

Cite this: *J. Mater. Chem. A*, 2025, 13, 35389

# Performance metrics for tensorial learning: prediction of $\text{Li}_4\text{Ti}_5\text{O}_{12}$ nuclear magnetic resonance observables at experimental accuracy

Angela F. Harper, <sup>a</sup> Simone S. Köcher, <sup>\*ba</sup> Karsten Reuter <sup>a</sup>  
and Christoph Scheurer <sup>ab</sup>

Predicting observable quantities from first principles calculations is the next frontier within the field of machine learning (ML) for materials modelling. While ML models have shown success for the prediction of scalar properties such as energetics or band gaps, models and performance metrics for the learning of higher order tensor-based observables have not yet been formalized. ML models for experimental observables, including tensorial quantities, are essential for exploiting the full potential of the paradigm shift enabled by machine learned interatomic potentials by mapping the structure–property relationship in an equally efficient way. In this work, we establish performance metrics for accurately predicting the electric field gradient tensor (EFG) underlying nuclear magnetic resonance (NMR) spectroscopy. We further demonstrate the superiority of a tensorial learning approach that fully encodes the corresponding symmetries over a separate scalar learning of individual tensor-derived observables. To this end we establish an extensive EFG dataset representative of real experimental applications and develop performance metrics for model evaluation which directly focus on the targeted NMR observables. Finally, by leveraging the computational efficiency of the ML method employed, we predict quadrupolar observables for 1512 atom models of  $\text{Li}_4\text{Ti}_5\text{O}_{12}$ , a high performance Li-ion battery anode material, which is capable of accurately distinguishing local atomic environments *via* their NMR observables. This workflow and dataset sets the standard for the next generation of tensorial based learning for spectroscopic observables.

Received 23rd June 2025  
Accepted 12th September 2025

DOI: 10.1039/d5ta05090a

rsc.li/materials-a

## 1 Introduction

Experimental solid-state NMR provides powerful, yet non-destructive methods to characterize atomic structures and dynamics over several length and time scales.<sup>1–3</sup> In quadrupolar nuclei, such as <sup>7</sup>Li, <sup>27</sup>Al, or <sup>17</sup>O, additional information is gained by directly probing the electric field gradient (EFG) tensor at the nucleus. However, in most state-of-the-art materials of interest, such as high-performance battery materials, local defects, disorder or amorphous regions are not only key to their function, but result at the same time in complex spectra, which are impossible to interpret unambiguously from experiment alone. This often renders a complementary predictive-quality modeling of EFG tensors indispensable for a comprehensible NMR crystallography approach.<sup>4–7</sup> Corresponding first-principles calculations, typically based on density-functional theory (DFT), are well established.<sup>7–12</sup> Despite their intrinsic high computational cost and unfavorable scaling with system size, they are routinely combined with random structure searches,<sup>7,13</sup> systematic sampling of structural disorder,<sup>13</sup> and

molecular dynamics simulations<sup>14,15</sup> to study the NMR observables of structural or temporal ensembles. However, while the structural models become more complex and larger and the dynamic trajectories longer in order to approach the complexity of real samples and experiments, the limits of DFT become more apparent in particular when studying the interplay of high structural complexity and dynamics over various timescales.<sup>16,17</sup>

A now common approach to reduce the computational burden of quantum mechanical simulations while retaining their first-principles accuracy is to train a machine learning (ML) surrogate model from a suitably composed database of calculated data. This is an established field for scalar properties, primarily focused on predicting the potential energy surface in order to create an ML inter-atomic potential for a given system.<sup>18–23</sup> A limited set of studies also learn additional scalar properties such as isotropic chemical shifts, dipole moments, or band gaps.<sup>24–26</sup> However, many physical properties are tensorial in nature or are derived from tensorial quantities, just as all NMR quadrupolar observables derive from the EFG tensor. In principle, either the relevant experimental scalar

<sup>a</sup>Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, DE, Germany<sup>b</sup>Institute of Energy Technologies (IET-1), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, DE, Germany. E-mail: s.koecher@fz-juelich.de

observables<sup>27–29</sup> or the individual scalar tensor elements can also be learned with well-established, symmetry-invariant ML approaches.<sup>26</sup> Yet, as this neglects the inherent tensor symmetries one would intuitively expect this approach to be inferior to a full tensorial learning that uses appropriate symmetry-equivariant descriptors.<sup>16</sup>

Validating and quantifying this notion requires both a diverse and challenging database that is representative for the complexity of a real-life application and a standardized method for evaluating and optimizing tensorial ML approaches to the respective tensor-derived observables. To this end, we use NMR as a prominent showcase and introduce an extensive EFG database for the commercial, high-performance Li-ion battery material  $\text{Li}_4\text{Ti}_5\text{O}_{12}$  (LTO). LTO is particularly suitable for such benchmark purposes as all three of its constituent species are quadrupolar and because its ionic mobility has been repeatedly studied by advanced  $^7\text{Li}$  NMR experiments.<sup>30–32</sup> We correspondingly assemble a database of 68 880 EFG tensors calculated for all three atomic species and for a wide range of interatomic distances that reflect those probed in the measurements. We then derive a performance metric for key NMR observables that is used in the optimization and assessment of scalar and tensor EFG ML models. This indeed reveals an order of magnitude superiority in predictive performance for an explicit learning of the full EFG tensor. The failure of the symmetry-agnostic scalar approach is instead traced back to its inability of capturing especially the orientation of the EFG tensor, which the experimental observables sensitively depend on. We finally demonstrate the effectiveness of our approach by predicting the EFG tensors for over 11 000 Li sites within 1512 atom models of LTO, and successfully distinguishing different local Li environments in the material.

Our guiding principles for developing performance metrics are directly geared to measurable observables and using them in the ML optimization are readily transferable to multiple other applications where tensor interactions matter, *e.g.* dielectric interactions, atomic forces, stress and strain, or chemical shielding. By explicitly demonstrating its effectiveness over scalar learning, and providing a database suitable for benchmarking, we thus aim to present this work as a general guideline for further tensorial learning methods.

## 2 Methods

### 2.1 Computational details

All DFT calculations of the EFG tensors carried out to generate the LTO-EFG database were performed with the plane-wave pseudopotential code CASTEP v22.1,<sup>33</sup> using the Gauge-Including Projector Augmented Waves (GIPAW) implementation<sup>34,35</sup> for calculating NMR properties. The PBE functional<sup>36</sup> was used to describe electronic exchange and correlation, with test calculations indicating essentially unaltered EFG tensors when instead using the PBEsol<sup>37</sup> or RSCAN<sup>38</sup> functionals. At a plane wave cut-off of 1000 eV and a  $k$ -point spacing of  $0.03 \times 2\pi \text{ \AA}^{-1}$ , the change in individual EFG tensor components  $\Delta V_{ij}$  between energy cutoffs are converged to within  $4 \times 10^{-3} \text{ V \AA}^{-2}$ . We fix  $V_{zz}$  to be positive, to ensure that the sign of the

eigenvalues is consistent between the DFT and ML tensors. The accuracy of DFT-derived EFG tensors was benchmarked before.<sup>39</sup>

All ML tasks were carried out using the SA-GPR (Symmetry-Adapted Gaussian Process Regression) implementation of  $\lambda$ -SOAP openly available at ref. 40. In particular, the scalar learning was performed using this code, with a scalar  $\lambda = 0$  value, and the rank-2 tensorial learning was performed using the tensor  $\lambda = 2$  value. Full details of the computational and theoretical background for TENSOP are given in Grisafi *et al.*<sup>41</sup> We provide further details of the specific formulations of the SOAP kernels used for the scalar GPR, and  $\lambda$ -SOAP kernels for the tensorial GPR in Sections 2.4 and 2.5.

## 3 Results

### 3.1 NMR as a showcase for tensorial learning

NMR quadrupolar observables all derive from the EFG tensor  $\mathbf{V}$  that characterizes the gradient of the electric field  $V$  experienced by a nucleus due to the nearby charge distribution. The tensor is calculated as a spatial second derivative

$$V_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j}, \quad (1)$$

which at a fixed position can be written as a  $(3 \times 3)$  matrix,

$$\mathbf{V} = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{13} & V_{23} & V_{33} \end{pmatrix}. \quad (2)$$

This matrix is not simply a combination of nine unrelated values, but rather a set of components which must satisfy both the properties that the matrix is symmetric ( $V_{ij} = V_{ji}$ ) and that the trace is 0 ( $V_{11} + V_{22} + V_{33} = 0$ ). Therefore, while every EFG tensor can be written as a  $(3 \times 3)$  matrix, not every  $(3 \times 3)$  matrix is a valid tensor. In other words, any  $(3 \times 3)$  matrix that we might predict from ML methods, which does not have these properties, violates the basic underlying physical properties of the system.

An EFG tensor can be transformed into its principal axis system (PAS)  $\mathbf{V}_{\text{PAS}}$ , shown in Fig. 1a, by diagonalization at a given nuclear position. This yields the eigenvalues  $V_{xx}$ ,  $V_{yy}$ , and  $V_{zz}$ , where by convention  $|V_{zz}| \geq |V_{yy}| \geq |V_{xx}|$ .<sup>42</sup> In the PAS,  $V_{zz}$  describes the magnitude of the tensor, while  $V_{yy}$  and  $V_{xx}$  describe its width. Major standard NMR observables like the quadrupolar coupling constant  $C_Q$  and the quadrupolar asymmetry parameter  $\eta$  directly derive from these eigenvalues.  $C_Q$  defines the coupling strength between the EFG and the applied magnetic field  $\mathbf{B}_0$

$$C_Q = \frac{e}{h} Q V_{zz}, \quad (3)$$

where  $e$  is the charge of the electron,  $Q$  is the nuclear quadrupole moment of the specified nucleus, and  $h$  is Planck's constant.



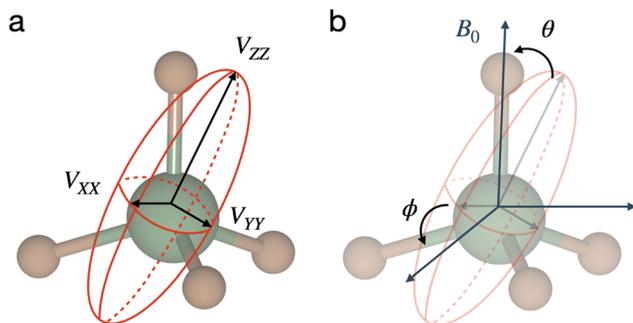


Fig. 1 Electric field gradient tensor visualized on an atomic site. (a) EFG tensor  $V_{\text{PAS}}$  in the principal axis system (PAS), where the eigenvalues  $V_{XX}$ ,  $V_{YY}$ , and  $V_{ZZ}$  describe its magnitude and width. (b) Relationship between the reference orientation of the magnetic field  $B_0$  and  $V_{\text{PAS}}$  using the polar and azimuthal angles  $\theta$  and  $\phi$ .

$$\eta = \frac{V_{XX} - V_{YY}}{V_{ZZ}} \quad (4)$$

describes the shape of the tensor, where  $\eta \in [0, 1]$  ranges from axial symmetry ( $\eta = 0$ ,  $V_{XX} = V_{YY}$ ) to a flat disk ( $\eta = 1$ ,  $V_{XX} = 0$ ).

Advanced NMR methods like Spin Alignment Echo (SAE) are also sensitive to the explicit orientation of the EFG tensor. SAE particularly tracks changes in the quadrupolar frequency

$$\omega_Q = \frac{3}{2I(2I-1)} C_Q \frac{1}{2} (3 \cos^2(\theta) - 1 - \eta \sin^2(\theta) \cos(2\phi)), \quad (5)$$

where  $I$  is the nuclear spin, and  $\theta \in [0, \pi]$  and  $\phi \in [-\pi/2, \pi/2]$  are the polar and azimuthal angles, respectively, between  $V_{\text{PAS}}$  and the lab frame with  $z$ -axis parallel to the magnetic field  $B_0$  as shown in Fig. 1b.<sup>43–45</sup> As ionic mobility can induce changes in  $\omega_Q$ , SAE has *e.g.* been repeatedly applied to battery materials like LTO.<sup>30–32</sup>

### 3.2 Introducing suitable performance metrics

The previous relations motivate to directly learn the EFG tensor as a whole, instead of separately or simultaneously learning a multitude of derived scalar observables or the individual EFG matrix elements. Consider, that even if each of the  $V_{ij}$  components are learned individually, for example, there is no guarantee that the final nine components when recombined to form a  $(3 \times 3)$  matrix will satisfy any of the properties of a rank-2 tensor. Furthermore, each of such trained models will have its own separate uncertainties with unspecified error propagation to the set of finally targeted NMR observables. Precisely, the latter limitation applies equally to the way tensorial learning models have thus far been evaluated and optimized. While in recent years, several tensorial learning approaches have been introduced, with neural networks,<sup>26</sup> structure–property maps<sup>46</sup> and with regression-based approaches<sup>25,41</sup> using a variety of descriptors,<sup>47,48</sup> they are typically only evaluated on the mean absolute error (MAE) of predicting the individual elements,  $V_{ij}$ , of the non-diagonalized tensor. For a total number of  $N$  data points for which reference values are available,  $\text{MAE}(V_{ij})$  is given by

$$\text{MAE}(V_{ij}) = \frac{\sum_{l=1}^N \sum_{i=1}^3 \sum_{j \geq i}^3 |V_{ij}^{\text{data},l} - V_{ij}^{\text{ML},l}|}{N}. \quad (6)$$

While this is indeed the direct error, *i.e.* loss function, of the tensorial learning model, it again provides no information on how well the actual observables of interest are predicted, and neither has the latter objective ever entered the very optimization process of the model.

To this end, we introduce suitable performance metrics on which we then base the model evaluation and hyperparameter optimization. These metrics are directly related to the targeted observables, but are universally applicable across nuclear species, less prone to noise, and rescaled to all lie within a comparable range. Use of such metrics instead of the observables themselves is particularly relevant for an efficient learning of diverse databases that comprise different quantities of different chemical species, like in the present showcase where we target the simultaneous learning of several key NMR observables for the three constituent LTO species  ${}^7\text{Li}$ ,  ${}^{47}\text{Ti}$ , and  ${}^{17}\text{O}$ . With the above introduced  $C_Q$  and  $\eta$ , we thereby specifically define a metric related to these two standard NMR observables sensitive to the magnitude and shape of the EFG tensor, while with the quadrupolar frequency  $\omega_Q$  probed in advanced SAE-NMR we choose an observable that is sensitive to the tensor orientation.

For  $C_Q$  and  $\eta$ , we define such rescaled and normalized metrics as

$$\tilde{C}_Q = \frac{V_{ZZ} - \mu(V_{ZZ})}{\sigma(V_{ZZ})} \quad (7)$$

and

$$\tilde{\eta} = \frac{V_{XX} - V_{YY}}{\mu(V_{ZZ})}, \quad (8)$$

using the mean  $\mu(V_{ZZ})$  and standard deviation  $\sigma(V_{ZZ})$  for each nucleus.  $\tilde{C}_Q$  thus preserves the magnitude of the original tensor while centering the values around zero. In turn,  $\tilde{\eta}$  reduces the noise introduced in  $\eta$  by the division by  $V_{ZZ}$ , *cf.* eqn (4), and reflects the fact that NMR measurements will only probe an average property over the system, such as  $\mu(V_{ZZ})$ , rather than an individual  $V_{ZZ}$ . At first glance, the polar and azimuthal angles  $\theta$  and  $\phi$  appear as useful additional metrics for  $\omega_Q$ , *cf.* eqn (5), as they have a fixed range, define the orientation of the EFG tensor in space, and hence are independent of the studied nuclear species. However, their discontinuity renders them a non-ideal representation. In tensorial learning, where the tensor orientation is explicitly available, this limitation can be overcome by the unit quaternion (*cf.* SI). This defines the orientation of any rank-2 tensor in terms of a normalized vector ( $\mathbf{q} = (q_0, q_1, q_2, q_3)$ ) and the dot product between two quaternions defines how closely oriented the two respective tensors are.<sup>49</sup> When using quaternions as metric, the performance would thus be evaluated by the calculation of

$$\tilde{q} = \mathbf{q}^{\text{data}} \cdot \mathbf{q}^{\text{ML}}, \quad (9)$$



which tells how well the learned ML tensors are aligned with the reference tensors. In short, inspired by the physical observables ( $C_Q$ ,  $\eta$ ,  $\theta$ ,  $\phi$ ), we introduce a new set of adapted parameters ( $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\mathbf{q}$ ) as ensemble adapted performance metrics, which still represents the magnitude, shape, and orientation of the EFG tensor albeit in a way that allows for a continuous, homogeneously weighted, less noisy, more stable description at the negligible cost of slightly softening some exact bound on the individual tensors (e.g.  $\eta \in [0, 1]$ ).

### 3.3 The LTO-EFG database

Previous databases of tensorial NMR properties including both chemical shielding and EFG tensors focused on providing a breadth of structures across varying chemical compositions.<sup>2,26,50,51</sup> Yet, they used locally geometry-optimized structures throughout, such that only NMR data of stable and metastable states (global and local minima) are included. By construction, learning on such datasets will have only a limited transferability to real (aka defected) or dynamic (aka finite temperature) systems for which information away from the minima would be required. Our aim in creating the LTO-EFG tensor database is therefore to provide a large (68 880 tensors in total) complementary database of DFT-calculated EFG tensors representative of real experimental applications.

In order to provide EFG tensors for the three different nuclei,  ${}^7\text{Li}$ ,  ${}^{47}\text{Ti}$ , and  ${}^{17}\text{O}$ , in a standardized way, all DFT calculations are based on the 42 atom  $R\bar{3}m$  supercell of stoichiometric  $\text{Li}_4\text{Ti}_5\text{O}_{12}$ , shown in Fig. 2a. The tetrahedral  $8a$  sites are occupied by Li and the octahedral  $16d$  sites by either Li or Ti in a ratio of Li : Ti, 1 : 5. The occupational disorder on the  $16d$  Wyckoff site allows for 6 symmetry-inequivalent crystal structures to be enumerated. By ‘rattling’ the atomic positions using a random displacement procedure<sup>52</sup> and by rescaling the crystalline lattice vectors, a total of 1640 structures are generated that we describe in further detail in the SI. The configurational sampling, rattling, and scaling procedure effectively distributes the interatomic distances across a wide range of 1–2 Å around the mean neighbor distance for every atom combination, as shown in Fig. 2b. The resulting diverse quadrupolar coupling constants  $C_Q$  and quadrupolar asymmetries  $\eta$ , as calculated from the eigenvalues of the EFG tensors, are shown across the entire database for each nucleus in Fig. 2c. Being only the second ever EFG tensor database available in literature,<sup>53</sup> this heterogeneity of the LTO-EFG database renders it a truly challenging benchmark for ML tensorial learning.

### 3.4 Scalar learning

Scalar learning focuses on the direct and separate learning of individual scalar quantities. Since the three target NMR observables  $C_Q$ ,  $\eta$  and  $\omega_Q$  as well as the quaternion vector  $\mathbf{q}$  are unsuitable for scalar machine learning purposes as described above, we therefore employ the scalar quantities  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$  and  $\phi$  as training objectives as well as as performance metrics for hyperparameter optimization. Technically and without loss of generality, we choose the established smooth overlap of atomic densities (SOAP) descriptor to encode the local atomic

environments in a symmetry-invariant way.<sup>54</sup> SOAP is an appropriate special case of the atomic cluster expansion<sup>55,56</sup> which combines both radial and spherical harmonic basis functions using a Gaussian smearing to approximate the local atomic density  $\rho(\mathbf{r})$  in a sphere of radius  $r_c$  around any atom of a given structure as depicted in Fig. 3a. The corresponding SOAP kernel ( $K_{\text{SOAP}}$ )<sup>54</sup> between two atomic configurations  $\mathcal{X}$  and  $\mathcal{X}'$  is then designed to enforce that the predicted scalar quantities are invariant under symmetry operations such as rotation, translation, and atomic permutation<sup>41,54</sup>

$$K_{\text{SOAP}}(\mathcal{X}, \mathcal{X}') = \left| \int \rho(\mathbf{r}) \rho'(\mathbf{r}) d\mathbf{r} \right|^2 \quad (10)$$

For learning and testing the scalar models, the entire LTO-EFG database for each nucleus is randomly divided into a training and test set with a ratio of 4 : 1. The learning of each scalar quantity,  $y$  ( $=\tilde{C}_Q, \tilde{\eta}, \theta, \phi$ ), over all configurations  $\mathcal{X}$  across the entire training set of the LTO-EFG database is performed using Gaussian Process Regression (GPR)<sup>20</sup>

$$y(\mathcal{X}) = \sum_{l=1}^N w_l K_{\text{SOAP}}(\mathcal{X}, \mathcal{X}_l), \quad (11)$$

solving for weights  $w_l$ . In practice, the SOAP representation is used here in its kernel form rather than exclusively as a descriptor vector. This choice is critical since the SOAP power spectrum can easily reach thousands of dimensions, which would make direct regression inefficient and prone to overfitting. For optimization of the SOAP hyperparameters, a global MAE minimization is efficiently achieved for each separate scalar quantity using a Box-Behnken design-of-experiment approach<sup>57</sup> with five-fold cross-validation for each nucleus across the training set. Full details of the specific hyperparameters used for each nucleus are provided in the SI. The resulting kernel matrices were normalized and regularized during GPR training to ensure numerical stability. All scalar kernels were generated using the  $\lambda = 0$  implementation of SA-GPR available at ref. 40, which is the generalized version of the original SOAP kernel introduced in.<sup>54</sup>

The achieved predictions of  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$ , and  $\phi$  for all  ${}^7\text{Li}$  in the test set are compiled in Fig. 3b, with comparable findings for the other two nuclei, as well as results for the MAE ( $V_{ij}$ ) provided in the SI.  $\tilde{C}_Q$  and  $\tilde{\eta}$  show an acceptable correlation between DFT and ML data. The Pearson correlation for  $\tilde{C}_Q$  and  $\tilde{\eta}$  is  $r = 0.96$  and  $r = 0.84$ , and the MAE is 0.17 and 0.14, respectively. In contrast, the predictions of  $\theta$  and  $\phi$  are of a substantially lower quality, with no visible linear trend in the correlation plots shown in Fig. 3b. This is reflected in the corresponding poor Pearson correlations of  $r = 0.46$  and  $r = 0.28$ , and MAEs of 0.47 and 0.75 radians, for  $\theta$  and  $\phi$ , respectively. This corresponds to a staggering MAE of  $27^\circ$  for  $\theta$  and  $43^\circ$  for  $\phi$ .

### 3.5 Tensorial learning

The tensorial learning targets the full EFG tensor  $\mathbf{V}$  and trains on the loss function defined in eqn (6). Correspondingly, we can now employ  $\tilde{C}_Q$ ,  $\tilde{\eta}$  and the quaternion  $\mathbf{q}$  as performance metrics



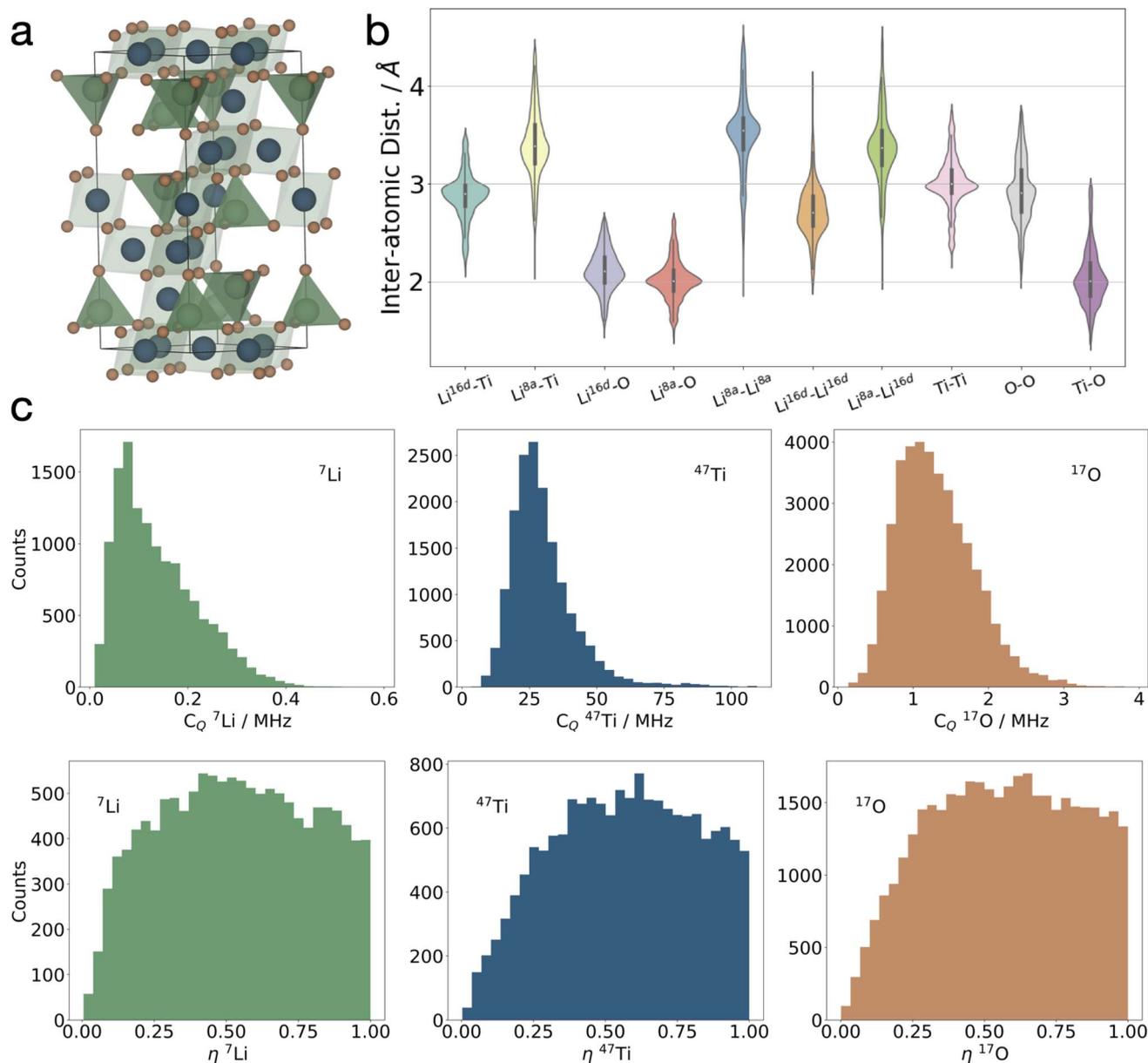


Fig. 2 Distribution of bond lengths,  $C_Q$ , and  $\eta$  for all nuclei in the LTO-EFG database. (a) LTO unit cell with labelled Wyckoff sites, Li 8a (green), Li/Ti 16d (blue spheres and green polygons), and O 32e sublattice (orange). (b) Violin plot showing the distribution of inter-atomic distances in the database, where Li is separated into the 16d and 8a sites. (c) Histograms showing the distribution of the DFT-calculated reference  $C_Q$  (top) and  $\eta$  (bottom) for each nucleus.

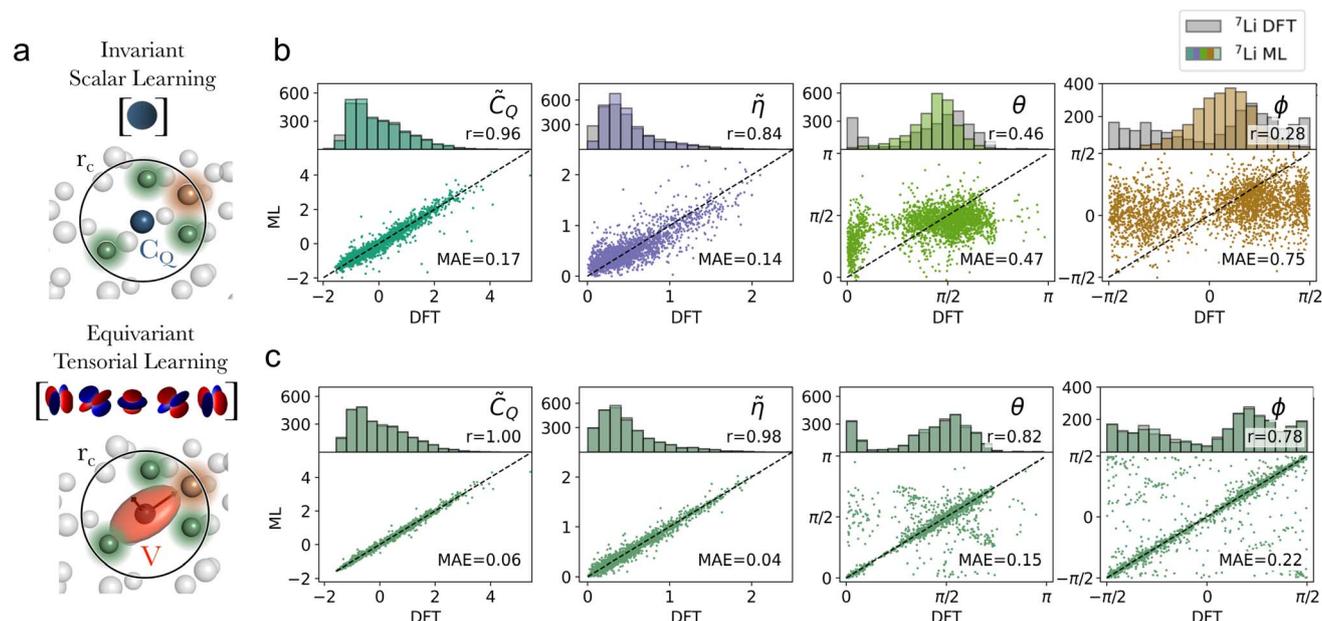
for hyperparameter optimization. The observable-targeted performance metrics ( $\hat{C}_Q$ ,  $\hat{\eta}$ ,  $\mathbf{q}$ ) puts the final observables of interest, namely  $\omega_Q$ , into focus and allows for an evaluation of the accuracy of the ML model with respect to the observable to be predicted. The extension to predicting tensors requires to additionally encode the rotational symmetry of tensorial properties, *i.e.* how the individual tensor elements change under rotations, *cf.* Fig. 3a. A corresponding equivariant description is achieved by the  $\lambda$ -SOAP<sup>41</sup> kernels, which are built on the definition of the scalar SOAP descriptor. These kernels exploit that by transforming the tensor of interest into its irreducible spherical tensor (IST) representation,<sup>58</sup> the GPR procedure can be simplified from

learning on a tensor of order  $\lambda$  into a vector of length  $k = 2\lambda + 1$ . In this representation, all symmetry operations follow the same transformations as the spherical harmonics, shown visually in Fig. 3a, and thus all kernel transformations can be written as Wigner matrices,  $\mathbf{D}^\lambda$ . The corresponding  $k$ -component vector kernel  $\mathbf{K}_{\text{SOAP}}^\lambda$  is then written as

$$\mathbf{K}_{\text{SOAP}}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{\mathbf{R}} \mathbf{D}^\lambda(\hat{\mathbf{R}}) \left| \int \rho(\mathbf{r}) \rho'(\hat{\mathbf{R}}\mathbf{r}) d\mathbf{r} \right|^2, \quad (12)$$

and the GPR learning for the  $k$ -component vector quantities  $\mathbf{y}$  arising in the IST representation of the target tensor generalizes to





**Fig. 3** <sup>7</sup>Li EFG performance metrics with scalar and tensorial learning approaches. (a) Schematic representation of the two learning approaches: scalar learning, encoding invariances of scalar observables such as  $C_Q$ , and tensorial learning, with equivariant encoding of the tensor symmetries of the full EFG tensor  $V$ . Spherical harmonics illustrate the symmetries captured by each descriptor. (b) Results of the scalar learning approach on the <sup>7</sup>Li test set from the LTO-EFG database, using separately trained models for the metrics  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$ , and  $\phi$  (lower panels correlation plots, upper panels value distributions with DFT values in grey and ML predicted values in the corresponding color). (c) Corresponding results of the tensorial learning approach, where a single model is trained on the complete EFG tensor and the four quantities are subsequently derived from the model's predictions, see text.

$$\mathbf{y}(\mathcal{X}) = \sum_{l=1}^N w_l \mathbf{K}_{\text{SOAP}}^{\lambda}(\mathcal{X}, \mathcal{X}_l). \quad (13)$$

For  $\lambda = 0$ , one can show that eqn (12) is equivalent to the scalar representation shown in eqn (10),<sup>41</sup> while for the present case of the EFG rank-2 tensor, five-component vector quantities are being learned that by construction preserve the full tensorial symmetries and physical properties.

For comparability, training and hyperparameter optimization follows the same 4 : 1 training-test set separation and five-fold cross validation scheme as in the scalar learning approach. As only one model is trained this time though, the global Box-Behnken minimization instead uses the average MAE of  $\tilde{C}_Q$ ,  $\tilde{\eta}$ , and the quaternion dot product,  $\tilde{q}$  as performance metrics. The resulting hyperparameters used for the  $\lambda = 2$  kernels are given in the SI, and all  $\lambda = 2$  kernels were constructed with the equivalent SA-GPR framework as for the scalar kernels.<sup>41</sup> As in the scalar learning framework, the models were trained using the explicit  $\lambda$ -SOAP kernels within the GPR framework, to ensure preservation of symmetry throughout the learning procedure. The full derivation of the  $\lambda$ -SOAP hierarchy of tensorial kernels is given in ref. 41.

Again, we obtain a comparable learning performance for all three nuclei, with the results for <sup>47</sup>Ti and <sup>17</sup>O provided in the SI. Fig. 3c summarizes the achieved predictions for all <sup>7</sup>Li in the test set, where for a direct comparison with the scalar learning results, the polar and azimuthal angles  $\theta$  and  $\phi$  were extracted from the learned quaternions. The correlation for  $\tilde{C}_Q$  and  $\tilde{\eta}$  is

now excellent, with Pearson correlations of  $r = 1.00$  and  $r = 0.98$ , respectively, far surpassing the already good correlation achieved in the scalar learning case. The MAE for  $\tilde{C}_Q$  and  $\tilde{\eta}$  is 0.06 and 0.04, respectively. Even more impressive is the improvement in the case of the angles  $\theta$  and  $\phi$ . They now also show a good agreement at only slightly lower Pearson correlations of  $r = 0.82$  and  $r = 0.78$ , respectively. The corresponding MAE is 0.15 radians or an acceptable  $8.6^\circ$  for  $\theta$ , and 0.22 radians or  $13^\circ$  for  $\phi$ , which despite the marked improvement is still significant. The parity plots for  $\theta$  and  $\phi$  in Fig. 3c show systematic outliers, which originate from EFG tensors, where the definition  $|V_{zz}| \geq |V_{yy}| \geq |V_{xx}|$  becomes more ambiguous since the eigenvalues are close in absolute value. When correcting the assignment of the eigenvalues and corresponding eigenvectors as described in the SI, the correlation between DFT reference values and ML predicted values improve considerably with MAE of 0.06 radians ( $3.4^\circ$ ) and 0.05 radians ( $2.9^\circ$ ) for  $\theta$  and  $\phi$ , respectively (*cf.* SI Fig. S6).

## 4 Discussion

The presented results clearly demonstrate the superiority of tensorial learning. At a closer look, they especially reveal that scalar learning struggles with predicting the orientation of the tensor as expressed by the two angles  $\theta$  and  $\phi$ . Admittedly, this seems generally a harder task as compared to the tensor magnitude and shape, as the tensorial learning also exhibits a worse performance for the angles than for  $\tilde{C}_Q$  and  $\tilde{\eta}$ .



Nevertheless, the scalar learning is not even able to qualitatively capture the bimodal value distribution of  $\theta$  and  $\phi$  across the  ${}^7\text{Li}$  test set in the LTO-EFG database, *cf.* Fig. 3b. This bimodality in  $\theta$  and  $\phi$  is even more pronounced for the  ${}^{47}\text{Ti}$  and  ${}^{17}\text{O}$  nuclei as shown in the SI, and there, scalar learning fails completely. This is particularly worrisome, as any NMR experiment will always measure an ensemble average over the atomic sites in the studied material. Now, the distribution of sites with their respective different atomic environments in the highly diverse LTO-EFG database is of course not truly representative of any distribution of atomic environments encountered in even a largely disordered real LTO material. Still, we consider the capability of tensorial learning to reproduce the bimodal angular distributions for all three nuclei across the wide site variety in the challenging LTO-EFG database an important feature that suggests that NMR simulations performed with this surrogate model will indeed yield correct structure–property relationships when applied to an ensemble of realistic structures derived from *e.g.* molecular dynamics simulations. Hence, from an application point of view, the quantitative accuracy of the ML model with respect to the individual data points (MAE) is secondary to the reliability of reproducing the correct distribution of observables. The tensor ML model outperforms the scalar learning in either case, but in particular with regard to the distribution especially of  $\theta$  and  $\phi$ , which indicates improved transferability.

The limitations of scalar learning with respect to capturing the tensor orientation imply that it will be particularly poor for any NMR observable sensitive to this orientation. It thereby does not matter if the tensor orientation is actually explicitly learned as through the scalar performance metrics (of  $\theta$  and  $\phi$ ) introduced above or only implicitly contained in the observable (such as  $\omega_Q$ ) itself. We illustrate this in Fig. 4 for the tensor-orientation sensitive quadrupolar frequency  $\omega_Q$ , *cf.* eqn (5). Fig. 4a compiles the predictions across the  ${}^7\text{Li}$  LTO-EFG test set

when the observable  $\omega_Q$  is scalar learned directly, and following exactly the analog scalar training and optimization protocol as before. The performance is largely comparable and equally poor as the performance obtained when scalar learning the performance metrics and then using eqn (5) to derive  $\omega_Q$ , *cf.* Fig. 4b.

A reliable description is instead only achieved through the full tensorial learning and subsequent extraction of all quantities from the learned  $\mathbf{V}$  to compute  $\omega_Q$  as shown in Fig. 4c. When referring to reliable, we hereby acknowledge that SAE-NMR experiments probe a dynamically averaged value of  $\omega_Q$  and for *e.g.*  ${}^7\text{Li}$  in LTO are sensitive to within roughly 10 kHz.<sup>30</sup> As indicated in Fig. 4, with an MAE of 4.7 kHz only the average accuracy of the tensorial learning approach falls well within this experimental uncertainty range and thus allows to make meaningful predictions. In contrast, both scalar learning approaches to  $\omega_Q$  exhibit MAEs above 10 kHz.

Intriguingly, the tensorial learning approach is not only more accurate, but also more data efficient. This extends to having to only train one versatile model to predict any tensor-derived observables, as well as to the required amount of training data. As shown in Fig. 5a, the scalar learning approach for the four quantities  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$ , and  $\phi$  exhibits a low learning rate across the number of  ${}^7\text{Li}$  data used in the training. This learning rate is much higher when predicting these quantities from the learned tensor  $\mathbf{V}$ , which in turn lowers the amount of training data required to arrive at the same or even better model accuracy. The MAE for the individual tensor components  $V_{ij}$ , which is additionally shown in Fig. 5b, has a percent error in the same order of magnitude as the quantities  $\tilde{C}_Q$  and  $\tilde{\eta}$ ,  $\theta$ , and  $\phi$  of the performance metrics. Yet, only assessing the performance of the model with respect to  $V_{ij}$  does not give any information about the accuracy for the quantities relevant to experimental applications.

Having ensured that our tensorial learning approach for EFG tensors is both accurate and efficient, we now extend its use to

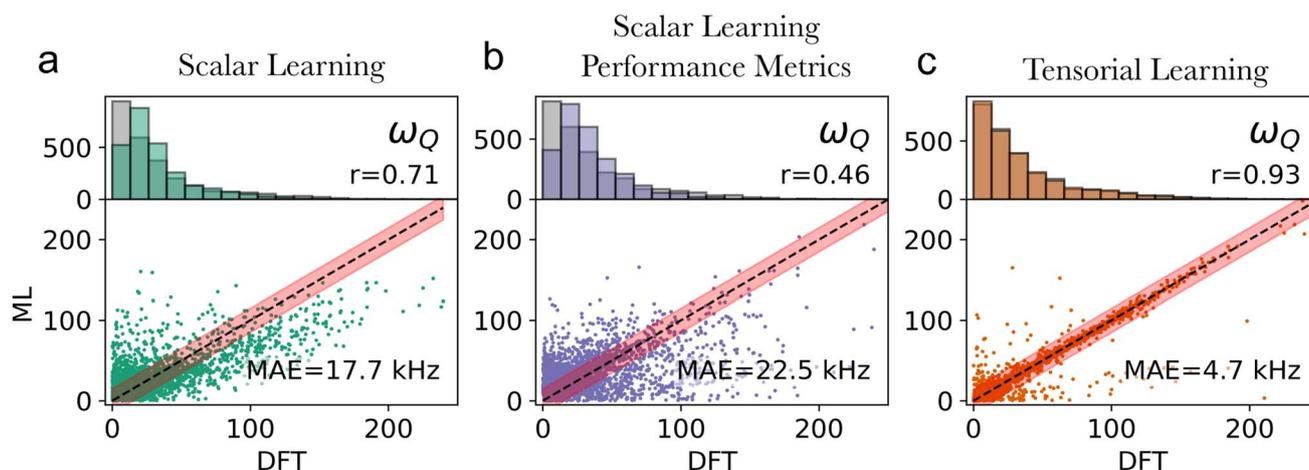


Fig. 4 Predictions for the tensor-orientation sensitive quadrupolar frequency  $\omega_Q$ . Results for the NMR observable  $\omega_Q$  across the  ${}^7\text{Li}$  LTO-EFG test set withheld from the training for three learning models: (a) direct scalar learning of  $\omega_Q$ , (b)  $\omega_Q$  derived from scalar learning the performance metrics  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$ ,  $\phi$  as in Fig. 3b, (c)  $\omega_Q$  derived from tensorial learning of  $\mathbf{V}$  as in Fig. 3c (lower panels correlation plots, upper panels value distributions with DFT values in grey and ML predicted values in the color). The red stripe in the correlation plots roughly denotes the error range of SAE-NMR experiments, see text.



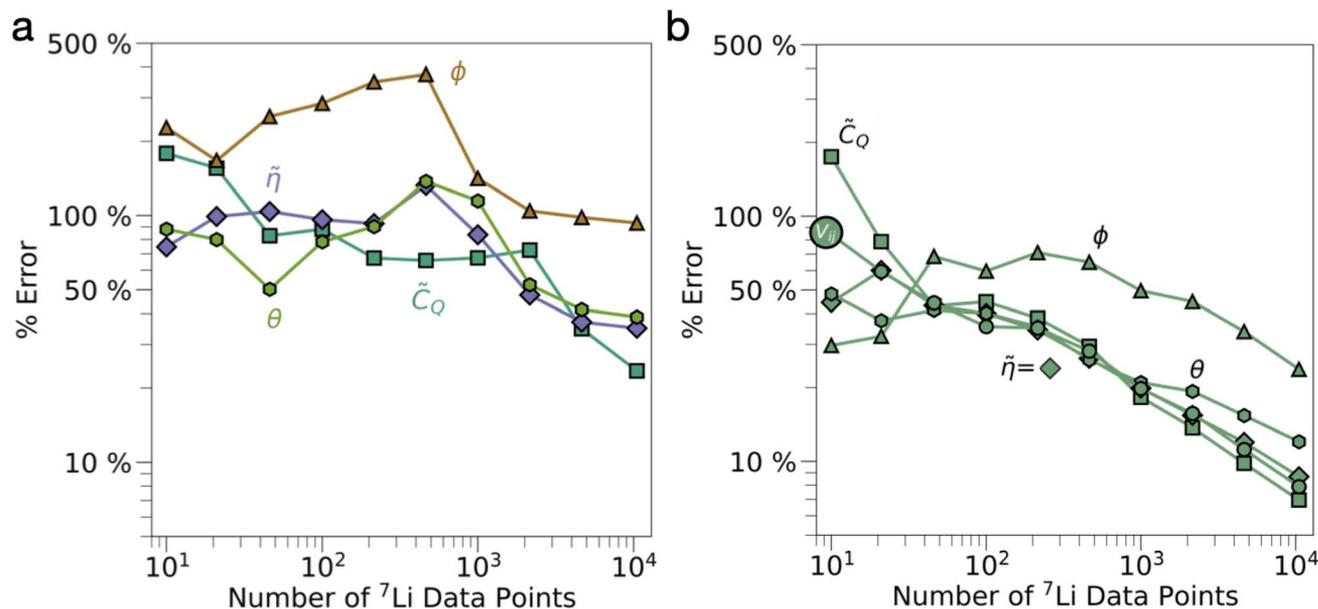


Fig. 5 Learning curves of the <sup>7</sup>Li EFG performance metrics for scalar and tensorial learning approaches. (a) Learning curves for the four independent scalar learning models for  $\tilde{C}_Q$ ,  $\tilde{\eta}$ ,  $\theta$ ,  $\phi$  shown in Fig. 3b. (b) Corresponding curves when extracting these quantities from the tensorial learning model in Fig. 3c. The line color corresponds to the model colors introduced in Fig. 3(b and c).

larger, realistic 1512 atom structures of LTO, in order to establish whether the model is also size extensive and effective at distinguishing local Li environments at scale. We chose to select a set of 20 low-energy structures of LTO from the database created by Heenen *et al.* in ref. 59. These structures were generated using a combination of Metropolis Monte Carlo and Wang Landau sampling<sup>60,61</sup> over the disordered Li/Ti sites

within the  $Fd\bar{3}m$  bulk unit cells. Given the computational efficiency of the EFG tensorial model, we are able to successfully predict full EFG tensors for the 11 000 Li sites within these 20 structures in under 24 hours on a single CPU.

As shown in Fig. 6, we distinguish different Li local environments based on similar predicted  $\omega_Q$  values. The colored regions in Fig. 6a are selected by first identifying the 5 most

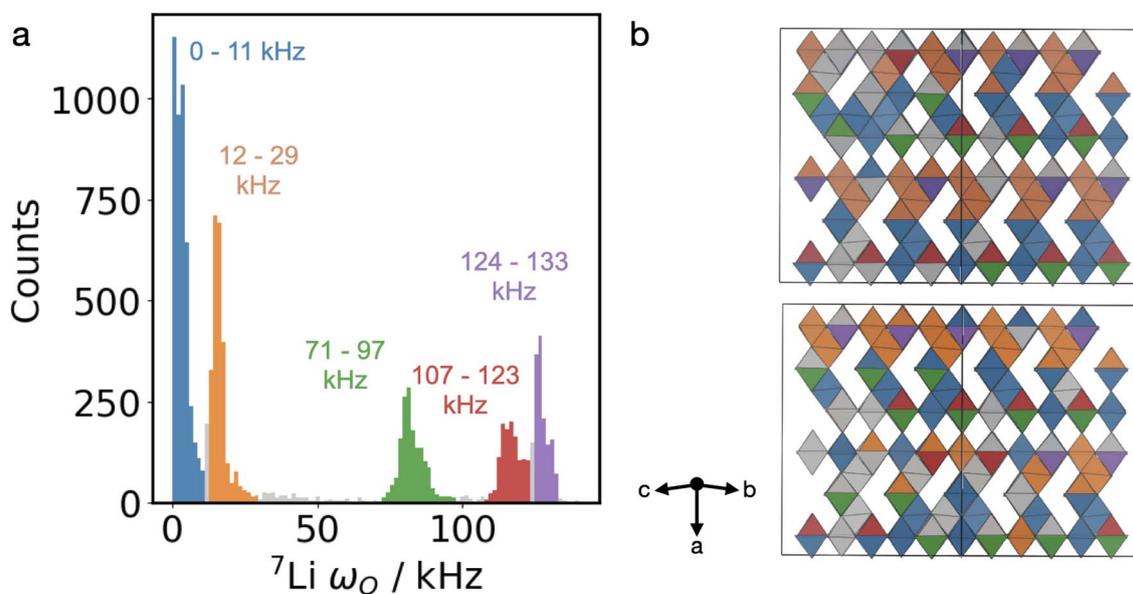


Fig. 6 Histogram of predicted  $\omega_Q$  for 1512-atom structures of  $\text{Li}_4\text{Ti}_5\text{O}_{12}$ . (a) The histogram shows the predictions for all 11 808 Li sites in the 20 lowest energy structures of LTO extracted from (ref. 59). The colored regions distinguish 5 different regions of  $\omega_Q$  which are separated based on the 5 most prominent peaks in the histogram. Any regions in grey were not classified into a distinct region. (b) Two sample structures of LTO, where the O and Ti atoms are not shown, and the Li local environments are colored based on their corresponding region in the histogram in (a).



prominent peaks in the histogram, and selecting the surrounding regions down to a minimum threshold of 10 counts per bin, and enforced that no two regions could overlap. This allows us to visualize the large LTO structures as shown in Fig. 6b, where we can see that local regions with similar Li  $\omega_Q$  have similar local environments.

Validating the simulated EFG tensors and the derived structure–property relationship with experimental measurements is unfortunately not possible, since NMR measures ensemble and temporal averages of all the  $^7\text{Li}$  nuclei present. Experimental results from SAE experiments on LTO yield an averaged, residual  $\omega_Q$  in the range of 10 kHz to 40 kHz,<sup>30</sup> while the computed values range up to 130 kHz (cf. Fig. 6) and the full dataset ranges up to 500 kHz. In order to derive the experimentally accessible observables, the tensor ML-model has to be combined with dynamics simulations such as molecular dynamics or kinetic Monte Carlo to reproduce the correct physical averaging observed in experiments.<sup>17</sup>

Because each of the identified local environments are distinguishable by their quadrupolar frequencies, it is furthermore possible to study Li diffusion between the sites in LTO using SAE NMR, and extract different Li ion mobilities between each of these 5 sites. There is strong evidence to suggest that the local configurational disorder within LTO is the main driver for Li ion mobility within the fast-ion conductor.<sup>59</sup> Therefore by studying LTO experimentally with SAE NMR, we suggest that it would be possible to experimentally confirm the modes of Li diffusion within the material between each of the local, distinct  $\omega_Q$  environments. This ability to predict experimental observables on realistic structures by combining complex structural models with extended dynamical trajectories is only made possible using our tensorial learning approach.

## 5 Conclusion

From the nature of all our results, it is clear that analogous conclusions would be obtained when using both invariant and equivariant equivalents to SOAP.<sup>62–66</sup> The decisive factor for the performance of the ML model is in whether it preserves the symmetry of the tensor or not, rather than in the style of learning or exact formulation of the descriptor employed. With an appropriate encoding of physical symmetries and the use of observable-targeting performance metrics for hyperparameter optimization and model evaluation, accurate predictions of experimental tensor-derived observables are reachable. We have exemplified this for the important case of NMR observables, achieving results within experimental precision over an extensive DFT-computed EFG database for the LTO battery material that is representative of the diversity of atomic environments encountered in real experiments. The tensorial learning excels particularly in the prediction of angular-dependent observables such as the quadrupolar frequency. This is especially impactful for the simulation and interpretation of advanced solid-state NMR experiments, which measure correlations between tensors on long length and time scales and are therefore highly sensitive to tensor orientation. Tensorial ML approaches are indispensable in order to keep pace with machine learned

interatomic potentials (MLIP), which are currently revolutionizing the capabilities of structural ensemble simulations and molecular dynamics approaching spectroscopically relevant time and length scales.<sup>17</sup> To bridge the gap between theoretical simulation models and experimental results, the MLIP derived structural and temporal ensembles have to be mapped to experimental observables with an approach of comparable computational efficiency to MLIPs so as not to nullify the computational advantage gained by MLIPs.

By predicting on a 1512 atom model of LTO, we have shown that this method is scalable and efficient for realistic systems. Beyond NMR quadrupolar observables, this method is general and applies to any tensorial property. Chemical shielding tensors, stress and strain tensors, and polarizability tensors are some of many, commonly used experimental properties which would now be readily predictable using this ML workflow. We expect that our positive results and guiding principles for context-aware performance metrics encourage the adoption of tensor-based ML approaches for all tensorial spectroscopy.

## Author contributions

AFH data curation, formal analysis, methodology, visualization, writing – original draft, writing – review and editing. SSK supervision, validation, visualization, writing – review and editing. KR writing – review and editing, funding acquisition, supervision, resources. CS supervision, writing – review and editing, resources, project administration, conceptualization.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The LTO-EFG dataset will be published open-access on the EDMOND MPG data service at <https://doi.org/10.17617/3.TXBK3G> after acceptance.

Supplementary information is available. See DOI: <https://doi.org/10.1039/d5ta05090a>.

## Acknowledgements

All computational resources were provided by the Max Planck Computing and Data Facility (MPCDF). All the authors would like to thank Prof. Rüdiger-A. Eichel and Prof. Josef Granwehr for scientific discussions and valuable input. AFH would like to acknowledge the support of the Alexander von Humboldt Foundation (Henriette-Hertz Programm). CS acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the cluster of excellence EXC 2089: e-conversion, project number 390776260.

## Notes and references

- 1 K. Märker, C. Xu and C. P. Grey, *J. Am. Chem. Soc.*, 2020, **142**, 17447.



- 2 H. Sun, S. Dwaraknath, H. Ling, X. Qu, P. Huck, K. A. Persson and S. E. Hayes, *npj Comput. Mater.*, 2020, **6**, 53.
- 3 O. Pecher, J. Carretero-Gonzalez, K. J. Griffith and C. P. Grey, *Chem. Mater.*, 2017, **29**, 213.
- 4 W. Shin, J. C. Garcia, A. Vu, X. Ji, H. Iddir and F. Dogan, *J. Phys. Chem. C*, 2022, **126**, 4276.
- 5 A. F. Harper, S. P. Emge, P. C. Magusin, C. P. Grey and A. J. Morris, *Chem. Sci.*, 2023, **14**, 1155.
- 6 I. Chubak, L. Alon, E. V. Silletta, G. Madelin, A. Jerschow and B. Rotenberg, *Nat. Commun.*, 2023, **14**, 84.
- 7 S. E. Ashbrook and D. McKay, *Chem. Commun.*, 2016, **52**, 7186.
- 8 D. L. Bryce, *IUCr*, 2017, **4**, 350.
- 9 J. Serrano-Sevillano, D. Carlier, A. Saracibar, J. M. Lopez del Amo and M. Casas-Cabanas, *Inorg. Chem.*, 2019, **58**, 8347.
- 10 M. Seifrid, G. M. Reddy, B. F. Chmelka and G. C. Bazan, *Nat. Rev. Mater.*, 2020, **5**, 910.
- 11 B. Thomas, B. S. Chang, J. J. Chang, M. Thuo and A. J. Rossini, *Chem. Mater.*, 2022, **34**, 7678.
- 12 Y. Yasui, M. Tansho, K. Fujii, Y. Sakuda, A. Goto, S. Ohki, Y. Mogami, T. Iijima, S. Kobayashi, S. Kawaguchi, *et al.*, *Nat. Commun.*, 2023, **14**, 2337.
- 13 R. F. Moran, D. McKay, P. C. Tornstrom, A. Aziz, A. Fernandes, R. Grau-Crespo and S. E. Ashbrook, *J. Am. Chem. Soc.*, 2019, **141**, 17838–17846.
- 14 T. Charpentier, M. C. Menziani and A. Pedone, *RSC Adv.*, 2013, **3**, 10550–10578.
- 15 J. M. Griffin, J. R. Yates, A. J. Berry, S. Wimperis and S. E. Ashbrook, *J. Am. Chem. Soc.*, 2010, **132**, 15651–15660.
- 16 T. Charpentier, *Faraday Discuss.*, 2025, **255**, 370–390.
- 17 A. F. Harper, T. Huss, S. S. Köcher and C. Scheurer, *Faraday Discuss.*, 2025, **255**, 411–428.
- 18 M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen and B. Hammer, *J. Phys. Chem. A*, 2018, **122**, 1504.
- 19 C. Nyshadham, M. Rupp, B. Bekker, A. V. Shapeev, T. Mueller, C. W. Rosenbrock, G. Csányi, D. W. Wingate and G. L. Hart, *npj Comput. Mater.*, 2019, **5**, 51.
- 20 V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti and G. Csányi, *Chem. Rev.*, 2021, **121**, 10073.
- 21 W. Xu, K. Reuter and M. Andersen, *Nat. Comput. Sci.*, 2022, **2**, 443.
- 22 H. Jung, L. Sauerland, S. Stocker, K. Reuter and J. T. Margraf, *npj Comput. Mater.*, 2023, **9**, 114.
- 23 K. Chen, C. Kunkel, B. Cheng, K. Reuter and J. T. Margraf, *Chem. Sci.*, 2023, **14**, 4913.
- 24 Y. Zhuo, A. Mansouri Tehrani and J. Brgoch, *J. Phys. Chem. Lett.*, 2018, **9**, 1668.
- 25 M. Veit, D. M. Wilkins, Y. Yang, R. A. DiStasio and M. Ceriotti, *J. Chem. Phys.*, 2020, **153**, 024113.
- 26 M. C. Venetos, M. Wen and K. A. Persson, *J. Phys. Chem. A*, 2023, **127**, 2388.
- 27 H. Sun, S. Dwaraknath, H. Ling, K. A. Persson and S. E. Hayes, *Sci. Rep.*, 2025, **15**, 26456.
- 28 F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti and L. Emsley, *Nat. Commun.*, 2018, **9**, 4501.
- 29 E. A. Engel, V. Kapil and M. Ceriotti, *J. Phys. Chem. Lett.*, 2021, **12**, 7701–7707.
- 30 M. F. Graf, H. Tempel, S. S. Köcher, R. Schierholz, C. Scheurer, H. Kungl, R.-A. Eichel and J. Granwehr, *RSC Adv.*, 2017, **7**, 25276.
- 31 K. Hogrefe, N. Minafra, W. G. Zeier and H. M. R. Wilkening, *J. Phys. Chem. C*, 2021, **125**, 2306.
- 32 P. P. M. Schleker, C. Grosu, M. Paulus, P. Jakes, R. Schlögl, R.-A. Eichel, C. Scheurer and J. Granwehr, *Commun. Chem.*, 2023, **6**, 113.
- 33 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. Probert, K. Refson and M. C. Payne, *Z. Kristallogr. Cryst. Mater.*, 2005, **220**, 567.
- 34 J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **76**, 024401.
- 35 M. Profeta, F. Mauri and C. J. Pickard, *J. Am. Chem. Soc.*, 2003, **125**, 541.
- 36 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 37 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 136406.
- 38 A. P. Bartók and J. R. Yates, *J. Chem. Phys.*, 2019, **150**, 161101.
- 39 J. Valenzuela Reina, F. Civaia, A. F. Harper, C. Scheurer and S. S. Köcher, *Faraday Discuss.*, 2025, **255**, 266–287.
- 40 D. Wilkins, A. Grisafi, A. Anelli, G. Fraux, J. Nigam, E. Baldi, L. Folkmann, M. Ceriotti, *TENSOAP - SA-GPR*, 2023, <https://github.com/dilkins/TENSOAP>.
- 41 A. Grisafi, D. M. Wilkins, G. Csányi and M. Ceriotti, *Phys. Rev. Lett.*, 2018, **120**, 036002.
- 42 R. K. Harris, E. D. Becker, S. M. Cabral De Menezes, P. Granger, R. E. Hoffman and K. W. Zilm, *Pure Appl. Chem.*, 2008, **80**, 59.
- 43 S. Stoll and D. Goldfarb, *EPR Spectroscopy: Fundamentals and Methods*, 2018, vol. 95.
- 44 J. Emsley and J. Feeney, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2007, **50**, 179.
- 45 P. P. Man, *Encycl. Anal. Chem.*, 2000, **10**, 12224.
- 46 A. Lunghi and S. Sanvito, *J. Phys. Chem. C*, 2020, **124**, 5802.
- 47 M. Domina, M. Cobelli and S. Sanvito, *Phys. Rev. B*, 2022, **105**, 214439.
- 48 V. H. A. Nguyen and A. Lunghi, *Phys. Rev. B*, 2022, **105**, 165131.
- 49 S. W. Shepperd, *J. Guid. Control Dyn.*, 1978, **1**, 223.
- 50 W. Gerrard, L. A. Bratholm, M. J. Packer, A. J. Mulholland, D. R. Glowacki and C. P. Butts, *Chem. Sci.*, 2020, **11**, 508.
- 51 P. A. Unzueta, C. S. Greenwell and G. J. Beran, *J. Chem. Theory Comput.*, 2021, **17**, 826.
- 52 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, *J. Phys. Condens. Matter*, 2017, **29**, 273002.
- 53 K. Choudhary, J. N. Ansari, I. I. Mazin and K. L. Sauer, *Sci. Data*, 2020, **7**, 362.
- 54 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
- 55 R. Drautz, *Phys. Rev. B*, 2019, **99**, 014104.



- 56 Y. Lysogorskiy, C. V. D. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, *et al.*, *npj Comput. Mater.*, 2021, **7**, 97.
- 57 G. E. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, 1987.
- 58 U. Weinert, *Arch. Ration. Mech. Anal.*, 1980, **74**, 165.
- 59 H. H. Heenen, C. Scheurer and K. Reuter, *Nano Lett.*, 2017, **17**, 3884–3888.
- 60 D. Frenkel and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications*, Elsevier, 2023.
- 61 F. Wang and D. Landau, *Phys. Rev. E*, 2001, **64**, 056101.
- 62 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.
- 63 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.
- 64 F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi and M. Ceriotti, *Chem. Rev.*, 2021, **121**, 9759.
- 65 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360.
- 66 K. T. Schütt, H. E. Saucedo, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.

