Learning crystal field parameters using convolutional neural networks

Noah F. Berthusen,[1, 2] Yuriy Sizyuk,[1, 3] Mathias S. Scheurer,[4, 5] and Peter P. Orth[1, 3, *]

[1]*Ames Laboratory, Ames, Iowa 50011, USA*
[2]*Department of Electrical and Computer Engineering,*
Iowa State University, Ames, Iowa 50011, USA.
[3]*Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA*
[4]*Department of Physics, Harvard University, Cambridge MA 02138, USA*
[5]*Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria*
(Dated: December 2, 2020)

We present a deep machine learning algorithm to extract crystal field (CF) Stevens parameters from thermodynamic data of rare-earth magnetic materials. The algorithm employs a two-dimensional convolutional neural network (CNN) that is trained on magnetization, magnetic susceptibility and specific heat data that is calculated theoretically within the single-ion approximation and further processed using a standard wavelet transformation. We apply the method to crystal fields of cubic, hexagonal and tetragonal symmetry and for both integer and half-integer total angular momentum values J of the ground state multiplet. We evaluate its performance on both theoretically generated synthetic and previously published experimental data on $PrAgSb_2$ and $PrMg_2Cu_9$, and find that it can reliably and accurately extract the CF parameters for all site symmetries and values of J considered. This demonstrates that CNNs provide an unbiased approach to extracting CF parameters that avoids tedious multi-parameter fitting procedures.

I. INTRODUCTION

Rare-earth magnets often exhibit rich magnetic behaviors as a result of various competing energy scales that include spin-orbit coupling, crystal field (CF) and Zeeman energies as well as magnetic exchange interactions [1–4]. CFs play an important role as they cause magnetocrystalline anisotropies and in many cases determine the level degeneracies of the localized f electron states [5–9]. This strongly influences thermodynamic observables such as the magnetization, magnetic susceptibility, and specific heat [10], but it can also have important ramifications for the nature of the Kondo effect in the system [11–14].

CFs arise from time-reversal-even interactions between electrons (in f orbitals for rare-earth elements) and charges in their crystalline environment and are conveniently described by an effective electrostatic potential. The form of the CF potential is dictated by the point symmetry at the rare-earth site and contains a variable number of independent parameters [6, 9, 15]. For example, while the CF potential for f electrons is fully described by only two independent parameters for the cubic point groups $\mathcal{G} = m\bar{3}m, 432, \bar{4}3m$, there are 26 independent parameters for the lowest symmetry groups 1 and $\bar{1}$ [16, 17]. These CF parameters are notoriously difficult to determine in first-principle calculations [18], and are therefore best regarded as phenomenological parameters that are found from comparison to experimental results. While most accurate values of CF parameters are obtained from analyzing inelastic neutron scattering results [19, 20], much insight can already be gained by much more straightforward measurements of thermodynamic observables such as the (magnetic part of the)

specific heat $c_M(T)$ as a function of temperature T, the magnetic susceptibility $\chi_a(T)$ along direction a, and the magnetization $\mu_a(\boldsymbol{B}, T)$ in a finite magnetic field $\boldsymbol{B}$. This approach allows investigating whole series of rare-earth compounds, which often provides a more complete understanding of a material class, as was demonstrated, for example, in Refs. [21–23].

Here, we focus on the method of extracting CF parameters from thermodynamic measurements that are performed in a regime above possible Kondo and magnetic ordering temperatures, where the rare-earth ion can be treated within the single-ion approximation [8, 9]. We will also assume that the Russell-Saunders approximation is valid and spin-orbit coupling is stronger than CF, Zeeman and magnetic exchange energy scales: $E_{\text{Coulomb}} \gg E_{\text{SOC}} \gg E_{\text{CF}}, E_{\text{Zeeman}}, E_{\text{ex}}$. Note that we will further focus on the case where the CF and Zeeman energies are larger than the exchange energy: $E_{\text{CF}}, E_{\text{Zeeman}} \gg E_{\text{ex}}$. Here, E_{Coulomb} and E_{SOC} refer to the isotropic Coulomb and spin-orbit interaction between N electrons within the $4f^N$ electronic configuration of a single rare-earth ion, and $E_{\text{Zeeman}} = -\mu_B(\boldsymbol{L} + 2\boldsymbol{S}) \cdot \boldsymbol{B}$ with total orbital and spin angular momentum operators $\boldsymbol{L}$ and $\boldsymbol{S}$. Under these assumptions, one can restrict the attention to the ground state J multiplet of the $4f^N$ configuration that is derived from the three Hund's rules [10]. Its $2J + 1$ sub-levels are only degenerate for spherical symmetry and split in a crystalline environment into a sequence of lower order multiplets. While their multiplicity is fully determined by site symmetry, the energies of the different levels as well as their wave functions depend in general on the values of the CF parameters.

To obtain the CF parameters from measurements of thermodynamic observables, one traditionally proceeds as follows. Starting from an initial guess of the CF parameters, one determines the energy levels and wave

---

functions by diagonalizing the CF Hamiltonian $H_{\mathrm{CF}} = \sum_{q,k} \mathcal{B}_k^q \widetilde{C}_q^{(k)}(\boldsymbol{J})$. Here, the summation runs over a symmetry-allowed set of quantum numbers k and q with $0 \leq k \leq 2\ell$, $-k \leq q \leq k$ for a single-ion with orbital quantum number ℓ ($\ell = 3$ for f-electrons). The coefficients $\mathcal{B}_k^q$ are CF Stevens parameters and the CF operator "equivalents" $\widetilde{C}_q^{(k)}$ are expressed in terms of angular momentum operators $\boldsymbol{J}$ acting on the ground state J multiplet of the ion [5, 8, 9, 24, 25]. Various forms for the operators, which differ in their normalization convention, have been used in the literature and will be discussed below. Once the energies and wave functions are known, it is straightforward to calculate thermodynamic observables such as c_M, χ_a and μ_a from the partition function in finite magnetic field (details are shown below). The theoretical result is then compared to experiment and the complete procedure is iterated with updated CF parameters until sufficient agreement is reached.

While this iterative process is straightforward in principle, it can be tedious and time consuming in practice, in particular for lower than cubic symmetries, where several fit parameters need to be optimized simultaneously. This is complicated by the fact that the impact on the thermodynamic response that is caused by modifying the CF parameters $\{\mathcal{B}_k^q\}$ is in most cases unknown and not straightforward to derive. This is a typical example of an "inverse problem" [26] that often occurs in science, where one wants to estimate parameters p characterizing the system (here the CF parameters) based on observations O (in our case thermodynamic observables). Given a model P (for us, the crystal-field Hamiltonian), it is straightforward to derive observables $O = F_P(p)$, but the inverse mapping $p = F_P^{-1}(O)$ is difficult to perform, in particular when the relation is non-linear as in our case; often, the inverse mapping is ill-conditioned or unstable and, thus, requires regularization.

Motivated by the multitude of recent explorations of machine-learning (ML) techniques in physics [27–29], in general, and the success of artificial neural networks and other ML approaches to attack complex inverse problems of physics [30–34], in particular, we here study how ML can be used to extract Stevens CF parameters from thermodynamic measurements. This data-driven approach to inverse problems is based on first computing a large set of training data $\{(p_j, F_P(p_j))|j = 1, 2, \dots\}$, which requires solving the (simple) forward problem for many values of $p = p_j$. With this data set, a non-linear function is trained to reconstruct p_j from $O_j = F_P(p_j)$; the key challenge is to find a model that generalizes well for feasible training data sizes, i.e., that works on physically relevant samples that are not part of the original training set.

More specifically, we here employ a convolutional neural network (CNN) to parametrize the non-linear function performing the inverse operation: it relates thermodynamic observables, $O = \{c_M(T), \chi_a(T), \mu_a(\boldsymbol{B}, T)\}$, to a set of CF parameters $p = \{\mathcal{B}_k^q\}$. We train the CNN on thermodynamic data for different site symmetries (cubic

m$\bar{3}$m, hexagonal $\bar{6}$m2, tetragonal 4mm) and different values of angular momentum $J = 4$ and $J = 15/2$. This corresponds to the rare-earth ions Pr^{3+} ($J = 4$) and Er^{3+} ($J = 15/2$) in different crystalline environments. The training data is obtained within the single-ion approximation, and further processed using a standard wavelet transformation before being fed into the CNN. We test the performance of the CNN on both calculated and previously published experimental data on PrAgSb_2 [21] and $\mathrm{PrMg}_2\mathrm{Cu}_9$ [23]. We find that our CNN architecture generalizes well for moderately large training data sets and for all site symmetries and values of J considered. It also provides good estimates of the Stevens parameters from experimental data.

The remainder of the paper is organized as follows. In Sec. II, we review the single-ion approximation, define our notation of the Stevens CF parameters, and explain how the relevant thermodynamic observables are computed. Readers already familiar with this, can proceed directly to Sec. III, where we detail our proposed ML framework to estimate Stevens parameters from thermodynamic quantities. In Sec. IV and Sec. V, we demonstrate and test our ML approach on synthetic and experimental data, respectively, and Sec. VI provides a summary.

II. CRYSTAL FIELD THERMODYNAMICS IN RARE-EARTHS

In this section, we provide the necessary background to perform a quantitative analysis of CF effects on thermodynamic observables in rare-earth materials. We begin by describing the single-ion approximation, which assumes that interactions between different rare-earth ions are negligible. This approximation is often justified by the hierarchy of interactions that exist in rare-earth intermetallics [9]. Focusing on the ground state multiplet of a single-ion with a definite total angular momentum J, we show how to expand the CF Hamiltonian for a given J and point symmetry group $\mathcal{G}$ in terms of operator equivalents, as first introduced by Stevens [5].

Straightforward diagonalization of the Hamiltonian matrix together with elementary statistical mechanics calculations, then yield the thermodynamic observables, (i) specific heat c_M, (ii) magnetic susceptibility χ_a (along direction a), and (iii) magnetization μ_a in finite applied magnetic field B_a. This calculation explicitly shows the (forward) mapping from a set of CF parameters to thermodynamic observables. These thermodynamic observables are then fed into the input nodes of a CNN that "learns" the inverse mapping from the observables to the CF parameters as output.

A.   Single-ion approximation

In the single-ion approximation one neglects the interaction between different rare-earth ions, which is often justified because the $4f$ electrons are strongly localized. This leads to a relative weakness of $4f$-$4f$ exchange interactions compared to $3d$-$3d$ and $3d$-$4f$ interactions [9], and an often weak hybridization between the localized $4f$ electrons and delocalized conduction electrons. The single-ion description breaks down, for example, when Kondo or Rudermann-Kittel-Kasuya-Yosida (RKKY) interactions play an important role in the magnetism of the system. Our analysis in the following is therefore restricted to parameter regimes, where both Kondo and RKKY interactions are weak effects, which is typically the case at not too low temperatures $T \gg T_\mathrm{K}, T_\mathrm{RKKY}$, where T_K (T_RKKY) refer to Kondo and RKKY temperatures scales.

In the single-ion approximation, one describes the $4f$ electronic part of the system by a non-interacting collection of Hamiltonians for single rare-earth ions in a $4f^N$ configuration which each take the form [8, 9]

$$H_{4f} = H_\mathrm{Coulomb} + H_\mathrm{SOC} - \mu_B(\boldsymbol{L} + 2\boldsymbol{S}) \cdot \boldsymbol{B}$$
$$+ \sum_{i=1}^{N} V_\mathrm{CF}(r_i, \theta_i, \phi_i) \,. \qquad (2.1)$$

Here, H_Coulomb and H_SOC describe the isotropic Coulomb and spin-orbit interactions among the N $4f$ electrons, which are the dominant energy scales. They enforce the three Hund's rules in the $4f^N$ configuration of the rare-earth ion, $S = \frac{1}{2}(2\ell+1 - |2\ell+1-N|)$, $L = S(2\ell+1-2S)$, and $J = L \pm S$. The resulting ground state is then a $2J+1$ degenerate multiplet. Here, $\ell = 3$ is the orbital angular momentum of a single f electron, S (L) are the total spin (orbital) angular momentum quantum numbers and J is the total angular momentum quantum number. The third Hund's rule enforces $J = L + S$ for more than half-filled $4f$ shells, $N \geq 2\ell + 1$ [8, 10].

The third term in Eq. (2.1) describes the Zeeman coupling to an external magnetic field $\boldsymbol{B}$, where μ_B is the Bohr magneton and $\boldsymbol{L} = \sum_{i=1}^{N} \boldsymbol{l}_i$ and $\boldsymbol{S} = \sum_{i=1}^{N} \boldsymbol{s}_i$ denote total orbital and spin angular momenta of the N electrons in the $4f^N$ configuration. In the following, we will assume that spin-orbit coupling dominates over Zeeman energy and use the Russell-Saunders LS-coupling scheme to express the Zeeman Hamiltonian using the total angular momentum $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$ as

$$H_\mathrm{Zeeman} = -\mu_B g_{JLS} \boldsymbol{J} \cdot \boldsymbol{B} \,. \qquad (2.2)$$

Here, we have introduced the g-factor

$$g_{JLS} = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \,. \qquad (2.3)$$

with angular momentum quantum numbers J, L, S corresponding to the magnitude of the operators $\boldsymbol{J}, \boldsymbol{L}, \boldsymbol{S}$, respectively.

Finally, the last term in Eq. (2.1) denotes the CF potential, which can be expanded in a series of (single-particle) irreducible tensor operators as [8]

$$V_\mathrm{CF}(r, \theta, \phi) = \sum_{k=2,4,6} \sum_{q=-k}^{k} B_k^q(r) C_q^{(k)}(\theta, \phi) \,. \qquad (2.4)$$

Here, the functions $B_k^q(r)$ depend on the radial coordinate only, and $C_q^{(k)}(\theta, \phi) = \sqrt{\frac{4\pi}{2k+1}} Y_k^q(\theta, \phi)$ are related to the spherical harmonics. Both sets of operators, $B_k^q(r)$ and $C_q^{(k)}(\theta, \phi)$, act on the coordinates $\boldsymbol{r}_i$ of individual electrons in the f-shell. Note that the summation of k is restricted to $k = 2, 4, 6$, as we anticipate to evaluate matrix elements of V_CF only within a single $4f^N$ configuration. This excludes odd values of k by parity considerations. Higher values of $k > 6$ are excluded from the triangular condition $k \leq 2\ell$ of the Clebsch-Gordon coefficients (or Wigner $3j$ symbols), which arise when performing an integration over products of three spherical harmonics [35]. Finally, we have also excluded the $k = 0$ term as it amounts to an unimportant constant energy shift.

B.   Operator equivalents in crystal field Hamiltonians

The evaluation of matrix elements of the CF Hamiltonian

$$H_\mathrm{CF} = \sum_{i=1}^{N} \sum_{k=2,4,6} \sum_{q=-k}^{k} B_k^q(r_i) C_q^{(k)}(\theta_i, \phi_i) \qquad (2.5)$$

in the limited subspace of a $4f^N$ electronic configuration of a single rare-earth ion is made easier by the method of operator equivalents introduced by Stevens [5]. First, within a fixed $4f^N$ manifold, the radial operators can be replaced by their expectation values in the $4f$ states, which defines the (single-particle) Stevens coefficients $B_k^q \equiv \langle B_k^q(r_i) \rangle_{4f}$. Since the precise form of the wavefunction is difficult to determine, a theoretical calculation of the Stevens coefficients from first-principles is notoriously challenging [18]. The B_k^q are therefore best regarded as phenomenological coefficients that are obtained from a comparison of calculated physical observables to experimental data.

The method of operator equivalents [5, 7, 36–38] relates matrix elements of (the sum over) irreducible tensor operators within a $4f^N$ configuration to matrix elements of expressions that depend on angular momentum operators $\boldsymbol{l}_i$:

$$\langle \{l_i, m_i\} | \sum_{i=1}^{N} C_q^{(k)}(\theta_i, \phi_i) | \{l_i, m_i'\} \rangle$$
$$= \mathfrak{a}_k \langle \{l_i, m_i\} | \sum_{i=1}^{N} \widetilde{C}_q^{(k)}(\boldsymbol{l}_i) | \{l_i, m_i'\} \rangle \,. \qquad (2.6)$$

Here, $\mathfrak{a}_k$ is an k (and l_i) dependent coefficient. The operator expressions on both sides transform under rotation according to the same irreducible representation of the continuous rotation group. This condition in fact defines the "operator equivalent" of the irreducible tensor operator on the left-hand side. The operator equivalents $\widetilde{C}_q^{(k)}(\boldsymbol{l}_i)$ can be obtained by converting the functions $C_q^{(k)}(\theta, \phi)$ into Cartesian coordinates, $(x, y, z) = (\sin\theta\sin\phi, \sin\theta\cos\phi, \cos\theta)$, symmetrizing monomials (e.g., $xy \to (xy + yx)/2$), and replacing $(x_i/r_i, y_i/r_i, z_i/r_i) \to (l_x, l_y, l_z)_i$. The proportionality of the matrix elements in Eq. (2.6) relies on the fact that the rotation group is continuous and, loosely speaking, matrix elements for any point on the sphere can thus be obtained from those at a single, fixed point via rotation. The proportionality factors $\mathfrak{a}_k$ essentially account for the difference of the matrix elements at the single reference point. The $\mathfrak{a}_k$ are independent of q due to the Wigner-Eckart theorem [35]. In the literature, the proportionality coefficients are typically denoted as $\mathfrak{a}_2 = \alpha_l$, $\mathfrak{a}_4 = \beta_l$, and $\mathfrak{a}_6 = \gamma_l$. For $l_i = 3$ corresponding to $4f$ rare-earth ions, one finds the values $\mathfrak{a}_2 = -2/45$, $\mathfrak{a}_4 = 2/495$, and $\mathfrak{a}_6 = -4/3861$ [6, 9, 25].

Here, we restrict our analysis to the $(2J + 1)$-dimensional ground state multiplet of the $4f^N$ configuration that obeys the three Hund's rules. Combining individual angular momenta to the total orbital angular momentum $\boldsymbol{L} = \sum_{i=1}^N \boldsymbol{l}_i$ and considering spin-orbit coupling within a fixed LS term, leading to total angular momentum $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, one can derive a similar "operator equivalent" relation as Eq. (2.6) for matrix elements taken within a particular J multiplet

$$\langle L, S, J, M_J| \sum_{i=1}^N C_q^{(k)}(\theta_i, \phi_i)|L, S, J, M_J'\rangle$$
$$= \mathfrak{b}_k \langle L, S, J, M_J|\widetilde{C}_q^{(k)}(\boldsymbol{J})|L, S, J, M_J'\rangle . \quad (2.7)$$

Here, the coefficients $\mathfrak{b}_k$ depend on k as well as on the quantum numbers l_i, L, S, J. Like the $\mathfrak{a}_k$, they are independent of m due to the Wigner-Eckart theorem. The values of the $\mathfrak{b}_k$ for the ground state multiplets of the R^{3+} rare-earth ions can be found in the literature, where they are commonly denoted as $\mathfrak{b}_2 = \alpha_J = \theta_2$, $\mathfrak{b}_4 = \beta_J = \theta_4$ and $\mathfrak{b}_6 = \gamma_J = \theta_6$ [8, 9].

Using the operator equivalence in Eq. (2.7), one can thus express the CF Hamiltonian acting within the $(2J+1)$-dimensional ground state multiplet as

$$H_{\mathrm{CF}} = \sum_{k=2,4,6} \sum_{q=-k}^{k} \mathcal{B}_k^q \widetilde{C}_q^{(k)}(\boldsymbol{J}) . \quad (2.8)$$

Here, we have introduced the Stevens coefficients $\mathcal{B}_k^q = \mathfrak{b}_k B_k^q$ that depend on the radial expectation values through $B_k^q \equiv \langle B_k^q(r_i)\rangle_{4\mathrm{f}}$ (see Eq. (2.4)). We regard both B_k^q and $\mathcal{B}_k^q$ as phenomenological parameters that are determined by comparing theoretical calculations of physical observables to experimental results.

In the following, we will use the CF Hamiltonian of the from in Eq. (2.8). We note that in the literature it is common to use the so-called Stevens operator equivalents $O_k^q(\boldsymbol{J})$ [9, 16, 25, 39], which are based on the tesseral harmonics (real and imaginary parts of the spherical harmonics). The Stevens operators employ a different normalization convention than the irreducible tensor operator equivalents $\widetilde{C}_q^{(k)}(\boldsymbol{J})$. This requires using k and q dependent factors K_k^q relating the $\widetilde{C}_q^{(k)}$ and O_q^q operators: $O_k^q = \frac{1}{K_k^q}\frac{2k+1}{4\pi}\frac{1}{\sqrt{2}}\big[\widetilde{C}_{-q}^{(k)} + (-1)^q\widetilde{C}_q^{(k)}\big]$ for $q \neq 0$ and $O_k^0 = \frac{1}{K_k^0}\frac{2k+1}{4\pi}\widetilde{C}_q^{(k)}$ [39]. The factors K_k^q can be found, for example, in Ref. [39], but can also be easily derived by direct comparison of the operator matrices [24, 25]. The main disadvantage of the Stevens operators $O_k^q(\boldsymbol{J})$ is that they do *not* obey the Wigner-Eckart theorem. Their matrix elements are explicitly tabulated [25, 37, 40].

C. Stevens crystal field parameters

In this section, we describe the convention of Stevens parameters that we use in the following and their relation to other common definitions. Following Lea, Leask, Wolf [16] and Walter [17], it is convenient to perform a transformation from the Stevens parameters $\{\mathcal{B}_k^q\}$ in Eq. (2.8) to a set of Stevens coefficients $\{x_0, \ldots, x_{N_{\mathrm{St}}-1}\}$. Here, x_0 describes the overall energy scale of the CF splitting (note that x_0 can be negative). The dimensionless parameters $\{x_1, \ldots, x_{N_{\mathrm{St}}-1}\}$ fulfill $|x_i| \leq 1$ and describe the relative weight of the different Stevens parameters $\mathcal{B}_k^q$.

1. Cubic symmetry

Let us explicitly describe the transformation from $\{\mathcal{B}_k^q\} \to \{x_i\}$ for cubic symmetry. The derivation easily generalizes to arbitrary point groups $\mathcal{G}$. For the cubic point groups $\mathcal{G} = \{\mathrm{m\bar{3}m}, 432, \bar{4}3\mathrm{m}\}$, the CF Hamiltonian contains two independent Stevens parameters, $N_{\mathrm{St}} = 2$, and reads

$$H_{\mathrm{CF}} = \mathcal{B}_4^4\Big(\widetilde{C}_4^{(4)} + \widetilde{C}_{-4}^{(4)} + \sqrt{\frac{14}{5}}\widetilde{C}_0^{(4)}\Big)$$
$$+ \mathcal{B}_6^4\Big(\widetilde{C}_4^{(6)} + \widetilde{C}_{-4}^{(6)} - \sqrt{\frac{2}{7}}\widetilde{C}_0^{(6)}\Big) . \quad (2.9)$$

Let us first normalize each operator that multiplies a particular Stevens coefficient

$$\widetilde{\mathcal{O}}^{(4)} = \widetilde{C}_4^{(4)} + \widetilde{C}_{-4}^{(4)} + \sqrt{\frac{14}{5}}\widetilde{C}_0^{(4)} \quad (2.10)$$

$$\widetilde{\mathcal{O}}^{(6)} = \widetilde{C}_4^{(6)} + \widetilde{C}_{-4}^{(6)} - \sqrt{\frac{2}{7}}\widetilde{C}_0^{(6)} \quad (2.11)$$

Normalization of $\widetilde{\mathcal{O}}^{(4)}$ and $\widetilde{\mathcal{O}}^{(6)}$ can be achieved by dividing by the sum of squared eigenvalues $\Lambda^{(k)} =$

$\sqrt{\sum_{i=1}^{2J+1}|\lambda_i^{(k)}|^2}$, where $\lambda_i^{(k)}$ are the eigenvalues of the operator $\widetilde{\mathcal{O}}^{(k)}$. Specifically, we define the scaled operators

$$O^{(k)} = \frac{2J+1}{\Lambda^{(k)}}\widetilde{\mathcal{O}}^{(k)} \tag{2.12}$$

and express the CF Hamiltonian as

$$\begin{aligned}
H_{\mathrm{CF}} &= \mathcal{B}_4^4\widetilde{\mathcal{O}}^{(4)} + \mathcal{B}_6^4\widetilde{\mathcal{O}}^{(6)} \\
&= x_0\Big[x_1\mathcal{O}^{(4)} + (|x_1|-1)\mathcal{O}^{(6)}\Big].
\end{aligned} \tag{2.13}$$

As anticipated above, the scale parameter x_0, which can be positive or negative, sets the overall energy scale of the CF splitting. The dimensionless weight parameter x_1 describes the ratio of Stevens coefficients

$$\frac{\mathcal{B}_4^4}{\mathcal{B}_6^4} \propto \frac{x_1}{|x_1|-1}, \tag{2.14}$$

and lies in the interval $-1 \le x_1 \le 1$. The ratio $\mathcal{B}_4^4/\mathcal{B}_6^4 = 0$ corresponds to $x_1 = 0$, whereas $\mathcal{B}_4^4/\mathcal{B}_6^4 \to \pm\infty$ corresponds to $x_1 = \pm1$. The exact relation between the two sets of Stevens parameters $\{\mathcal{B}_k^q\}$ and $\{x_i\}$ depends on the value of J and can b easily rederived from Eq. (2.13).

2. Hexagonal symmetry

For hexagonal site symmetry with point symmetry groups $\mathcal{G} = \{\bar{6}m2, 6/mmm, 6mm, 622\}$, the CF Hamiltonian contains four independent Stevens parameters, $N_{\mathrm{St}} = 4$, and reads

$$H_{\mathrm{CF}} = \mathcal{B}_2^0\widetilde{C}_0^{(2)} + \mathcal{B}_4^0\widetilde{C}_0^{(4)} + \mathcal{B}_6^0\widetilde{C}_0^{(6)} + \mathcal{B}_6^6\big(\widetilde{C}_6^{(6)} + \widetilde{C}_{-6}^{(6)}\big). \tag{2.15}$$

Defining $\widetilde{\mathcal{O}}_q^{(k)} = \widetilde{C}_q^{(k)}$ for $m \ne 6$ and $\widetilde{\mathcal{O}}_6^{(6)} = \widetilde{C}_6^{(6)} + \widetilde{C}_{-6}^{(6)}$ for $l = m = 6$, we again normalize the $\widetilde{\mathcal{O}}_q^{(k)}$ via

$$\mathcal{O}_q^{(k)} = \frac{2J+1}{\Lambda_q^{(k)}}\widetilde{\mathcal{O}}_q^{(k)}. \tag{2.16a}$$

Here,

$$\Lambda_q^{(k)} = \sqrt{\sum_{i=1}^{2J+1}|\lambda_{q,i}^{(k)}|^2}, \tag{2.16b}$$

where $\lambda_{q,i}^{(k)}$ are the eigenvalues of $\widetilde{\mathcal{O}}_q^{(k)}$. Finally, we express the CF Hamiltonian in terms of the normalized operators as

$$\begin{aligned}
H_{\mathrm{CF}} = |x_0|\Big[&x_1\mathcal{O}_0^{(2)} + x_2\mathcal{O}_0^{(4)} + x_3\mathcal{O}_6^{(6)} \\
&+ \mathrm{sign}(x_4)\Big|1 - |x_1| - |x_2| - |x_3|\Big|\mathcal{O}_0^{(6)}\Big].
\end{aligned} \tag{2.17}$$

As before, x_0 describes the overall energy scale, whereas the weight parameters $-1 \le x_1, \ldots x_3 \le 1$ describe the relative weight of the four Stevens parameters $\mathcal{B}_k^q$. Note that we have split off the sign of x_0. This turns out to be advantageous in the ML calculation described below as it makes the overall scale prefactor $|x_0|$ strictly positive. This comes at the cost of introducing an additional parameter, x_4, defined as $\mathrm{sign}(x_4) := \mathrm{sign}(x_0)$. Only the sign of x_4 enters the Hamiltonian.

3. Tetragonal symmetry

For tetragonal site symmetry with point symmetry groups $\mathcal{G} = \{4mm, 4/mmm\}$, the CF Hamiltonian contains five independent Stevens parameters, $N_{\mathrm{St}} = 5$, and reads

$$\begin{aligned}
H_{\mathrm{CF}} &= \mathcal{B}_2^0\widetilde{C}_0^{(2)} + \mathcal{B}_4^0\widetilde{C}_0^{(4)} + \mathcal{B}_4^4\big(\widetilde{C}_4^{(4)} + \widetilde{C}_{-4}^{(4)}\big) + \mathcal{B}_6^0\widetilde{C}_0^{(6)} \\
&\quad + \mathcal{B}_6^4\big(\widetilde{C}_4^{(6)} + \widetilde{C}_{-4}^{(6)}\big) \\
&= \mathcal{B}_2^0\widetilde{\mathcal{O}}_0^{(2)} + \mathcal{B}_4^0\widetilde{\mathcal{O}}_0^{(4)} + \mathcal{B}_4^4\widetilde{\mathcal{O}}_4^{(4)} + \mathcal{B}_6^0\widetilde{\mathcal{O}}_0^{(6)} + \mathcal{B}_6^4\widetilde{\mathcal{O}}_4^{(6)}.
\end{aligned} \tag{2.18}$$

We have defined the operators $\widetilde{\mathcal{O}}_q^{(k)}$ in the second line, which we then normalize as in Eq. (2.16).

Finally, the CF Hamiltonian is expressed in terms of the normalized operators as

$$\begin{aligned}
H_{\mathrm{CF}} = |x_0|\Big[&x_1\mathcal{O}_0^{(2)} + x_2\mathcal{O}_4^{(4)} + x_3\mathcal{O}_0^{(4)} + x_4\mathcal{O}_4^{(6)} \\
&+ \mathrm{sign}(x_5)\Big|1 - |x_1| - |x_2| - |x_3| - |x_4|\Big|\mathcal{O}_0^{(6)}\Big].
\end{aligned} \tag{2.19}$$

In addition to the scale parameter x_0, the Hamiltonian contains four bounded Stevens parameters $-1 \le x_1, \ldots, x_4 \le 1$. Like in the hexagonal case, we have split off the sign of x_0 explicitly and introduced an additional parameter, x_5, as $\mathrm{sign}(x_5) := \mathrm{sign}(x_0)$. The Hamiltonian only depends on the sign of x_5.

D. Thermodynamic observables

In this section, we describe how to obtain the thermodynamic observables of interest: magnetization (per rare-earth ion), $\boldsymbol{\mu}(T, \boldsymbol{B})$, in finite magnetic field, magnetic susceptibility $\chi_a(T)$ along direction a, and specific heat $c_M(T)$. We calculate these quantities starting from the Hamiltonian (2.1) of a single rare-earth ion in a magnetic field $\boldsymbol{B}$ and exposed to a CF with point symmetry $\mathcal{G}$. From our discussion above, we know that the Hamiltonian projected onto the ground state multiplet with total angular momentum J reads

$$H_J = -\mu_B g_{JLS}\boldsymbol{J}\cdot\boldsymbol{B} + \sum_{l=2,4,6}\sum_{m=-l}^{l}\mathcal{B}_k^q\widetilde{C}_q^{(k)}(\boldsymbol{J}). \tag{2.20}$$

Here, $\boldsymbol{J} = (J_x, J_y, J_z)$ denotes the total angular momentum operator, the g-factor g_{JLS} is explicitly given in Eq. (2.3), and the form of the CF Hamiltonian is constrained by the point group $\mathcal{G}$. The method we describe in the following can be used for any point group $\mathcal{G}$, but we will focus on the experimentally common cases of cubic, hexagonal and tetragonal crystal symmetry with point groups $\mathcal{G}$ that were discussed in Sec. II C.

In the basis of J_z eigenstates, $J_z |m_J\rangle = m_J |m_J\rangle$, the Hamiltonian H_J is a $(2J+1) \times (2J+1)$ dimensional matrix that can be easily diagonalized,

$$H_J |n\rangle = E_n |n\rangle, \qquad (2.21)$$

with field-dependent energies E_n and eigenstates

$$|n\rangle = \sum_{m_J=-J}^{J} a_{m_J} |m_J\rangle. \qquad (2.22)$$

In Fig. 1(a), we show the resulting energy spectrum for $J = 4$ and cubic CF ($\mathcal{G} = \mathrm{m\bar{3}m}$) as a function of the dimensionless Stevens parameter x_1, see Eq. (2.13). The level spectrum is presented both at zero magnetic field and at a finite field of $B_z = 1$ T. In zero field, we observe that for $x_0 > 0$, the ground state is a singlet (triplet) for $x_1 < 0.34$ ($x_1 > 0.34$). For $x_0 < 0$, the ground state is a triplet for $x_1 < -0.57$, a doublet for $-0.57 < x_1 < 0.74$ and a singlet for $x_1 > 0.74$. The level degeneracy is split in an applied magnetic field and one observes the emergence of several (avoided) level crossings. A similar behavior is observed for other integer values of the angular momentum quantum number J with lower degeneracies in the case of lower-symmetry CFs. For half-integer J, the levels are at least doubly degenerate in the absence of an external magnetic field due to Kramers theorem.

1.   Magnetization and magnetic susceptibility

The magnetization per single rare-earth ion along direction a is given by

$$\mu_a(T, \boldsymbol{B}) = \frac{\mu_B g_{JLS}}{Z} \sum_n \langle n|J_a|n\rangle\, e^{-E_n/k_B T} \qquad (2.23)$$

with partition function $Z = \mathrm{Tr}\, e^{-\beta H_J}$ and magnetic field along direction a. In Fig. 1(b), we show the magnetization μ_z as a function of $\boldsymbol{B} = (0,0,B)^T$ for different temperatures T in a cubic CF ($\mathcal{G} = \mathrm{m\bar{3}m}$). The Stevens parameters are chosen to be $x_0 = 20$ K and $x_1 = 0.5$ such that the ground state is a triplet with $\langle J_z \rangle = \pm\frac{5}{2}, 0$. The magnetization μ_z thus increases linearly at low fields with a slope that increases Curie-like as $1/T$. The magnetization saturates at a saturation magnetic field value B_{sat} that increases with temperature. At the lowest temperature, $T = 1$ K, the saturation occurs at $B_{\mathrm{sat}}(1\text{ K}) \simeq 3$ T. The saturation value of the magnetization is given by $\mu_z^{\mathrm{sat}} = \mu_B g_{JLS} \langle 0|J_z|0\rangle$, where $|0\rangle$ is the ground state in

magnetic field. Here, we have chosen $L = 5$ and $S = 1$ such that $g_{JLS} = 4/5$, which corresponds to the rare-earth ion Pr^{3+}. Note that μ_z^{sat} deviates slightly from the expected value of $5/2 \times 4/5 = 2$, where $5/2$ is the expectation value of J_z in the triplet ground state in small fields, due to field induced mixing of higher levels.

The magnetic susceptibility is obtained at small magnetic fields from the slope

$$\chi_a(T) = \mu_a(T, B_a)/B_a. \qquad (2.24)$$

Its behavior at low temperatures is determined by the ground state degeneracy [10]. If the ground state is a singlet, it is of van-Vleck type, $\chi_a \propto \sum_{i \neq 0} \frac{|\langle i|J_a|0\rangle|^2}{E_i - E_0}$, and becomes temperature independent at temperatures much smaller than the energy gap to the first excited state, $k_B T \ll E_1 - E_0$. In contrast, χ_a is Curie-like $\chi_a \propto \frac{g_{JLS}^2 \mu_B^2 \langle 0|J^2|0\rangle}{T}$, if the ground state degeneracy is larger than one. In Fig. 1(c), we show the susceptibility χ_z as a function of temperature for the case of a triplet ground state, where it follows a characteristic Curie-like behavior $\chi_a \propto 1/T$ [see inset of Fig. 1(c)].

2.   Specific heat

The specific heat in zero magnetic field is calculated from

$$c_M(T) = \frac{1}{k_B T^2} \Big(\langle H_{\mathrm{CF}}^2 \rangle - \langle H_{\mathrm{CF}} \rangle^2 \Big), \qquad (2.25)$$

where the average is performed with respect to the CF eigenstates $H_{\mathrm{CF}} |n^{(0)}\rangle = E_n^{(0)} |n^{(0)}\rangle$:

$$\langle \mathcal{O} \rangle = \frac{1}{Z_{\mathrm{CF}}} \sum_n \langle n^{(0)}|\mathcal{O}|n^{(0)}\rangle e^{-E_n^{(0)}/k_B T} \qquad (2.26)$$

with $Z_{\mathrm{CF}} = \mathrm{Tr}\, e^{-H_{\mathrm{CF}}/k_B T} = \sum_n e^{-E_n^{(0)}/k_B T}$. The specific heat vanishes exponentially, $c_M \propto \left(\frac{\Delta}{k_B T}\right)^2 e^{-\Delta/k_B T}$, at temperatures below the gap to the first excited state, $k_B T \ll \Delta$. As shown in Fig. 1(c) for $J = 4$ and $\mathcal{G} = \mathrm{m\bar{3}m}$, it exhibits a Schottky anomaly peak at higher temperatures, whose position and weight yields direct information about the size of the gap to the excited states and the relative degeneracies of the ground and excited state levels. Note that excited state levels higher than the first often occur nearby in energy and thus contribute to the specific heat as well.

III.   CNN APPROACH FOR FINDING CRYSTAL FIELD PARAMETERS

In this section, we describe the method of using a two-dimensional CNN to determine the Stevens parameters $\{x_i\}$ for a given angular momentum J and CF symmetry group $\mathcal{G}$ from thermodynamic observables. Our goal is
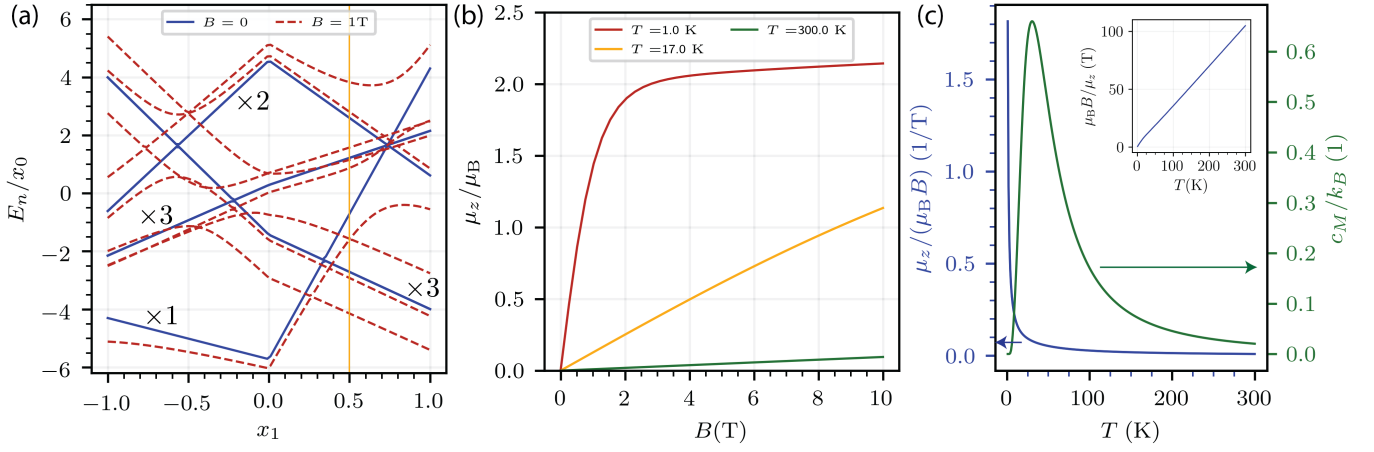
FIG. 1. (a) Energy levels E_n/x_0 of a $J = 4$ rare-earth ion (Pr^{3+}) in cubic CF with point symmetry $\mathcal{G} = m\bar{3}m$ as a function of Stevens parameter x_1. The degeneracy of the levels in the absence of a magnetic field is indicated in the figure, and splits in applied magnetic field B along the z axis. The g-factor is set to $g_{JLS} = 4/5$. The vertical yellow line indicates the parameter choice, $x_1 = 0.5$, for panels (b, c), where we further set $x_0 = 20$ K as absolute energy scale. (b) Magnetization per rare-earth ion μ_z/μ_B as a function of applied magnetic field B along the z-axis for different temperatures. The Stevens parameters are $x_0 = 20$ K and $x_1 = 0.5$. The fully saturated moment is reduced from the value in the ground multiplet at $x_1 = 0.5$ is $\mu_z/\mu_B = 5/2$ due to field-induced mixing into other states. (c) Magnetic susceptibility along z-axis (per rare-earth ion) $\chi_z/\mu_B = \mu_z/(\mu_B B)$ (blue, left y-axis) and specific heat c_M/k_B (green, right y-axis) as a function of temperature T. The magnetic field is fixed to $B = 10^{-4}$ T when computing χ_z. The Stevens parameters are identical to panel (b). The inset shows the inverse susceptibility $\chi_z^{-1} = \mu_B B/\mu_z$, highlighting the Curie-like behavior that occurs over the full temperature range due to the triplet ground state. The specific heat c_M/k_B shows a Schottky-like peak at a position proportional to the level splitting between the ground multiplet and the excited states $\Delta = 40$ K. Note that here the first excited state is a singlet and the contribution of the next higher triplet level (at about 80 K) is significant.

to build a ML model that can be fed with experimental data and accurately predict the underlying Stevens coefficients that characterize the material, thereby circumventing a time-consuming data fitting procedure. One therefore places thermodynamic data on the input nodes of the network and obtains the set of Stevens coefficients as output. We choose the input data of the network to be from observables that are experimentally readily available: magnetization $\mu_a(T, B_a)$, magnetic susceptibility $\chi_a(T)$, and magnetic specific heat $c_M(T)$. To train the network, we require a sufficiently large dataset that we generate by calculating the thermodynamic observables for random choices of Stevens parameters within the single-ion approximation as described in Sec. II. Comparing different network architectures, we found it to be advantageous to perform a wavelet transformation on the data before feeding it into the network. In the following, we describe the details of the training data generation and the network architecture and parameters.

A. Training data generation

A training data set contains the following three types of observables, which are calculated for a fixed choice of angular momentum J and point group $\mathcal{G}$, and randomly sampled Stevens coefficients $\{x_i\}$, using the approach detailed in Sec. II:

(i) *Magnetization* per single rare-earth ion, μ_a, along direction a as a function of external magnetic field B_a for fixed temperature T. We choose a magnetic field range between $B_{\min} = 0$ T and $B_{\max} = 10$ T, and three temperatures $T_j = 1, 17, 300$ K to represent the behavior in typical field and temperature ranges which are easily accessible experimentally. We use $N_{\text{steps}}^B = 64$ equally spaced magnetic field points

$$B_a = B_{\min} + \frac{B_{\max} - B_{\min}}{N_{\text{steps}}^B - 1} b_a \qquad (3.1)$$

with $b_a \in \{0, \ldots, N_{\text{steps}}^B - 1\}$. Depending on the CF point symmetry group $\mathcal{G}$, we choose different high-symmetry directions—only one high-symmetry direction $a = [001] \equiv z$ (two high-symmetry directions, $a = \{[100], [001]\}$) for cubic (tetragonal and hexagonal) symmetry. Combined with the three temperature values T_j, this corresponds to three (six) sets of magnetization data: $\mu_a(B_a, T_j)$. Other choices of directions are of course possible, but we wanted to keep the size of the input data set as small as possible to keep the experimental work necessary to obtain it at a minimum.

(ii) *Magnetic susceptibility* $\chi_a(T)$ along direction a as a function of temperature T. We choose a temperature range between $T_{\min} = 1$ K and $T_{\max} = 300$ K using $N_{\text{steps}}^T = 64$ equally spaced temperature points

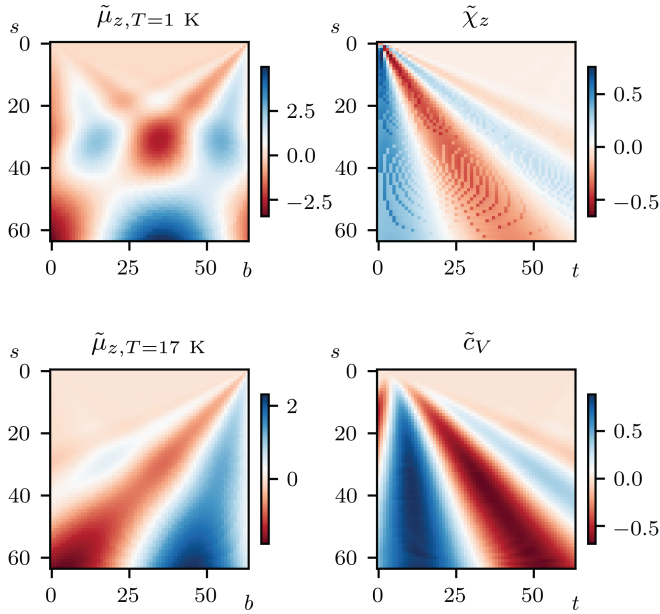$$T = T_{\min} + \frac{T_{\max} - T_{\min}}{N_{\text{steps}}^T - 1} t. \qquad (3.2)$$

FIG. 2.   Continuous wavelet scaleograms of the data of Fig. 1(b,c) for which $J = 4$, $\mathcal{G} = \mathrm{m\bar{3}m}$, $x_0 = 20$ K, and $x_1 = 0.5$. The plots show the CWT coefficients, using the real Morlet mother wavelet (3.4), as a function of scale s and physical index b (or t). We choose linearly spaced scales from $s_{\min} = 1$ to $s_{\max} = 64$, and also use 64 equally spaced points along the "physical" dimension, corresponding to the magnetic field, b (see Eq. (3.1)), and the temperature, t see Eq. (3.2)). In addition to these four scaleograms we also provide $\tilde{\mu}_{z,T=300\,\mathrm{K}}$ to the CNN, resulting in a total of five scaleograms to be layered into one training sample. For the lower symmetry point groups, we provide magnetization and susceptibility along both [100] and [001] directions, resulting in a total of nine scaleograms in one training sample.

We use the same directions a for the susceptibility and the magnetization, corresponding to one (or two) sets of susceptibility data.

(iii) *Magnetic specific heat* $c_M(T)$ as a function of temperature T. We use the same temperature range and step size as for the susceptibility, see Eq. (3.2).

One training sample therefore consists of five (nine) different sets of thermodynamic data. A complete training sample for $J = 4$, $\mathcal{G} = \mathrm{m\bar{3}m}$ and $x_0 = 20$ K, $x_1 = 0.5$ is shown in Fig. 1(b, c). To obtain the training data set, we draw the Stevens parameters randomly from a uniform distribution and, for each of these sampled values, compute the aforementioned observables. Note that within our convention there exist N_{St} CF parameters for the cubic and $N_{\mathrm{St}} + 1$ CF parameters for the hexagonal and tetragonal cases. While x_0 can take either sign in the cubic case, it is strictly positive for hexagonal and tetragonal systems. In the latter cases, only the sign of the last Stevens parameter $\mathrm{sign}(x_{N_{\mathrm{St}}})$ enters the Hamiltonian.

After comparing different network architectures (see more details below in Sec. III B), we have found it to be

advantageous to first perform a continuous wavelet transformation (CWT) of the "raw" thermodynamic data before feeding it into a two-dimensional (2D) CNN. The reason is that CNNs are well suited to model data with an image-like structure like the wavelet scaleograms that are produced by the CWT. Similar to a Fourier transform, a CWT is used to perform a harmonic analysis and decompose a signal into its fundamental frequencies. The advantage of a CWT is that it produces a sparse representation of the data by providing localization in both frequency and "time" domain, with the main features of the data appearing in only a (small) subset of all CWT coefficients. This property is key for applications in data compression and denoising [41]. We find that it also enables superior performance of a 2D CNN compared to placing a "raw" data vector of linear size $5 \times 64 = 320$ (or $9 \times 64 = 576$) on the network input nodes. Here, 5(9) corresponds to the number of thermodynamic observables and 64 to number $N_{\mathrm{steps}}^B = N_{\mathrm{steps}}^T$ of values of B_a and T, respectively.

A CWT of a discrete and equally spaced 1D data set of size N_f,

$$\{f_0, f_1, \ldots f_{N_f - 1}\} = \{f(t_{\min}), f(t_{\min} + \Delta), \ldots, f(t_{\max})\},$$

corresponds to performing the following transformation

$$\tilde{f}(t, s) = \frac{1}{\sqrt{s}} \sum_{i=0}^{N_f - 1} f_i \, \psi\left(\frac{(i - t)\Delta}{s}\right). \qquad (3.3)$$

Here, $\psi(t)$ is the so-called mother wavelet function, which is translated by parameter t and scaled by the scale parameter s. The scale s can be regarded as a period or inverse frequency. We choose a mother wavelet function of the real Morlet form [42]

$$\psi(t) = e^{-\frac{t^2}{2}} \cos(5t). \qquad (3.4)$$

We use a linearly spaced set of scales, $1 \le s \le 64$, which emphasizes the low frequency behavior of the data in comparison to using a geometric spacing. We have explicitly checked for a few cases that the CNN performs equally well if we use geometrically spaced scales. By calculating the convolution of this family of translated and scaled wavelets with our original data, we perform a frequency analysis that provides additional insight into the changes over the "time" domain, which corresponds to temperature or magnetic field in our data.

We apply a CWT to each of five (or nine) 1D observable data sets, $\{\mu_a(T, b_i), \chi_a(t_i), c_M(t_i)\}$, to produce a total of five (or nine) 2D wavelet scaleograms. In Fig. 2, we show four of the five scaleograms corresponding to the "raw" data in Fig. 1(b,c), which depict the associated (real) CWT coefficients $\tilde{f}$ defined in Eq. (3.3). It clearly distinguishes regions with small and large wavelet coefficients, which is a characteristic of the underlying "raw" data set. The peak position and characteristic power-law behavior of the original data, for example,
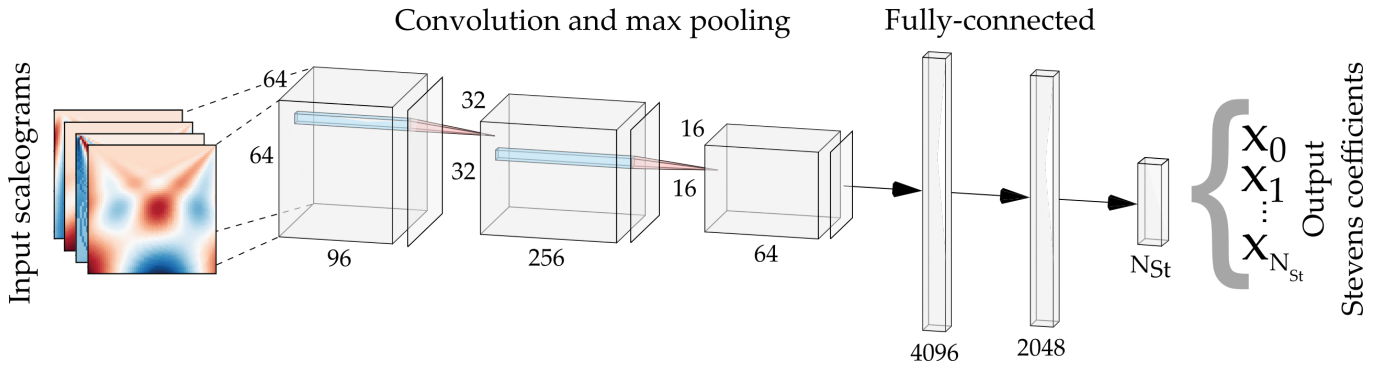
FIG. 3. Schematic architecture of the 2D CNN that is used in this work. The multi-channel input image comprises five (nine) stacked CWT scaleograms of thermodynamic observables. The image first passes through three convolution and max-pooling layers. This allows the network to extract features from the images. The two following fully connected layers then predict the $N_P + 1$ Stevens coefficients, $\{x_0, x_1, ..., x_{N_P}\}$ in the output nodes. Here, $N_P = N_{\rm St} - 1$ ($N_P = N_{\rm St}$) for the cubic (hexagonal, tetragonal) case.

$\mu_z \propto B, \chi_a \propto 1/T, c_M \propto 1/T^2$ that occurs in certain parameter regions, can also be clearly recognized in the scaleograms. All five (nine) scaleograms are stacked into a multi-channel image and then placed on the input nodes of a 2D CNN, whose architecture is described in the next subsection.

B.   Convolutional neural network

We employ a CNN architecture that is based off the LeNet-5 architecture [43], which we scale up to be appropriate for the form of our training dataset. Initial experimentation with alternative architectures, such as simple neural networks and 1D CNNs, yielded significantly worse results, leading to our decision to use a 2D CNN. As illustrated in Fig. 3, the input of the network is a five (nine) channel image containing the five (nine) scaleograms created using the CWT. We center the input data by subtracting the mean of each channel so that the distribution of the input "pixel" values has zero mean. We similarly normalize the target data $\{x_i\}$, as the coefficients x_0 and x_i with $i \neq 0$ have significant size differences and different dimensions. Typically, $|x_0| \in [0.5, 50]$ (in units of K), while the dimensionless $x_i \in [-1, 1]$ for $i > 0$.

The centered input CWT scaleograms are fed into three sets of convolution and max-pooling layers. Each convolution layer has two identical 2D convolution layers with a kernel size of 3×3 and a stride of 1×1. This increases the number of channels and allows extracting data features. The max-pooling layers have a pool-size and stride of 2×2, which essentially corresponds to a down-sampling of the image by a factor of two. Each convolution layer uses the ReLU activation function [44] and has batch normalization.

The final max-pooling layer is flattened and fed into two fully-connected layers, each with the ReLU activation function and 30% dropout. These layers feed into a fully-connected output layer whose width corresponds to the number of independent Stevens parameters $N_P + 1$. Here, $N_P = N_{\rm St} - 1$ ($N_P = N_{\rm St}$) for the cubic (hexagonal, tetragonal) case due to the procedure of splitting off the sign of x_0 into an additional parameter. The output layer uses a linear activation function. The $N_P + 1$ output values are the prediction of the CNN for the Stevens coefficients $\{x_0, \ldots, x_{N_{\rm P}}\}$,

The total number of trainable parameters in the network is $13,702,818 \simeq 1.3 \times 10^7$. This is substantially larger than the number of trainable parameters in the original LeNet-5 architecture [43]. This is due to the high dimensionality of the fully-connected layers in our version of the network. A larger number of parameters means that the network will be able to form more complex relationships between the features and the targets. However, we run the risk of over-parameterizing the network, resulting in a model that overfits—that is, it performs very well on the training data but poorly on unseen data. By applying normalization and dropout throughout the network we mitigate this issue. We build the network using Keras [45] and train it on Nvidia Volta V100S graphic processing unit (GPU). We use the Adam optimizer [46] with the recommended parameters to minimize the mean squared error (MSE) loss function,

$$C(\hat{\mathbf{x}}, \mathbf{x}) = \frac{1}{(N_P + 1)N_{\rm batch}} \sum_{j=1}^{N_{\rm batch}} \sum_{i=0}^{N_P} \left[\hat{x}_i(j) - x_i(j)\right]^2,$$
(3.5)

where $\hat{x}_i$ is the network's prediction for x_i. Using 10^5 training examples and 1.5×10^4 validation and 1.5×10^4 testing examples with a batch size of $N_{\rm batch} = 64$, the network converged after around 100 epochs. With the available GPU, training the network took around 70 seconds per epoch.

IV. CNN RESULTS

In this section, we present results and measure the performance of CNNs predicting Stevens coefficients for three different point groups. We choose groups in cubic, hexagonal and tetragonal crystal classes that are of experimental relevance: m$\bar{3}$m, $\bar{6}$m2 and 4mm. These groups allow for $N_{\text{St}} = 2, 4, 5$ independent Stevens parameters, respectively. The complexity of the task to find Stevens parameters from thermodynamic data increases when lowering the symmetry. We consider both integer and half-integer values of the total angular momentum quantum number J, and find that our method works equally well in both cases. For concreteness, we investigate $J = 4$ and $J = 15/2$, which correspond to the ground state values of the rare-earth ions Pr^{3+} ($J = 4$) and Er^{3+} ($J = 15/2$).

For a given point group $\mathcal{G}$ and value of J, we train a CNN using the training data described in Sec. III A. The input thus corresponds to five (nine) thermodynamic observables for cubic (hexagonal and tetragonal) point groups, and the output of the network are the $N_P + 1$ Stevens parameters, $\{x_0, \ldots, x_{N_P}\}$. We use two performance measures: (i) the *mean absolute error* (MAE) of the network's prediction of the Stevens coefficients

$$\text{MAE}(i) = \frac{1}{N_{\text{test}}} \sum_{j=1}^{N_{\text{test}}} |\hat{x}_i(j) - x_i(j)|. \quad (4.1)$$

Here, $\hat{x}_i$ is the prediction of the network, x_i is the true Stevens coefficient that is used to generate the data placed on the input nodes, and N_{test} is the size of the testing dataset. The MAE is related to the loss function (3.5) used to train the network. (ii) The *mean squared error* (MSE) of two thermodynamic data sets $\{\mu_a, \chi_a, c_M\}$ generated by $\hat{x}_i$ and x_i, respectively:

$$\text{MSE}(\{x_i\}) = \frac{1}{M} \sum_{\nu=1}^{M} \left[\mathcal{O}_\nu(\{\hat{x}_i\}) - \mathcal{O}_\nu(\{x_i\}) \right]^2. \quad (4.2)$$

Here, $M = 5 \times 64 = 320$ ($9 \times 64 = 576$) for cubic (hexagonal, tetragonal) point groups is the length of the thermodynamic dataset and $\mathcal{O}_\nu$ runs over the five (nine) experimental observables $\{\mu_{a,T_\alpha}(b_i), \chi_a(t_i), c_M(t_i)\}$ as a function of temperature t_i and magnetic field b_i (see Sec. III A) that are obtained for a given choice of Stevens parameters.

The MSE measures the performance of the CNN in reproducing the desired (input) thermodynamic data set that was generated using $\{x_i\}$. We include this metric as the sensitivity of the error in the observables (MSE) with respect to the error in the Stevens parameters (MAE) depends on the values of the $\{x_i\}$, and the MSE thus contains additional information about the networks performance. Unless otherwise noted, both MAE and MSE are evaluated on a testing data of size $N_{\text{test}} = 4 \times 10^3$ that was not shown to the network during training.

In the following, we separately discuss the performance of the CNNs for the cubic, hexagonal and tetragonal point groups.

A. Cubic point group symmetry

We consider the case of cubic point group $\mathcal{G} = $ m$\bar{3}$m and $J = 4$, which is applicable to cubic Pr rare-earth compounds. The energy level diagram for this case is shown in Fig. 1(a) and exhibits singlet, doublet and triplet ground states, depending on the sign of x_0 and the value of x_1. As shown in Fig. 4, the CNN accurately predicts the two independent Stevens coefficients $\{x_0, x_1\}$ with error values of MAE(0) = 0.321 K and MAE(1) = 0.012. Note that we choose the energy range of 0.5 K $\leq |x_0| \leq$ 50 K. The color code and the inset show the MSE, which lies at $\langle \text{MSE} \rangle = 2.7 \times 10^{-3}$ on average. The results in Fig. 4 show the predictions of two networks: one is trained with strictly positive $x_0 \in [0.5, 50]$, and a second one is trained with strictly negative $x_0 \in [-50, -0.5]$. When applied to a given testing example, which has a definite but unknown sign of x_0, the performance of the network that was trained on data with the same sign of x_0 as the testing example is typically much better and can be easily identified. Here, we show results for testing examples that have a known sign for simplicity, i.e., the positive (negative) network is tested on samples with positive (negative) x_0.

In Fig. 5, we visualize the distribution of MSE as a function of the two Stevens parameters x_0 and x_1. We clearly observe that the MSE is larger in regions where x_0 is small. This occurs as the energy level spectrum collapses in this limit, with all levels being smaller than (or comparable to) the minimal thermal energy $\simeq k_B T_{\min}$ at $T_{\min} = 1$ K. In this regime, thermodynamic data cannot resolve the order of the levels. We also find an increased MSE along the lines $x_1 \simeq 0.35$ for positive $x_0 > 0$ and $x_1 = -0.6$ and $x_1 = 0.75$ for negative $x_0 < 0$. This follows from the fact that the ground state energy exhibits level crossings in these parameter regions, as shown in Fig. 1(a). This makes the thermodynamic observables quite sensitive to small errors in the Stevens parameters as the nature of the ground state changes between singlet, doublet and triplet states. As a result, the MSE is enhanced even though the MAE is still small and the Stevens coefficients are predicted with high accuracy.

B. Hexagonal point group symmetry

We also consider the case of hexagonal point group $\mathcal{G} = \bar{6}$m2 and $J = 15/2$, which is applicable to hexagonal Er rare-earth compounds. Being a half-integer value of J, the energy level exhibits Kramers degeneracy in the absence of a magnetic field. The number of independent Stevens coefficients for this point group is $N_{\text{St}} = 4$ (see Eq. (2.15)). We split off the sign of x_0 into a separate
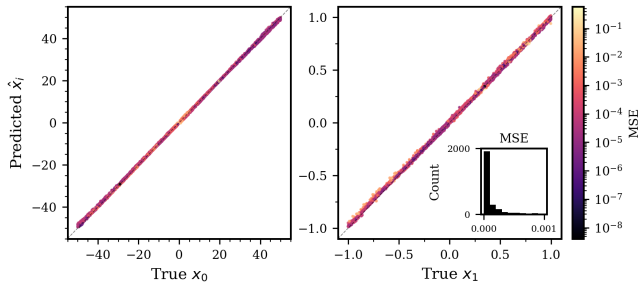
FIG. 4. True Stevens coefficients x_i versus predicted coefficients $\hat{x}_i$ for cubic point group m$\bar{3}$m and $J = 4$. The network was trained with the five thermodynamic data sets $\{\mu_{z,T_\alpha}(b_i), \chi_z(t_i), c_M(t_i)\}$ for $T_\alpha = \{1, 17, 300\}$ K. We choose the high-symmetry direction $z = [001]$. The Stevens coefficients are predicted with MAEs given by MAE(0) = 0.321 K and MAE(1) = 0.012. The inset and color coding shows the MSE of each data point, which lies at $\langle\text{MSE}\rangle = 2.7 \times 10^{-3}$ on average.
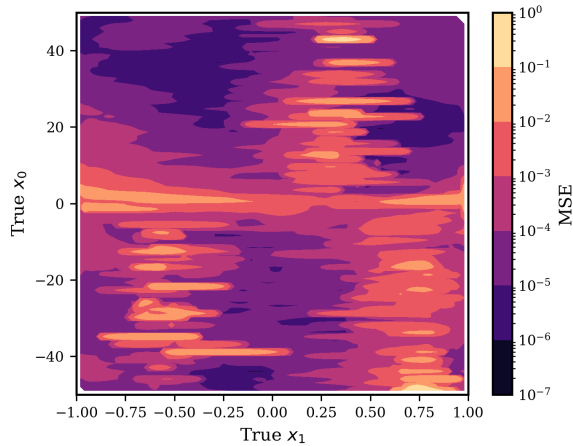
FIG. 5. Heat map showing the MSE as a function of the true Stevens coefficients x_0 and x_1 for $\mathcal{G} = $ m$\bar{3}$m and $J = 4$. The heat map shows a total number of $N_{\text{testing}} = 4 \times 10^3$ data points. Regions with larger MSE occur when x_0 becomes comparable to the temperature probed ($x_0 \lesssim T_{\min} = 1$ K) and when there is an energy level crossing involving the ground state [see Fig. 1(a)]. While it becomes impossible to predict the coefficients if $x_0 \lesssim T_{\min}$ as the spectrum collapses, the increased MSE at the position of level crossings rather indicate an enhanced sensitivity of the observables with respect to small errors in the $\{\hat{x}_i\}$, which are still accurately predicted by the network (see Fig. 4).

parameter x_4 such that $\text{sign}(x_4) = \text{sign}(x_0)$. This allows us to consider the parameter $x_0 \equiv |x_0|$ to be strictly positive. The training data sets thus contains the five coefficients, $\{x_0, \ldots, x_4\}$, with strictly positive $x_0 > 0$ and only the sign of x_4 entering the Hamiltonian. We consider the scale parameter to be in the region 0.5 K $\leq x_0 \leq 50$ K and $-1 \leq x_i \leq 1$ for $i \geq 1$.

As described in Sec. III A, the training data contains in addition to the specific heat c_M, the magnetization

μ_a and susceptibility χ_a along both $a = [100]$ and $a = [001]$ directions. This provides information about the anisotropy between the ab plane and the c axis in the system, and is necessary for the CNN to be able to predict the parameter x_3. A training data set thus consists of the nine observables $\{\mu_{a,T_\alpha}(B_a), \chi_a(T), c_M(T)\}$. Here, we set $T_\alpha = 1, 17, 300$ K and consider the magnetic field range $0 \leq B_a \leq 10$ T and temperature range 1 K $\leq T \leq 300$ K as described in Sec. III A.

As shown in Fig. 6, the CNN can accurately predict the Stevens parameters with $\{\text{MAE}(0), \ldots, \text{MAE}(3)\} = \{1.325 \text{ K}, 0.024, 0.031, 0.073\}$. The sign of the fifth parameter x_4 was correctly found in 96% of the cases (see inset in the left panel in Fig. 6). The color corresponds to the MSE of each testing data set. A histogram of the MSE values is included in the right most panel. The average MSE over all testing data sets is $\langle\text{MSE}\rangle = 0.272$. In general, the MAE increases slightly with larger values of x_0, which can be understood from the fact that the thermal energy is not sufficient to probe higher lying levels. This could likely be improved by enlarging the temperature range by increasing $T_{\max}$. The CNN performs worst for the x_3 coefficient, in particular when this parameter is small. This parameter contains information about the anisotropy between ab and c axis directions as well as between directions within the ab-plane. The CNN predictions of x_3 could thus likely be improved by providing additional magnetization data along a second, inequivalent direction in the ab plane. Finally, we note that quantitatively similar results were obtained for the integer case of $J = 4$, showing that the method works equally well for integer and half-integer values of J.

C. Tetragonal point group symmetry

We study the performance of the CNN for the tetragonal point group $\mathcal{G} = 4$mm and half-integer $J = 15/2$, which corresponds to tetragonal Er rare-earth systems. In this case, the CF allows for $N_{\text{St}} = 5$ independent Stevens parameters. Since we split off the sign of x_0, the CNN actually predicts six parameters $\{x_0, \ldots, x_5\}$, where $0.5 \leq x_0 \leq 50$ (in units of K), $-1 \leq x_i \leq 1$ for $i \geq 1$ and the training data depend only on the sign of x_5. A training data set contains nine scaleograms obtained from specific heat c_M, magnetization μ_a and susceptibility χ_a along $a = [100], [001]$ directions: $\{\mu_{a,T_\alpha}(B_a), \chi_a(T), c_M(T)\}$. The temperature and field ranges are 1 K $\leq T \leq 300$ K and $0 \leq B_a \leq 10$ T. Providing information about the anisotropy between the ab plane and the c axis is necessary for the CNN to be able to learn the dependence on the parameters x_2 and x_4.

As shown in Fig. 7, the CNN can accurately predict the Stevens parameters for the majority of the data points that it was tested on. The MAEs of the Stevens coefficients read $\{\text{MAE}(0), \ldots, \text{MAE}(4)\} = \{1.380 \text{ K}, 0.022, 0.038, 0.031, 0.059\}$. The sign of x_5 was correctly predicted by the network in 93% of the cases.
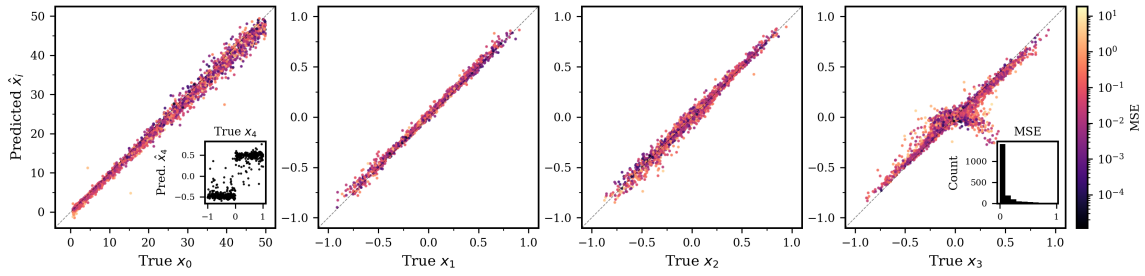
FIG. 6. True Stevens coefficients x_i versus predicted coefficients $\hat{x}_i$ for hexagonal $\bar{6}$m2 point group and $J = 15/2$ for $N_{\text{testing}} = 4000$ data sets. The color denotes the MSE and the inset histogram in the right most panel shows the distribution of the MSE. This network was trained using nine 1D observable data sets, $\{\mu_a(T, b_i), \chi_a(t_i), c_M(t_i)\}$. The MAEs of the Stevens coefficients are $\{\text{MAE}(0), \ldots, \text{MAE}(3)\} = \{1.325 \text{ K}, 0.024, 0.031, 0.073\}$. The sign of x_4 was correctly predicted in 96% of the cases (see inset in left most panel). The average MSE is $\langle \text{MSE} \rangle = 0.272$.
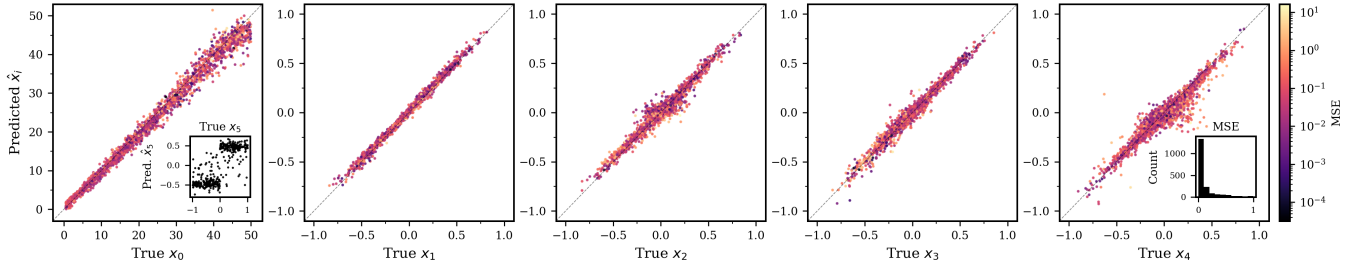

FIG. 7. True Stevens coefficients x_i versus predicted coefficients $\hat{x}_i$ for tetragonal 4mm point group and $J = 15/2$ for $N_{\text{testing}} = 4000$ data sets. The color denotes the MSE and the inset histogram in the right most panel shows the distribution of the MSE. This network was trained using nine 1D observable data sets, $\{\mu_a(T, b_i), \chi_a(t_i), c_M(t_i)\}$. The MAEs of the Stevens coefficients are $\{\text{MAE}(0), \ldots, \text{MAE}(4)\} = \{1.380 \text{ K}, 0.022, 0.038, 0.031, 0.059\}$. The sign of x_5 was correctly predicted correctly in 93% of the cases (see inset in left most panel). The average MSE is $\langle \text{MSE} \rangle = 0.255$.

The average MSE is given by $\langle \text{MSE} \rangle = 0.255$. The overall performance is comparable to the hexagonal case of $\bar{6}$m2, even though the tetragonal case exhibits one more Stevens parameter. Similar to the hexagonal case, the error is larger for larger values of x_0, which likely stems from the fact that the bandwidth of the spectrum becomes larger than the maximal thermal energy $k_B T_{\text{max}}$. This suggests increasing the temperature range in the training data. The MAE of different coefficients is comparable. The largest MAE occurs for the parameter x_4, which measures the anisotropy of the system, both between ab and c directions as well as within the ab plane. It could likely be better predicted by adding magnetization data along another inequivalent direction in the ab plane to the training sets. Finally, we note that we have applied the algorithm to the case of $J = 4$ and obtained quantitatively similar results.

V.   APPLICATION TO EXPERIMENTAL DATA

The ultimate application of the presented CNN algorithm is to extract CF parameters from real experimental data. Since experimental circumstances and parameters are different for each material, it is required to train a custom CNN for each case. We provide the necessary programs as open-source software [47].

In this section, we demonstrate this by applying the algorithm to two published experimental data sets of Praseodymium-based rare-earth intermetallics: (i) PrAgSb_2 [21], where the rare-earth ion Pr^{3+} ($J = 4$) experiences 4mm site symmetry, and (ii) PrMg_2Cu_9 [23], where it exhibits $\bar{6}$m2 symmetry. Importantly, published data of magnetization, magnetic susceptibility and magnetic specific heat on the same single crystal are available for these systems [21, 23].

When applying the CNN algorithm to experimental data, one must first select the set of thermodynamic observables that are given to the network. In general, it is best to include as much data as possible, for example, magnetization along different directions, which informs the network about the anisotropy in the material. In addition, one must choose the temperature and magnetic field ranges. These will in general be different for each observable to ensure that the assumptions of the modeling, in particular the single-ion approximation, are valid. Once the set of experimental input data is determined (together with J and $\mathcal{G}$), one generates a customized training data set using the same observables and parameter ranges.
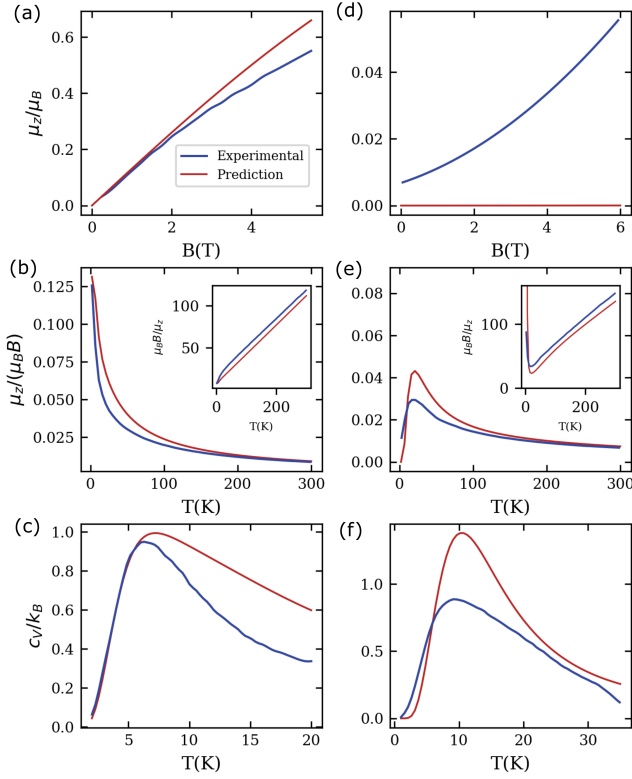
	x_i	(k, q)	$\mathcal{B}_k^q$	$B_{k,\text{Stevens}}^q$
x_0	39 K	–	–	–
x_1	0.19	$(2, 0)$	4.1 K	1.3 K
x_2	0.068	$(4, 4)$	0.059 K	0.052 K
x_3	-0.32	$(4, 0)$	-0.39 K	-0.041 K
x_4	-0.030	$(6, 4)$	-2.1×10^{-3} K	-1.5×10^{-3} K
$\text{sign}(x_5)$	-1	$(6, 0)$	-0.038 K	-2.4×10^{-3} K

TABLE I. Stevens parameters for PrAgSb_2 obtained from CNN using $\{\mu_{z,T=2\,\text{K}}, \chi_z(T), c_M(T)\}$ from Ref. [21] as input data. The coefficients x_i and $\mathcal{B}_k^q$ are defined in Sec. II, and the $B_{k,\text{Stevens}}^q$ are prefactors of Stevens operator equivalents O_k^q [24, 40]. Note that Myers *et al.* state $B_{2,\text{Stevens}}^0 = 1.8 \pm 0.3$ K [21], close to what we find.

FIG. 8. Panels (a-c) show experimental target data (blue) and CNN predictions (red) for PrAgSb_2, where the rare-earth ion Pr^{3+} ($J = 4$) exhibits tetragonal 4mm site symmetry. Experimental data is taken from Ref. 21 and transformed to the units used by the CNN as described in Appendix A. We train a network on the subset of experimentally available input data $\{\mu_{z,\text{T=2\,K}}, \chi_z(T), c_M(T)\}$ in the magnetic field and temperature ranges shown, where $z = [001]$ direction. The input thus consists of a three-channel scaleogram image. The network predicts the Stevens parameters $\{x_0, \ldots, x_4\} = \{38.657, 0.194, 0.068, -0.317, -0.030\}$ and $\text{sign}(x_5) = -1$, which describe the experimental data well with a MSE = 0.016. Panels (d-f) show experimental target data (blue) and CNN predictions (red) for PrMg_2Cu_9, where the rare-earth ion Pr^{3+} ($J = 4$) exhibits hexagonal $\bar{6}$m2 site symmetry. Experimental data is taken from Ref. 23 and transformed to the units used by CNN. The network is trained based on the type of available experimental data, which is identical to panels (a-c). The network predicts the Stevens parameters $\{x_0, \ldots, x_3\} = \{-57.862, -0.575, -0.269, 0.013\}$ and $\text{sign}(x_4) = -1$, which describes the experimental data well with a MSE = 0.019. Note that the magnetization μ_z/μ_B is very small due to a strong ab easy-plane anisotropy. In both cases, the network captures the initial slope of the magnetization [note that small y axis scale in panel (d)] and correctly predicts both the Curie-Weiss slope of the inverse susceptibility at higher temperatures as well as the location of the Schottky peak in the specific heat.

When selecting a suitable temperature window, one must ensure that the system remains paramagnetic at all T to avoid the occurrence of many-body phenomena such as the development of magnetic order, which are ne-

glected in the modeling that generates the training data. When including specific heat data, it is also important to realize that the magnetic part of c_M is typically experimentally approximated by subtracting off the specific heat of a corresponding nonmagnetic compound. A nonmagnetic analogue material can often be obtained by replacing the magnetic rare-earth ion by a nonmagnetic one such as La, Y or Lu. The subtraction procedure is only valid when the phonon and electronic contributions to the specific heat in the two materials are identical, which restricts the temperature regime that can be used for (magnetic) c_M in the algorithm. We emphasize that these caveats are related to the physical modeling of the forward problem of computing observables. Our proposed CNN-based approach of solving the inverse problem, however, is more generally valid and can, in principle, also be used in conjunction with more advanced and realistic physical models (that might, e.g., be able to capture magnetism or phonons).

Both Pr compounds that we investigate remain paramagnetic down to $T = 2$ K and can be well described within the single-ion approximation over the complete temperature range from $T_{\min} = 2$ K to $T_{\max} = 300$ K. Published data exists for the magnetization and susceptibility along the [001] axis, which will be used as input data for the CNN. In Appendix A, we describe in detail how to transform from the experimental units to the units used in the training data. While there also exist data for magnetic fields applied in the ab plane, these will not be included, because the exact in-plane direction was not experimentally determined. This implies that the Stevens coefficients that describe the anisotropy in the ab plane cannot be determined from the data. Finally, the subtraction procedure performed to obtain the magnetic specific heat is approximately valid in a low-temperature regime and will also be included in the input data set.

In Fig. 8, we show the complete experimental data, which is used as input for the CNNs (after performing the CWT of Sec. III A), together with the predictions that rely on the values $\{\hat{x}_i\}$ obtained from the CNN. Panels (a-c) are for PrAgSb_2 and panels (d-f) are for PrMg_2Cu_9.

	x_i	(k,q)	$\mathcal{B}_k^q$	$B_{k,\text{Stevens}}^q$
x_0	57.9 K	–	–	–
x_1	0.575	$(2,0)$	18.0 K	5.7 K
x_2	0.269	$(4,0)$	0.49 K	0.05 K
x_3	-0.013	$(6,6)$	-1.3×10^{-3} K	-1.3×10^{-3} K
$\text{sign}(x_4)$	-1	$(6,0)$	0.021 K	1.3×10^{-3} K

TABLE II. Stevens parameters for $PrMg_2Cu_9$ obtained from CNN using $\{\mu_{z,T=2\,\text{K}}, \chi_z(T), c_M(T)\}$ from Ref. [23] as input data.

The experimental data is taken from Refs. [21, 23], from which the network extracts the values of the Stevens coefficients given in Table I and II. To validate the predictions of the network, we use these coefficients to calculate the thermodynamic observables. As shown in Fig. 8, we observe an overall very good agreement between the experimental data and the theoretical results for both compounds. The MSEs are 0.016 and 0.019 for $PrAgSb_2$ and $PrMg_2Cu_9$, respectively. This justifies our approach to employ the single-ion approximation to describe the material properties. The slope of the inverse susceptibility is described correctly and there are only slight deviations in the susceptibility at intermediate temperatures. Finally, the position of the peak in the specific heat is correctly reproduced, which is a direct indication that the energy level splitting between the ground and excited state has been successfully extracted from the data.

VI. SUMMARY AND OUTLOOK

To summarize, we present a deep ML algorithm for extracting CF Stevens parameters from thermodynamic observables of local-moment materials that can be treated within the single-ion approximation. We focus on rare-earth intermetallics and train a CNN on input data of magnetization, susceptibility, and specific heat. The training data is obtained from straightforward statistical mechanics calculations for different, randomly sampled, values of the Stevens parameters. To exploit the ability of CNNs in image recognition, we process the raw thermodynamic data using a wavelet transform and feed the resulting multi-channel scaleogram image to the network. The presented algorithm provides a convenient and powerful tool for extracting CF parameters from experimental data that avoids a tedious multiple parameter fitting procedure. We provide all programs necessary to run the algorithm and to apply it to experimental situations as open software [47].

The CNN provides an unbiased solution to the inverse problem of finding the Stevens parameters for a given set of thermodynamic observables. Depending on the type and amount of input data, this inverse problem can be ill-defined and allow for multiple solutions. Our study is an explicit test on the performance of CNNs on this well-known inverse physics problem of wide interest. We systematically investigate the performance of the algorithm for different site symmetries in the cubic, hexagonal and tetragonal crystal classes. The point groups we consider are experimentally relevant and allow for 2, 4 and 5 independent Stevens parameters, thus testing the CNN in cases of increasing complexity. We find that the CNN can accurately predict all Stevens coefficients if one provides magnetization data both along the easy-axis as well as within the easy-plane. The network performs equally well for integer and half-integer values of the total angular momentum J. Finally, we demonstrate that the algorithm also works well when applied to real experimental data of two Praseodymium compounds, $PrAgSb_2$ and $PrMg_2Cu_9$, which we obtain from the literature [21, 23].

Interesting future directions are to consider lower symmetry CFs, to systematically investigate the stability of the algorithm with respect to input data noise, and to explore other choices of input observables, including low-symmetry magnetic field directions and direction averaged quantities. The latter would extend the applicability of the algorithm to polycrystalline materials.

ACKNOWLEDGMENTS

[1] R. Elliott, ed., *Magnetic Properties of Rare Earth Metals* (Springer, 1972).

[2] P. Fulde and M. Loewenhaupt, Magnetic excitations in crystal-field split 4f systems, Adv. Phys. **34**, 589 (1985).

[3] A. Szytuła, Chapter 2 Magnetic properties of ternary intermetallic rare-earth compounds, in *Handbook of Magnetic Materials*, Vol. 6 (Elsevier, 1991) pp. 85–180.

[4] P. C. Canfield and S. L. Bud'ko, Preserved entropy and fragile magnetism, Rep. Prog. Phys. **79**, 084506 (2016).

[5] K. W. H. Stevens, Matrix Elements and Operator Equivalents Connected with the Magnetic Properties of Rare Earth Ions, Proc. Phys. Soc. A **65**, 209 (1952).

[6] B. Bleaney and K. W. H. Stevens, Paramagnetic resonance, Rep. Prog. Phys. **16**, 108 (1953).

[7] S. A. Altshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance* (Academic Press, New York, 1964).

[8] B. G. Wybourne, *Spectroscopic Properties of Rare Earths* (John Wiley & Sons, New York, N.Y., USA, 1965).

[9] M. D. Kuz'min and A. M. Tishin, Chapter Three Theory of Crystal-Field Effects in 3d-4f Intermetallic Compounds, in *Handbook of Magnetic Materials*, Vol. 17, edited by K. H. J. Buschow (Elsevier, 2007) pp. 149–233.

[10] P. Fazekas, *Lecture Notes on Electron Correlation and Magnetism*, Series in Modern Condensed Matter Physics No. v. 5 (World Scientific, Singapore ; River Edge, N.J, 1999).

[11] D. L. Cox, Quadrupolar Kondo effect in uranium heavy-electron materials?, Phys. Rev. Lett. **59**, 1240 (1987).

[12] D. L. Cox and A. Zawadowski, Exotic Kondo effects in metals: Magnetic ions in a crystalline electric field and tunnelling centres, Advances in Physics **47**, 599 (1998).

[13] P. Chandra, P. Coleman, and R. Flint, Hastatic order in the heavy-fermion compound URu2Si2, Nature **493**, 621 (2013).

[14] J. S. Van Dyke, G. Zhang, and R. Flint, Field-induced ferrohastatic phase in cubic non-Kramers doublet systems, Phys. Rev. B **100**, 205122 (2019).

[15] H. Bethe, Termaufspaltung in Kristallen, Annalen der Physik **395**, 133 (1929).

[16] K. Lea, M. Leask, and W. Wolf, The raising of angular momentum degeneracy of f-Electron terms by cubic crystal fields, J. Phys. Chem. Solids **23**, 1381 (1962).

[17] U. Walter, Treating crystal field parameters in lower than cubic symmetries, J. Phys. Chem. Solids **45**, 401 (1984).

[18] M. Richter, Chapter 2 Density functional theory applied to 4f and 5f elements and metallic compounds, in *Handbook of Magnetic Materials*, Vol. 13 (Elsevier, 2001) pp. 87–228.

[19] M. Loewenhaupt and K. H. Fischer, Chapter 6 Neutron scattering on heavy fermion and valence fluctuation 4f-systems, in *Handbook of Magnetic Materials*, Vol. 7 (Elsevier, 1993) pp. 503–608.

[20] O. Moze, Chapter 4 Crystal field effects in intermetallic compounds studied by inelastic neutron scattering, in *Handbook of Magnetic Materials*, Vol. 11 (Elsevier, 1998) pp. 493–624.

[21] K. Myers, S. Bud'ko, I. Fisher, Z. Islam, H. Kleinke, A. Lacerda, and P. Canfield, Systematic study of anisotropic transport and magnetic properties of RAgSb2 (R=Y, La–Nd, sm, Gd–Tm), J. Magn. Magn. Mater **205**, 27 (1999).

[22] S. L. Bud'ko, Z. Islam, T. A. Wiener, I. R. Fisher, A. H. Lacerda, and P. C. Canfield, Anisotropy and metamagnetism in the RNi2Ge2 (R=Y, La–Nd, Sm–Lu) series, J. Magn. Magn. Mater **205**, 53 (1999).

[23] T. Kong, W. R. Meier, Q. Lin, S. M. Saunders, S. L. Bud'ko, R. Flint, and P. C. Canfield, Physical properties of single crystalline RMg_2Cu_9 (R=Y,Ce-Nd,Gd-Dy,Yb) and the search for in-plane magnetic anisotropy in hexagonal systems, Phys. Rev. B **94**, 144434 (2016).

[24] R. M. Geilhufe and W. Hergert, GTPack: A mathematica group theory package for application in solid-state physics and photonics, Frontiers in Physics **6**, 86 (2018).

[25] W. Hergert and R. M. Geilhufe, *Group Theory in Solid State Physics and Photonics: Problem Solving with Mathematica* (Wiley-VCH, Weinheim, Germany, 2017).

[26] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Vol. 375 (Springer Science & Business Media, 1996).

[27] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[28] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to Machine Learning for physicists, Phys. Rep. **810**, 1 (2019).

[29] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[30] L.-F. Arsenault, R. Neuberg, L. A. Hannah, and A. J. Millis, Projected regression method for solving Fredholm integral equations arising in the analytic continuation problem of quantum physics, Inverse Problems **33**, 115007 (2017).

[31] R. Fournier, L. Wang, O. V. Yazyev, and Q. Wu, Artificial neural network approach to the analytic continuation problem, Phys. Rev. Lett. **124**, 056401 (2020).

[32] P. Z. Hanakata, E. D. Cubuk, D. K. Campbell, and H. S. Park, Forward and inverse design of kirigami via supervised autoencoder, , arXiv:2008.05298 (2020), arXiv:2008.05298 [cond-mat.mtrl-sci].

[33] Y. H. Teoh, M. Drygala, R. G. Melko, and R. Islam, Machine learning design of a trapped-ion quantum spin simulator, Quantum Sci. Technol. **5**, 024001 (2020).

[34] N. Laanait, J. Romero, J. Yin, M. T. Young, S. Treichler, V. Starchenko, A. Borisevich, A. Sergeev, and M. Matheson, Exascale deep learning for scientific inverse problems, arXiv e-prints (2019), arXiv:1909.11150 [cs.LG].

[35] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton Univ. Press, Princeton, N.J., 1957).

[36] B. R. Judd, *Operator Techniques in Atomic Spectroscopy*, 1st ed. (McGraw-Hill, New York, 1963).

[37] H. A. Buckmaster, Tables of matrix elements for the operators, Can. J. Phys. **40**, 1670 (1962).

[38] D. Smith and J. H. M. Thornley, The use of 'operator equivalents', Proc. Phys. Soc. **89**, 779 (1966).

[39] O. Danielsen and P.-A. Lindgård, *Quantum Mechanical Operator Equivalents Used in the Theory of Magnetism*, Denmark. Forskningscenter Risoe. Risoe-r No. 259 (Risø National Laboratory, 1972).

[40] A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon Press ; Oxford University Press, Oxford, 1970).

[41] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. (Academic Press, Amsterdam ; Boston, 2008).

[42] G. R. Lee, R. Gommers, K. Wohlfahrt, F. Wasilewski, A. O'Leary, H. Nahrstaedt, D. M. Hurtado, A. Sauvé, T. Arildsen, H. Oliveira, D. M. Pelt, A. Agrawal, Sylvain-Lan, M. Pelletier, M. Brett, F. Yu, S. Choudhary, D. Tricoli, L. M. Craig, L. Ravindranathan, J. Dan, jakirkham, J. Antonello, D. Laszuk, D. Goertzen, C. Goldberg, B. Reczey, 0-tree, A. Smith, and asnt, PyWavelets/pywt: PyWavelets 1.1.1, Zenodo (2019).

[43] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, in *Proceedings of the IEEE* (1998) pp. 2278–2324.

[44] A. F. Agarap, Deep learning using rectified linear units (ReLU), CoRR (2018), arXiv:1803.08375.

[45] F. Chollet *et al.*, Keras, https://keras.io (2015).

[46] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (2015).

[47] P. P. Orth and N. Berthusen, Using convolutional neural networks to extract crystal field parameters from thermodynamic data, figshare (2020).

Appendix A: Transformation between units used in experiment and training data

In this section, we describe how to transform from the commonly used experimental units in Refs. [21, 23] to the units used in the training data generation. It is important to first perform a transformation of units in the experimental data set before giving it to the CNN.

1. Specific heat

The SI units for the (magnetic) specific heat are $[c_M] = $ J/(kg K). One often also uses units of $[c_M] = $ J/(mol K). The training data contains c_M/k_B per rare-earth ion, which is a dimensionless quantity that we denote by $c_{\text{train}} \equiv c/k_B$ per rare-earth ion. To transform experimental data c_{exp} given in units of J/(mol K) into the training data units, we need to perform

$$\frac{c_{\text{exp}}}{N_A k_B} = \frac{c_{\text{exp}}}{8.31445973} \mathrel{\widehat{=}} c_{\text{train}} \tag{A1}$$

where N_A is Avogadro's constant and k_B is the Boltzmann constant. In other words, we divide the numerical values obtained from experimental plots by a factor of 8.31445973 before feeding them into the CNN.

2. Magnetization

The training data uses the dimensionless quantity μ_α/μ_B, which is the magnetic moment per rare-earth ion. This is identical to the units used in the experimental plots, which we can thus directly input into the CNN. We note that we transform from Oersted (auxiliary field units) to Tesla (magnetic field units) via $\mu_0 \times 1 \text{ Oe} = 10^{-4}$ T.

3. Susceptibility

The magnetic susceptibility (or volume susceptibility) is experimentally obtained as $\chi = M/H$, where M is the magnetization and H is the auxiliary field. Its SI and cgs units are $[\chi] = 4\pi(1) = 1\text{emu/cm}^3$. Note that it is dimensionless in SI units. It is distinguished from the molar susceptibility χ_{mol} with units $[\chi_{\text{mol}}] = 4\pi 10^{-6} \text{m}^3/\text{mol} = 1\text{emu/mol}$. The two are related by $\chi = 4\pi \times 10^{-6} \frac{\rho}{\mathcal{M}} \chi_{\text{mol}}$, where ρ is the mass density, $\mathcal{M}$ is the molar mass.

The training data reads $\chi_{\text{train}} = \mu_\alpha/(\mu_B B)$, which is the ratio of the induced magnetic moment per rare-earth ion μ_a divided by the Bohr magneton μ_B and a small magnetic field $B = 10^{-4}$ T. Note that the result is independent of the value of B as we ensure that we are in the linear regime of $\mu_a(B)$. This is related to the experimentally measured (volume) susceptibility via $\chi_{\text{train}} = \frac{V_{\text{uc}}}{N_R \mu_B \mu_0} \chi$, where χ is the magnetic susceptibility in SI units and V_{uc}/N_R is an effective volume per rare-earth ion such that the magnetization $M_a = \mu_a/(V_{\text{uc}}/N_R)$. Here, V_{uc} is the unit cell volume and N_R is the number of rare-earth ions per unit cell.

Combining the two transformations discussed above leads to

$$\chi_{\text{train}} \mathrel{\widehat{=}} 4\pi \times 10^{-6} \frac{\rho}{\mathcal{M}} \frac{V_{\text{uc}}}{N_R} \frac{\chi_{\text{mol}}}{\mu_0 \mu_B} = 1.79053 \frac{N_{\text{f.u}}}{N_R} \chi_{\text{mol}}, \tag{A2}$$

where $N_{\text{f.u.}}$ is the number of formula units per unit cell. One thus needs to multiply the experimental data for M/H (in emu/mol) by a factor of $1.79053 \frac{N_{\text{f.u}}}{N_R}$, before feeding it into the CNN.