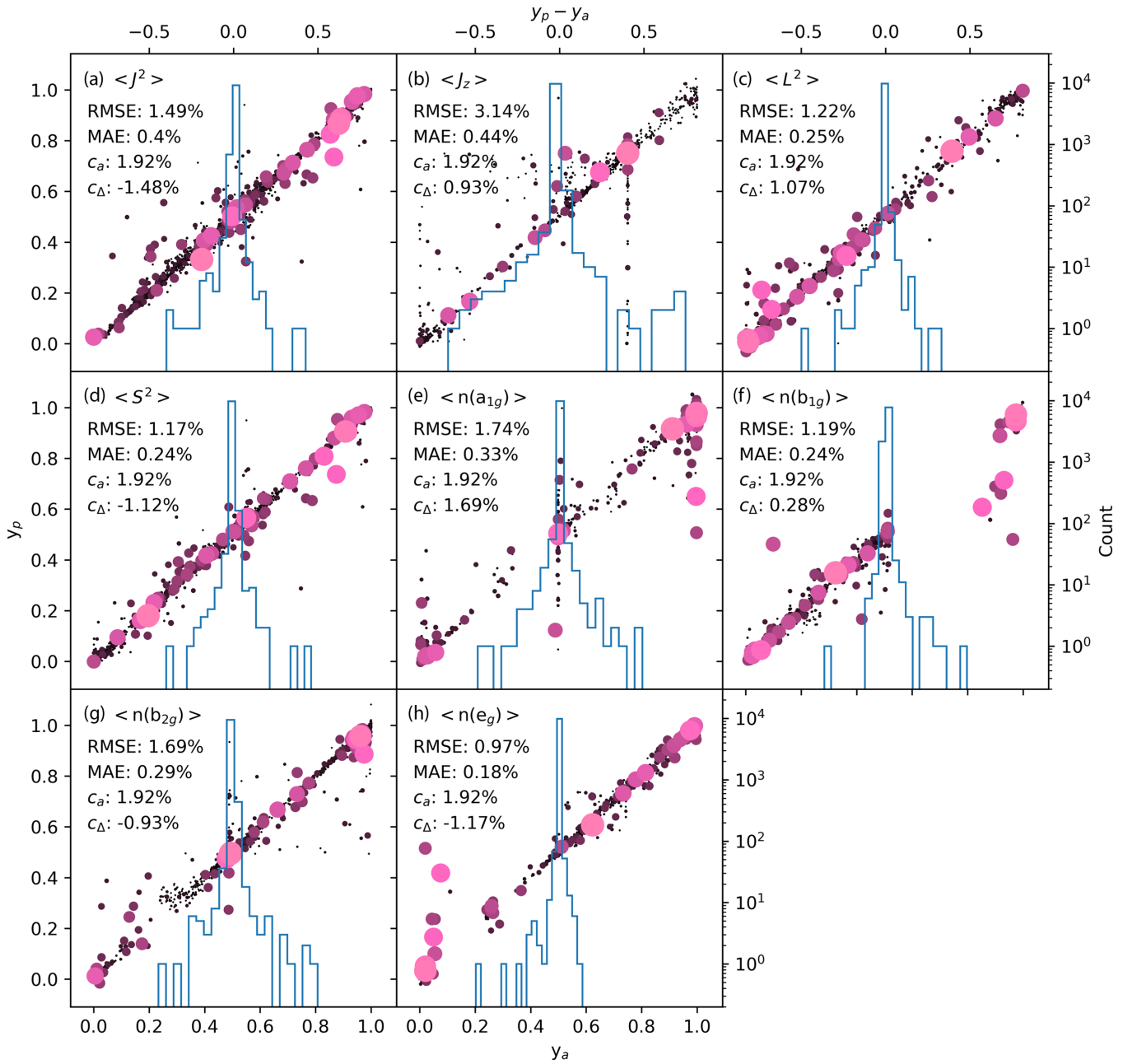


FIG. S7. Comparison of normalized  $y_p$  vs  $y_a$  for predicted properties by A2pX-T2EP for  $\text{Fe}^{2+}$  determined on test dataset (10 000 spectra). The overlaid histograms show the distribution of difference between  $y_p$  vs  $y_a$  on a logarithmic count scale. Pearson correlation to SE for  $\Delta_y = y_p - y_a$  ( $c_\Delta$ ) and  $y_a$  ( $c_a$ ). Color code and data point sizes are as in Fig. S4. The scales of the histograms are on the right top; the scales of predicted ( $y_p$ ) vs actual data ( $y_a$ ) are given on the bottom left.



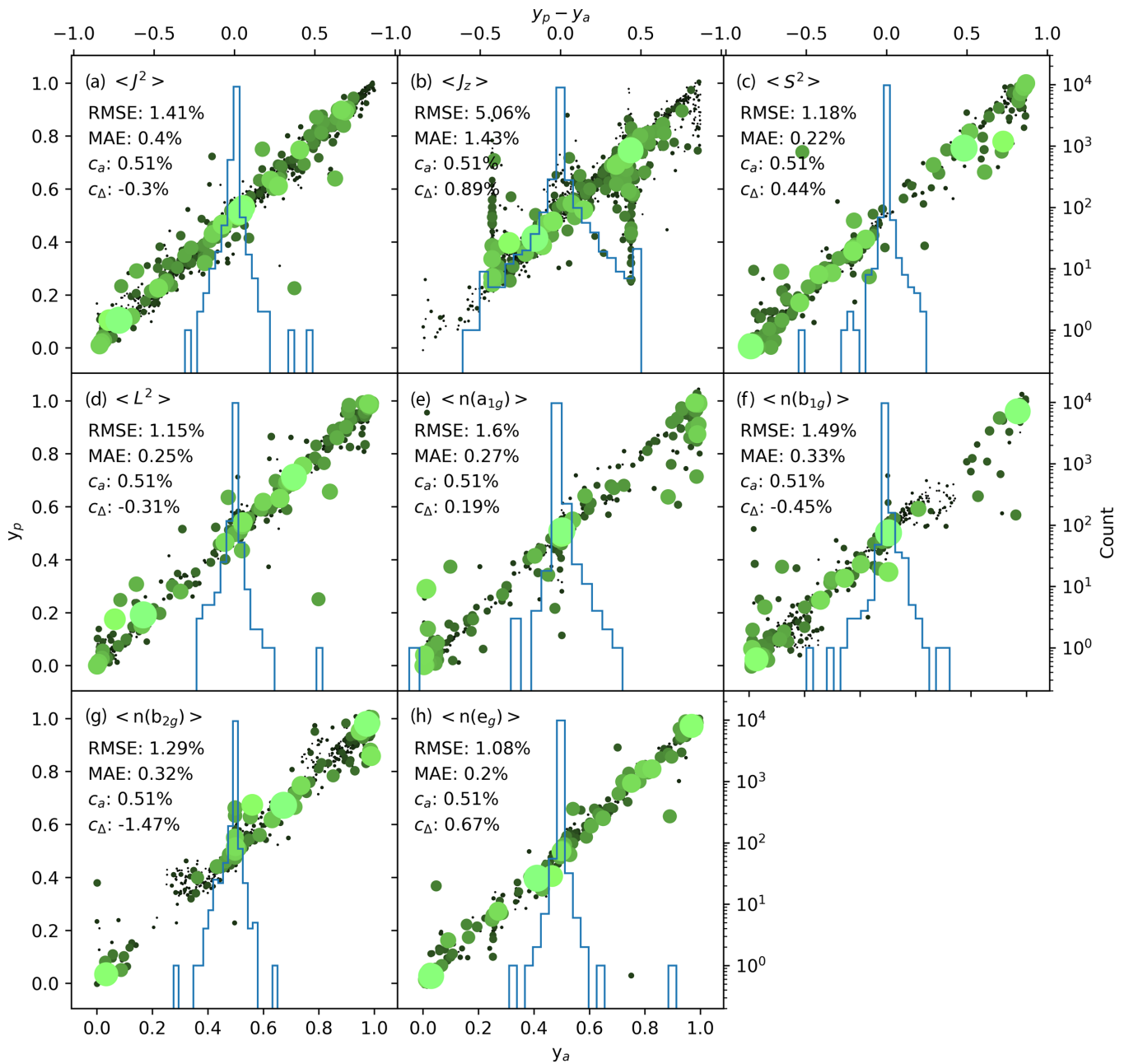


FIG. S8. Comparison of normalized  $y_p$  vs  $y_a$  for predicted properties by A2pX-T2EP for  $Mn^{2+}$  determined on the test dataset (10 000 spectra). The overlaid histograms show the distribution of difference between  $y_p$  vs  $y_a$  on a logarithmic count scale. Pearson correlation to SE for  $\Delta_y = y_p - y_a$  ( $c_\Delta$ ) and  $y_a$  ( $c_a$ ). Color code and data point sizes are as in Fig. S5. The scales of the histograms are on the right top; the scales of predicted ( $y_p$ ) vs actual data ( $y_a$ ) are given on the bottom left.

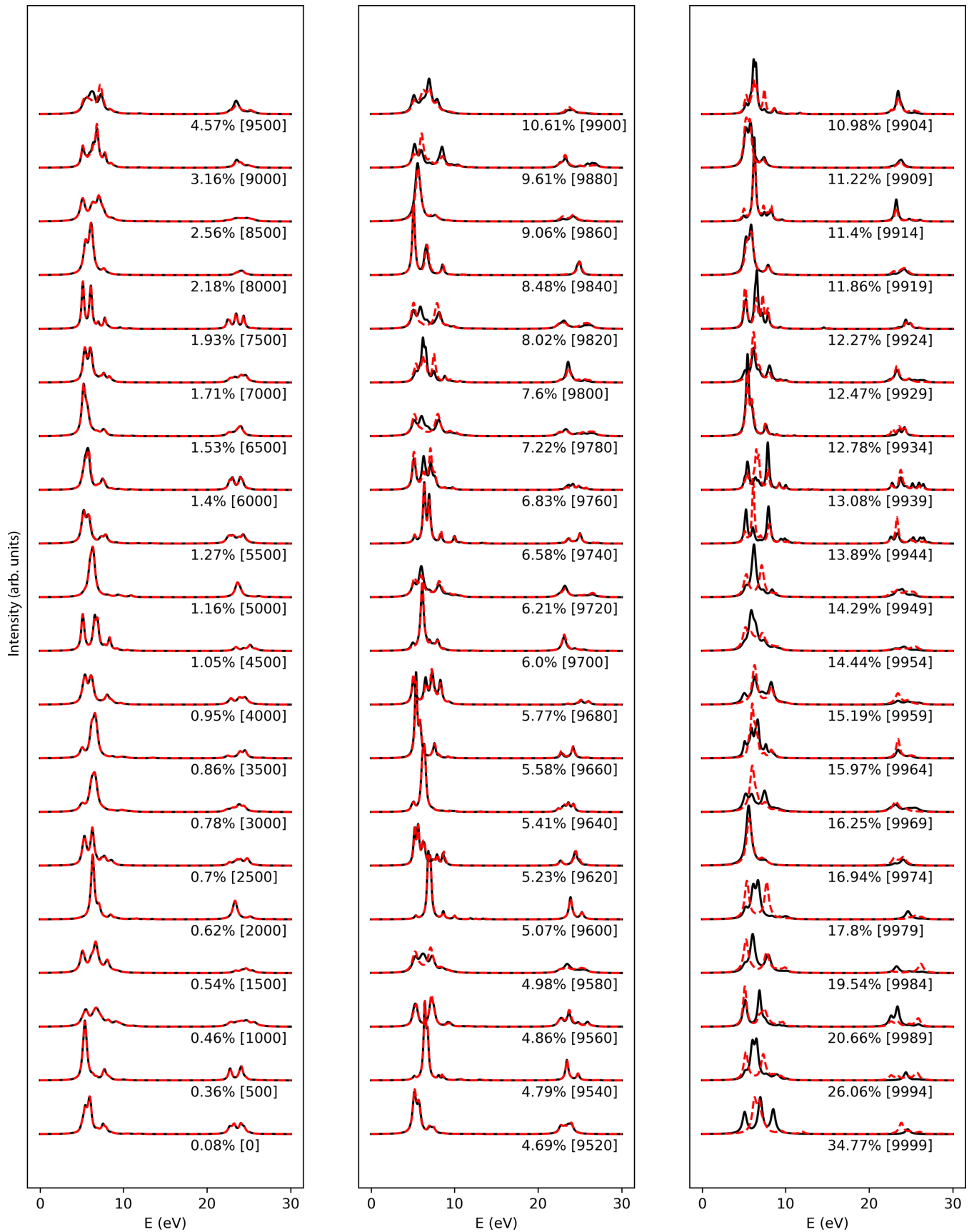


FIG. S9. Overview of selected  $\text{Ni}^{2+}$  reference spectra (black curve) and recomputed spectra (red dashed). The SE (%) is indicated as well as the ranked position according to SE in square brackets in the test dataset. The energy scale is shifted.

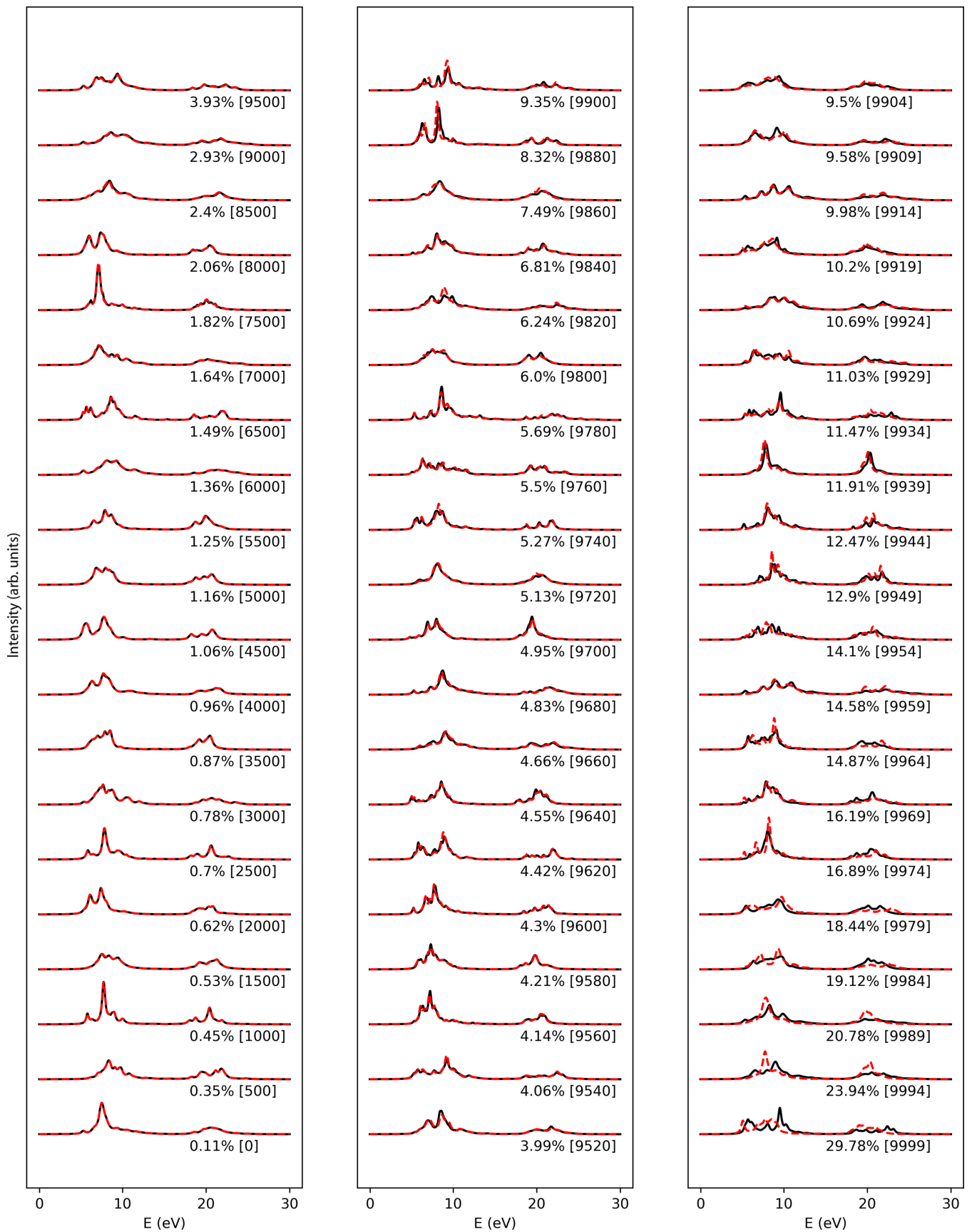


FIG. S10. Overview of selected Fe<sup>2+</sup> reference spectra (black curve) and recomputed spectra (red dashed). The SE (%) is indicated as well as the ranked position according to SE in square brackets in the test dataset. The energy scale is shifted.

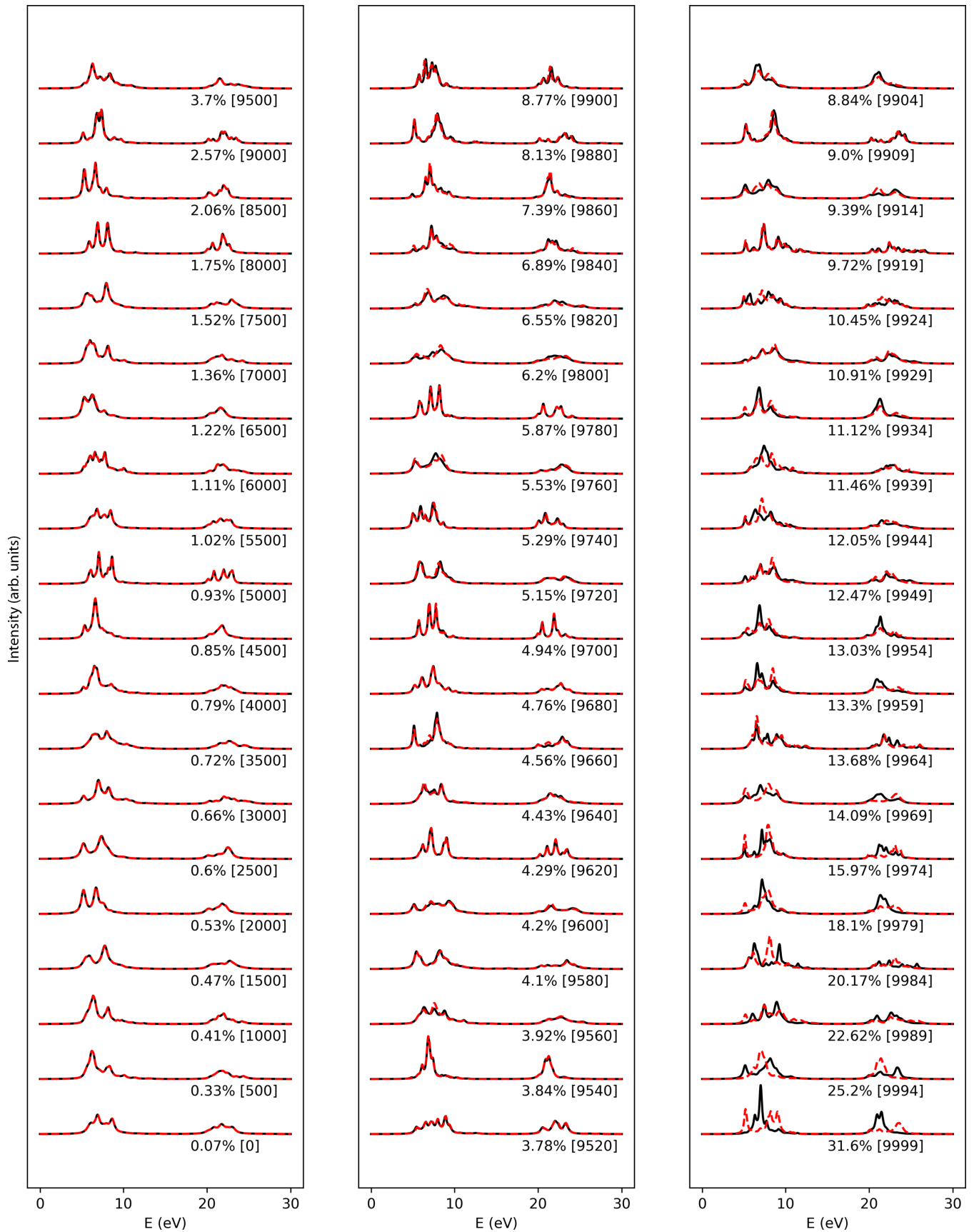


FIG. S11. Overview of selected  $\text{Co}^{2+}$  reference spectra (black curve) and recomputed spectra (red dashed). The SE (%) is indicated as well as the ranked position according to SE in square brackets in the test dataset. The energy scale is shifted.

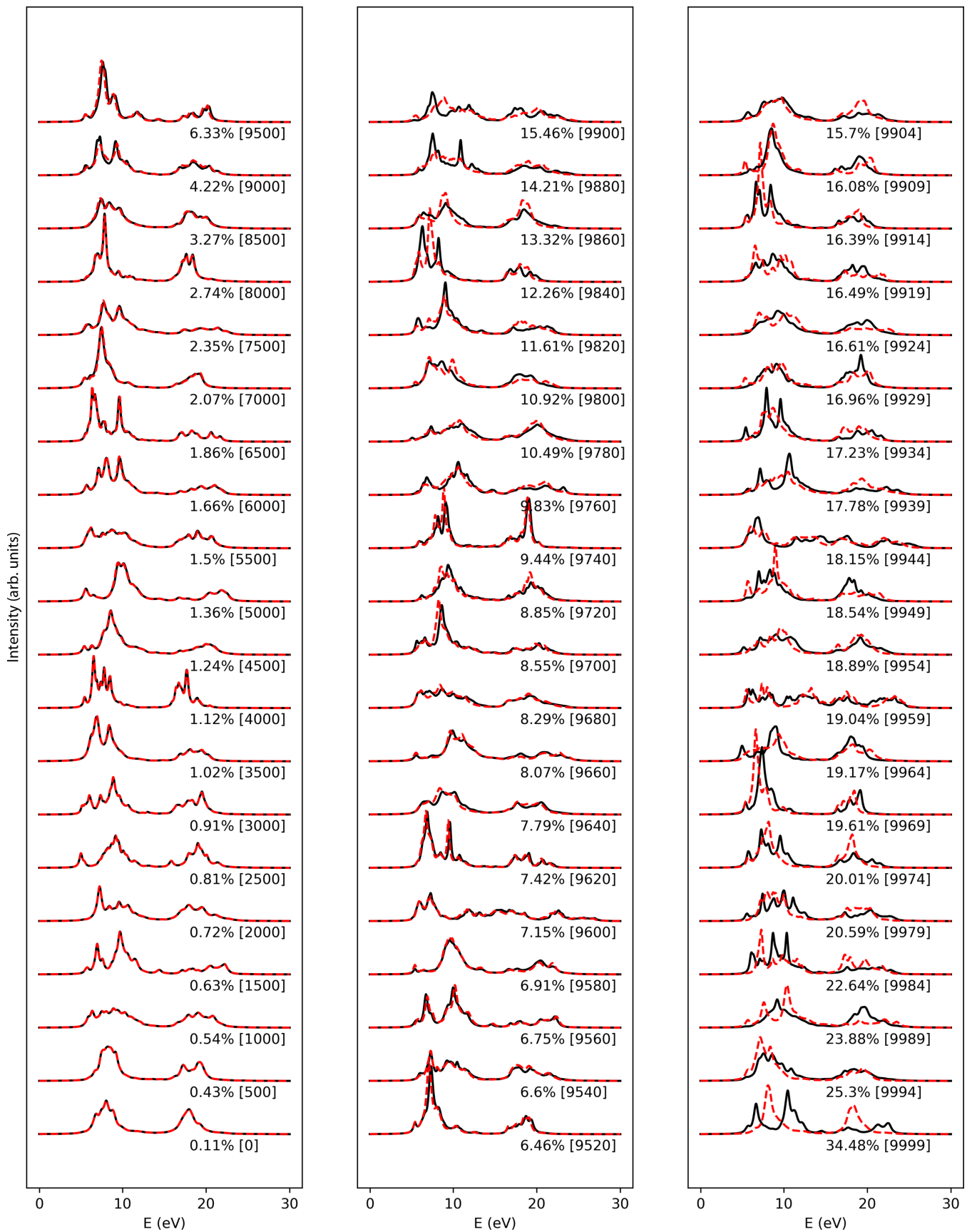


FIG. S12. Overview of selected  $Mn^{2+}$  reference spectra (black curve) and recomputed spectra (red dashed). The SE (%) is indicated as well as the ranked position according to SE in square brackets in the test dataset. The energy scale is shifted.

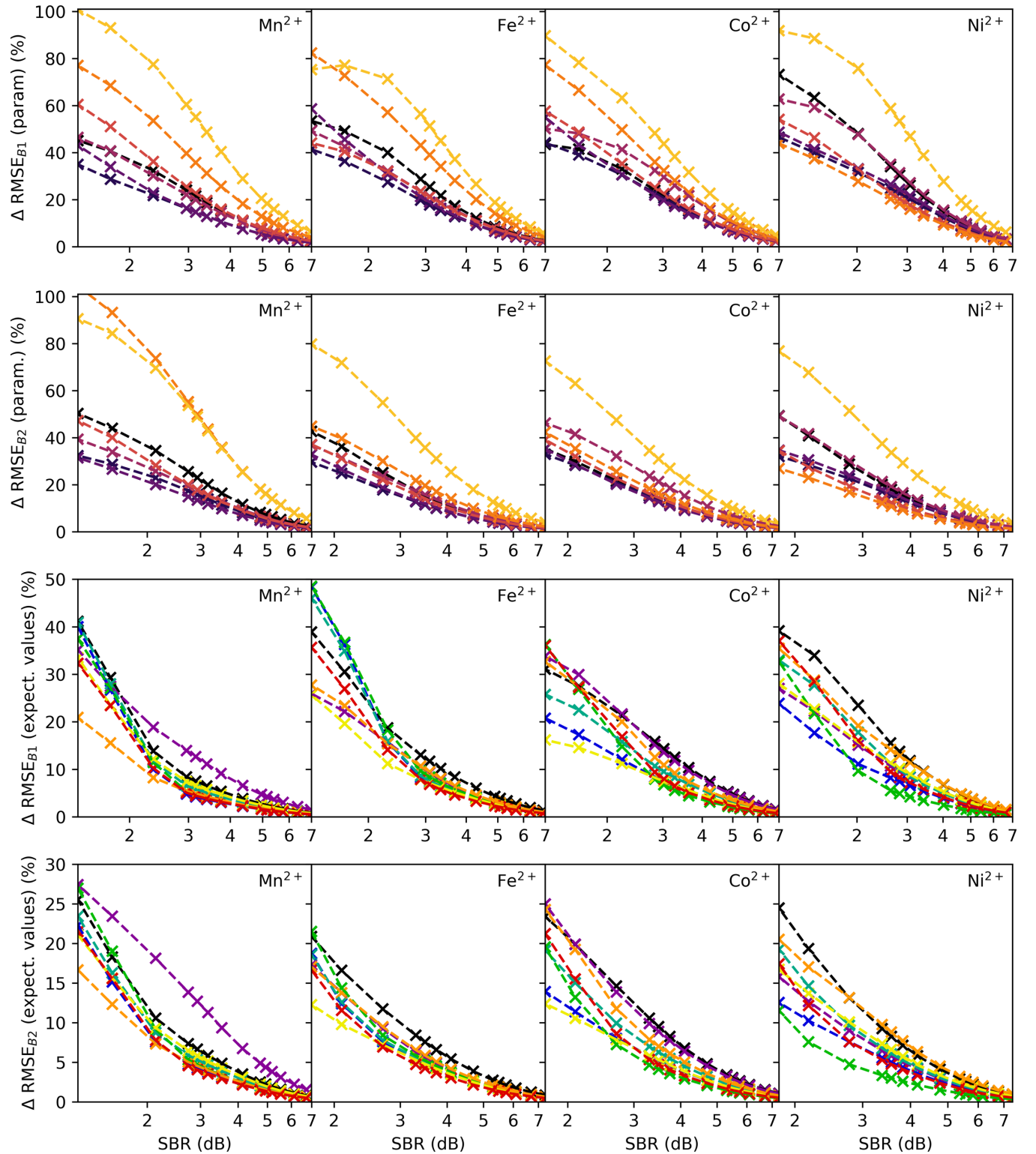


FIG. S13. Influence of step background centered at either  $L_3$  ( $B_1$ ) or  $L_2$  ( $B_2$ ) in test spectra on the A2pX-T2EP's analytic (parameter estimation) and predictive (expectation value estimation) performance measured in  $\Delta\text{RMSE}$ . The same curve labeling as in Fig. 5 applies.

- [1] A. Ponrouch, C. Frontera, F. Bardé, and M. R. Palacín, Towards a calcium-based rechargeable battery, *Nat. Mater.* **15**, 169 (2016).
- [2] P. Hansmann, A. Severing, Z. Hu, M. W. Haverkort, C. F. Chang, S. Klein, A. Tanaka, H. H. Hsieh, H. J. Lin, C. T. Chen, B. Fåk, P. Lejay, and L. H. Tjeng, Determining the Crystal-Field Ground State in Rare Earth Heavy Fermion Materials Using soft-X-ray Absorption Spectroscopy, *Phys. Rev. Lett.* **100**, 066405 (2008).
- [3] C. Wäckerlin, D. Chylarecka, A. Kleibert, K. Müller, C. Iacovita, F. Nolting, T. A. Jung, and N. Ballav, Controlling spins in adsorbed molecules by a chemical switch, *Nat. Commun.* **1**, 61 (2010).
- [4] P. Hartmann, C. L. Bender, M. Vračar, A. K. Dürr, A. Garsuch, J. Janek, and P. Adelhelm, A rechargeable room-temperature sodium superoxide (NaO<sub>2</sub>) battery, *Nat. Mater.* **12**, 228 (2012).
- [5] B. Stadtmüller, D. Lüftner, M. Willenbockel, E. M. Reinisch, T. Sueyoshi, G. Koller, S. Soubatch, M. G. Ramsey, P. Puschnig, F. S. Tautz, and C. Kumpf, Unexpected interplay of bonding height and energy level alignment at heteromolecular hybrid interfaces, *Nat. Commun.* **5**, 3685 (2014).
- [6] F. Sedona, M. Di Marino, D. Forrer, A. Vittadini, M. Casarin, A. Cossaro, L. Floreano, A. Verdini, and M. Sambì, Tuning the catalytic activity of Ag(110)-supported Fe phthalocyanine in the oxygen reduction reaction, *Nat. Mater.* **11**, 970 (2012).
- [7] K. Wakonig, A. Diaz, A. Bonnin, M. Stampanoni, A. Bergamaschi, J. Ihli, M. Guizar-Sicairos, and A. Menzel, X-ray Fourier ptychography, *Sci. Adv.* **5**, eaav0282 (2019).
- [8] I. E. Brumboiu, O. Eriksson, and P. Norman, Photoelectron spectroscopy of molecules beyond the electric dipole approximation, *J. Chem. Theory Comput.* **15**, 5483 (2019).
- [9] L. Ma, X. Zhang, Z. Xu, A. Späth, Z. Xing, T. Sun, and R. Tai, Three-dimensional focal stack imaging in scanning transmission X-ray microscopy with an improved reconstruction algorithm, *Opt. Express* **27**, 7787 (2019).
- [10] D. Golze, J. Wilhelm, M. J. Van Setten, and P. Rinke, Core-Level binding energies from GW: An efficient full-frequency approach within a localized basis, *J. Chem. Theory Comput.* **14**, 4856 (2018).
- [11] M. W. Haverkort, Theory of Resonant Inelastic X-ray Scattering by Collective Magnetic Excitations, *Phys. Rev. Lett.* **105**, 167404 (2010).
- [12] M. Daniel and D. Astruc, Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications toward Biology, Catalysis, and Nanotechnology, *Chem. Rev.* **104**, 293 (2004).
- [13] M. Ghiasi, A. Hariki, M. Winder, J. Kuneš, A. Regoutz, T. L. Lee, Y. Hu, J. P. Rueff, and F. M. De Groot, Charge-transfer effect in hard x-ray 1s and 2p photoemission spectra: LDA+DMFT and cluster-model analysis, *Phys. Rev. B* **100**, 075146 (2019).
- [14] Y. Wang, L. Zhu, N. Ba, F. Gao, and H. Xie, Effects of NH<sub>4</sub>F quantity on N-doping level, photodegradation and photocatalytic H<sub>2</sub> production activities of N-doped TiO<sub>2</sub> nanotube array films, *Mater. Res. Bull.* **86**, 268 (2017).
- [15] I. S. Lim, J. Ouyang, J. Luo, L. Wang, S. Zhou, and C.-J. Zhong, Multifunctional fullerene-mediated assembly of gold nanoparticles, *Chem. Mater.* **17**, 6528 (2005).
- [16] P. A. M. Dirac, The quantum theory of the emission and absorption of radiation, *Proc. R. Soc. Lond. A*, **114**, 243 (1927).
- [17] J. J. Rehr and R. C. Albers, Theoretical approaches to x-ray absorption fine structure, *Rev. Mod. Phys.* **72**, 621 (2000).
- [18] K. R. Wilson, M. Cavalleri, B. S. Rude, R. D. Schaller, T. Catalano, A. Nilsson, R. J. Saykally, and L. G. M. Pettersson, X-ray absorption spectroscopy of liquid methanol microjets: Bulk electronic structure and hydrogen bonding network, *J. Phys. Chem. B* **109**, 10194 (2005).
- [19] E. Rauls, W. G. Schmidt, T. Pertram, and K. Wandelt, Interplay between metal-free phthalocyanine molecules and Au(110) substrates, *Surf. Sci.* **606**, 1120 (2012).
- [20] M. Odelius, M. Cavalleri, A. Nilsson, and L. G. M. Pettersson, X-ray absorption spectrum of liquid water from molecular dynamics simulations: Asymmetric model, *Phys. Rev. B* **73**, 024205 (2006).
- [21] X. Wang, L. de' Medici, and A. J. Millis, Theory of oxygen K edge x-ray absorption spectra of cuprates, *Phys. Rev. B* **81**, 094522 (2010).
- [22] L. Triguero, Y. Luo, L. G. M. Pettersson, H. Ågren, P. Väterlein, M. Weinelt, A. Föhlisch, J. Hasselström, O. Karis, and A. Nilsson, Resonant soft-x-ray emission spectroscopy of surface adsorbates: Theory, computations, and measurements of ethylene and benzene on Cu(110), *Phys. Rev. B* **59**, 5189 (1999).
- [23] B. Brena and Y. Luo, Time-dependent DFT calculations of core electron shake-up states of metal-(free)-phthalocyanines, *Radiat. Phys. Chem.* **75**, 1578 (2006).
- [24] B. Brena, Y. Luo, M. Nyberg, S. Carniato, K. Nilson, Y. Alfredsson, J. Åhlund, N. Mårtensson, H. Siegbahn, and C. Puglia, Equivalent core-hole time-dependent density functional theory calculations of carbon 1s shake-up states of phthalocyanine, *Phys. Rev. B* **70**, 195214 (2004).
- [25] K. Weiss, S. Gebert, M. Wühh, H. Wadepohl, and C. Wöll, Near edge x-ray absorption fine structure study of benzene adsorbed on metal surfaces: Comparison to benzene cluster complexes, *J. Vac. Sci. Technol. A* **16**, 1017 (1998).
- [26] J. Janczak and R. Kubiak, X-ray single crystal investigations of magnesium phthalocyanine. The 4+1 coordination of the Mg ion and its consequence, *Polyhedron* **20**, 2901 (2001).
- [27] L. G. Pettersson, H. Ågren, Y. Luo, and L. Triguero, Benzene adsorbed on Cu(110): Theoretical X-ray absorption, emission and shake calculations, *Surf. Sci.* **408**, 1 (1998).
- [28] L. Triguero, L. G. M. Pettersson, and H. Ågren, Calculations of near-edge x-ray-absorption spectra of gas-phase and chemisorbed molecules by means of density-functional and transition-potential theory, *Phys. Rev. B* **58**, 8097 (1998).
- [29] T. Zhang, I. E. Brumboiu, V. Lanzilotto, J. Lüder, C. Grazioli, E. Giangrisostomi, R. Ovsyannikov, Y. Sassa, I. Bidermane, M. Stupar, M. de Simone, M. Coreno, B. Ressel, M. Pedio, P. Rudolf, B. Brena, and C. Puglia, Conclusively addressing the CoPc electronic structure: A joint gas-phase and solid-state photoemission and absorption spectroscopy study, *J. Phys. Chem. C* **121**, 26372 (2017).
- [30] F. De Groot and A. Kotani, *Core Level Spectroscopy of Solids* (CRC Press, Boca Raton, 2008).
- [31] F. D. Groot, Multiplet effects in X-ray spectroscopy, *Coord. Chem. Rev.* **249**, 31 (2005).
- [32] J. Rehr, Theory and calculations of X-ray spectra: XAS, XES, XRS, and NRIXS, *Radiat. Phys. Chem.* **75**, 1547 (2006).

- [33] E. L. Shirley, Ti 1s pre-edge features in rutile: A Bethe-Salpeter calculation, *J. Electron Spectrosc. Relat. Phenom.* **136**, 77 (2004).
- [34] E. L. Shirley, Bethe-Salpeter treatment of X-ray absorption including core-hole multiplet effects, *J. Electron Spectrosc. Relat. Phenom.* **144-147**, 1187 (2005).
- [35] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, Bethe-Salpeter equation calculations of core excitation spectra, *Phys. Rev. B* **83**, 115106 (2011).
- [36] Y. Kumagai, H. Ikeno, and I. Tanaka, All-electron CI calculations of 3d transition-metal L 2,3 XANES using zeroth-order regular approximation for relativistic effects, *J. Phys.: Condens. Matter* **21**, 104209 (2009).
- [37] I. Josefsson, K. Kunnus, S. Schreck, A. Föhlisch, F. de Groot, P. Wernet, and M. Odelius, *Ab Initio* calculations of x-ray spectra: Atomic multiplet and molecular orbital effects in a multiconfigurational SCF approach to the L-edge spectra of transition metal complexes, *J. Phys. Chem. Lett.* **3**, 3565 (2012).
- [38] F. M. de Groot, Ligand and metal X-ray absorption in transition metal complexes, *Inorganica Chimica Acta* **361**, 850 (2008).
- [39] E. Stavitski and F. M. F. de Groot, The CTM4XAS program for EELS and XAS spectral shape analysis of transition metal L edges, *Micron* **41**, 687 (2010).
- [40] M. W. Haverkort, M. Zwierzycki, and O. K. Andersen, Multiplet ligand-field theory using Wannier orbitals, *Phys. Rev. B* **85**, 165113 (2012).
- [41] J. Lüder, J. Schött, B. Brena, M. W. Haverkort, P. Thunström, O. Eriksson, B. Sanyal, I. Di Marco, and Y. O. Kvashnin, Theory of L -edge spectroscopy of strongly correlated systems, *Phys. Rev. B* **96**, 245131 (2017).
- [42] H. K. D. H. Bhadeshia, Neural networks in materials science, *ISIJ International* **39**, 966 (1999).
- [43] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science, *Nature (London)* **559**, 547 (2018).
- [44] H. Ji and Y. Jung, Artificial neural network for the configuration problem in solids, *J. Chem. Phys.* **146**, 064103 (2017).
- [45] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, 303 (1989).
- [46] R. Xia and S. Kais, Quantum machine learning for electronic structure calculations, *Nat. Commun.* **9**, 4195 (2018).
- [47] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.* **3**, e1603015 (2017).
- [48] R. M. Balabin and E. I. Lomakina, Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies, *J. Chem. Phys.* **131**, 074104 (2009).
- [49] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, and S. Manzhos, Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy, *J. Chem. Phys.* **148**, 241702 (2018).
- [50] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, and A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature (London)* **533**, 73 (2016).
- [51] U. Kumar, S. Nayak, S. Chakrabarty, S. Bhattacharjee, and S.-C. Lee, Gallium-Boron-Phosphide (GaBP<sub>2</sub>): A new III-V semiconductor for photovoltaics, *J. Mater. Sci.* **55**, 9448 (2020).
- [52] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Bypassing the Kohn-Sham equations with machine learning, *Nat. Commun.* **8**, 872 (2017).
- [53] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, From DFT to machine learning: recent approaches to materials science—a review, *J. Phys. Mater.* **2**, 032001 (2019).
- [54] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, Solving the electronic structure problem with machine learning, *npj Comput. Mater.* **5**, 22 (2019).
- [55] A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig, and A. Mar, High-throughput machine-learning-driven synthesis of full-heusler compounds, *Chem. Mater.* **28**, 7324 (2016).
- [56] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space, *J. Phys. Chem. Lett.* **6**, 2326 (2015).
- [57] K. Mills, M. Spanner, and I. Tamblin, Deep learning and the Schrödinger equation, *Phys. Rev. A* **96**, 042113 (2017).
- [58] R. L. Carlin, Introduction to ligand field theory (Ballhausen. Carl J.), *J. Chem. Education* **40**, 390 (1963).
- [59] M. W. Haverkort, Spin and orbital degrees of freedom in transition metal oxides and oxide thin films studied by soft x-ray absorption spectroscopy, [arXiv:cond-mat/0505214](https://arxiv.org/abs/cond-mat/0505214).
- [60] M. Shariati, J. Lüder, I. Bidermane, S. Ahmadi, E. Göthelid, P. Palmgren, B. Sanyal, O. Eriksson, M. N. Piancastelli, B. Brena, and C. Puglia, Photoelectron and absorption spectroscopy studies of metal-free phthalocyanine on Au(111): Experiment and theory, *J. Phys. Chem. C* **117**, 7018 (2013).
- [61] I. Bidermane, J. Lüder, S. Boudet, T. Zhang, S. Ahmadi, C. Grazioli, M. Bouvet, J. Ruzs, B. Sanyal, O. Eriksson, B. Brena, C. Puglia, and N. Witkowski, Experimental and theoretical study of electronic structure of lutetium bi-phthalocyanine, *J. Chem. Phys.* **138**, 234701 (2013).
- [62] I. E. Brumboiu, A. Anselmo, B. Brena, A. Dzwilewski, K. Svensson, and E. Moons, Near-edge X-ray absorption fine structure study of the C60-derivative PCBM, *Chem. Phys. Lett.* **568-569**, 130 (2013).
- [63] S. Manzhos, R. Dawes, and T. Carrington, Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces, *Int. J. Quantum Chem.* **115**, 1012 (2015).
- [64] M. Majumder, S. E. Hegger, R. Dawes, S. Manzhos, X. G. Wang, C. Tucker, J. Li, and H. Guo, Explicitly correlated MRCI-F12 potential energy surfaces for methane fit with several permutation invariant schemes and full-dimensional vibrational calculations, *Mol. Phys.* **113**, 1823 (2015).
- [65] G. B. Goh, N. O. Hodas, and A. Vishnu, Deep learning for computational chemistry, *J. Comput. Chem.* **38**, 1291 (2017).
- [66] A. V. Sinitskiy and V. S. Pande, Deep Neural Network Computes Electron Densities and Energies of a Large Set of Organic Molecules Faster than Density Functional Theory (DFT), [arXiv:1809.02723](https://arxiv.org/abs/1809.02723).
- [67] J. Timoshenko and A. I. Frenkel, “Inverting” x-ray absorption spectra of catalysts by machine learning in search for activity descriptors, *ACS Catal.* **9**, 10192 (2019).
- [68] M. Retegan, [Crispy](https://github.com/Retegan/Crispy): v0.7.3, 2019, [doi:10.5281/zenodo.1008184](https://doi.org/10.5281/zenodo.1008184).



- [69] F. de Groot, X-ray absorption and dichroism of transition metal compounds, in *The 17th International Conference on X-Ray and Inner-Shell Processes*, Vol. 389 (ASCE, 1997), pp. 497–520.
- [70] S. I. Csiszar, M. W. Haverkort, Z. Hu, A. Tanaka, H. H. Hsieh, H.-J. Lin, C. T. Chen, T. Hibma, and L. H. Tjeng, Controlling Orbital Moment and Spin Orientation in CoO Layers by Strain, *Phys. Rev. Lett.* **95**, 187205 (2005).
- [71] H. Huo and M. Rupp, Unified Representation of Molecules and Crystals for Machine Learning, [arXiv:1704.06439](https://arxiv.org/abs/1704.06439).
- [72] F. Chollet, keras (2015), <https://github.com/fchollet/keras>.
- [73] M. D. Zeiler, ADADELTA: An Adaptive Learning Rate Method, [arXiv:1212.5701](https://arxiv.org/abs/1212.5701).
- [74] M. Kubin, J. Kern, M. Guo, E. Källman, R. Mitzner, V. K. Yachandra, M. Lundberg, J. Yano, and P. Wernet, X-ray-induced sample damage at the Mn L-edge: A case study for soft X-ray spectroscopy of transition metal complexes in solution, *Phys. Chem. Chem. Phys.* **20**, 16817 (2018).
- [75] E. Kreyszig, H. Kreyszig, and E. J. Norminton, *Advanced Engineering Mathematics, Tenth* (Wiley, Hoboken, NJ, 2011).
- [76] D. Alders, L. H. Tjeng, F. C. Voigt, T. Hibma, G. A. Sawatzky, C. T. Chen, J. Vogel, M. Sacchi, and S. Iacobucci, Temperature and thickness dependence of magnetic moments in NiO epitaxial films, *Phys. Rev. B* **57**, 11623 (1998).
- [77] F. M. F. de Groot, M. Abbate, J. van Elp, G. A. Sawatzky, Y. J. Ma, C. T. Chen, and F. Sette, Oxygen 1s and cobalt 2p X-ray absorption of cobalt oxides, *J. Phys.: Condens. Matter* **5**, 2277 (1993).
- [78] T. J. Regan, H. Ohldag, C. Stamm, F. Nolting, J. Lüning, J. Stöhr, and R. L. White, Chemical effects at metal/oxide interfaces studied by x-ray-absorption spectroscopy, *Phys. Rev. B* **64**, 214422 (2001).
- [79] B. Gilbert, B. H. Frazer, A. Belz, P. G. Conrad, K. H. Neilson, D. Haskel, J. C. Lang, G. Srajer, G. D. Stasio, and G. De Stasio, Multiple scattering calculations of bonding and x-ray absorption spectroscopy of manganese oxides, *J. Phys. Chem. A* **107**, 2839 (2003).
- [80] P. S. Miedema, S. Stepanow, P. Gambardella, and F. M. F. de Groot, 2p x-ray absorption of iron-phthalocyanine, *J. Phys.: Conf. Ser.* **190**, 012143 (2009).
- [81] P. Coppens, L. Li, and N. J. Zhu, Electronic ground state of iron(II) phthalocyanine as determined from accurate diffraction data, *J. Am. Chem. Soc.* **105**, 6173 (1983).
- [82] T. Kroll, R. Kraus, R. Schönfelder, V. Y. Aristov, O. V. Molodtsova, P. Hoffmann, and M. Knupfer, Transition metal phthalocyanines: Insight into the electronic structure from soft x-ray spectroscopy, *J. Chem. Phys.* **137**, 054306 (2012).
- [83] O. Snezhkova, J. Lüder, A. Wiengarten, S. R. Burema, F. Bischoff, Y. He, J. Ruzs, J. Knudsen, M.-L. Bocquet, K. Seufert, J. V. Barth, W. Auwärter, B. Brena, and J. Schnadt, Nature of the bias-dependent symmetry reduction of iron phthalocyanine on Cu(111), *Phys. Rev. B* **92**, 075428 (2015).