THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Machine Learning Full NMR Chemical Shift Tensors of Silicon Oxides with Equivariant Graph Neural Networks

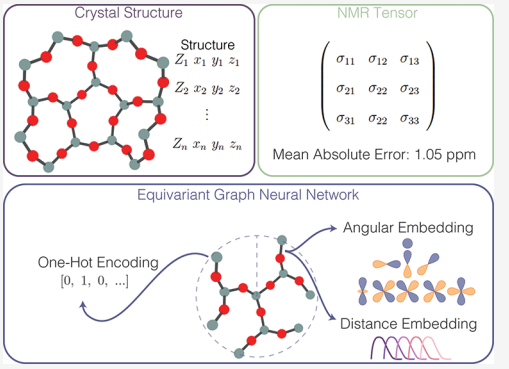
Maxwell C. Venetos, Mingjian Wen, and Kristin A. Persson*

Cite This: https://doi.org/10.1021/acs.jpca.2c07530

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The nuclear magnetic resonance (NMR) chemical shift tensor is a highly sensitive probe of the electronic structure of an atom and furthermore its local structure. Recently, machine learning has been applied to NMR in the prediction of isotropic chemical shifts from a structure. Current machine learning models, however, often ignore the full chemical shift tensor for the easier-to-predict isotropic chemical shift, effectively ignoring a multitude of structural information available in the NMR chemical shift tensor. Here we use an equivariant graph neural network (GNN) to predict full ^{29}Si chemical shift tensors in silicate materials. The equivariant GNN model predicts full tensors to a mean absolute error of 1.05 ppm and is able to accurately determine the magnitude, anisotropy, and tensor orientation in a diverse set of silicon oxide local structures. When compared with other models, the equivariant GNN model outperforms the state-of-the-art machine learning models by 53%. The equivariant GNN model also outperforms historic analytical models by 57% for isotropic chemical shift and 91% for anisotropy. The software is available as a simple-to-use open-source repository, allowing similar models to be created and trained with ease.



INTRODUCTION

Many useful properties of materials manifest from the precise structure of a given composition. Traditional structure determination techniques such as X-ray diffraction (XRD) are suitable for atoms of moderate to high atomic number; however, they can lead to ambiguous structure assignments for materials containing atoms with low atomic number.[1] In addition, XRD relies heavily on long-range order for correct measurement, but such long-range order is often lacking in many classes of materials, e.g., nanostructures, amorphous materials, and materials with a tetrahedral network, making structural characterization of such materials via XRD difficult. Such a class of materials are exemplified by silicates, which consist of a tetrahedral structure of silicons and oxygens (which have low atomic number and are difficult to observe via XRD).

Silicate materials are ubiquitous, from naturally occurring rocks and minerals like quartz[2−5] and garnet[6] to diverse manufacturing applications such as glasses,[7,8] cements,[9] and zeolite catalysts.[10,11] To discover new potential applications of silicates, accurate elucidation of their structures is a prerequisite.

Nuclear magnetic resonance (NMR) spectroscopy has become a reliable tool for structural investigations in such materials. As a spectroscopy technique, NMR is highly sensitive to the electron density about an atom and relies on local structure rather than any long-range order. NMR measurements are typically combined with powder XRD measurements and *ab initio* simulations to obtain refined crystal structures in a technique termed *NMR crystallography*.[12−22] These refinement procedures, however, often take an expensive iterative approach, as many *ab initio* NMR calculations are repeated until the results converge. Despite advances in computational power and algorithmic efficiencies, the NMR calculation is still expensive and time-consuming.[23]

Machine learning (ML) has increasingly been shown to be useful for providing high-quality predictions for material properties but with orders of magnitude less computational demand.[24−28] ML techniques have recently been applied to NMR, with most applications focused on organic molecules[29−34] and a few focused on ^{29}Si chemical shift prediction.[35,36] With ^{29}Si NMR databases becoming more widely available[37,38] machine learning will begin to take a greater role in structural studies. While chemical shifts are useful, as they correlate to the average electronic environment about an atom, shifts are only one piece of the spectrum. The line shape observed in an NMR measurement is described by a tensor, of which the chemical shift is the isotropic part. By

Received: October 26, 2022
Revised: January 19, 2023

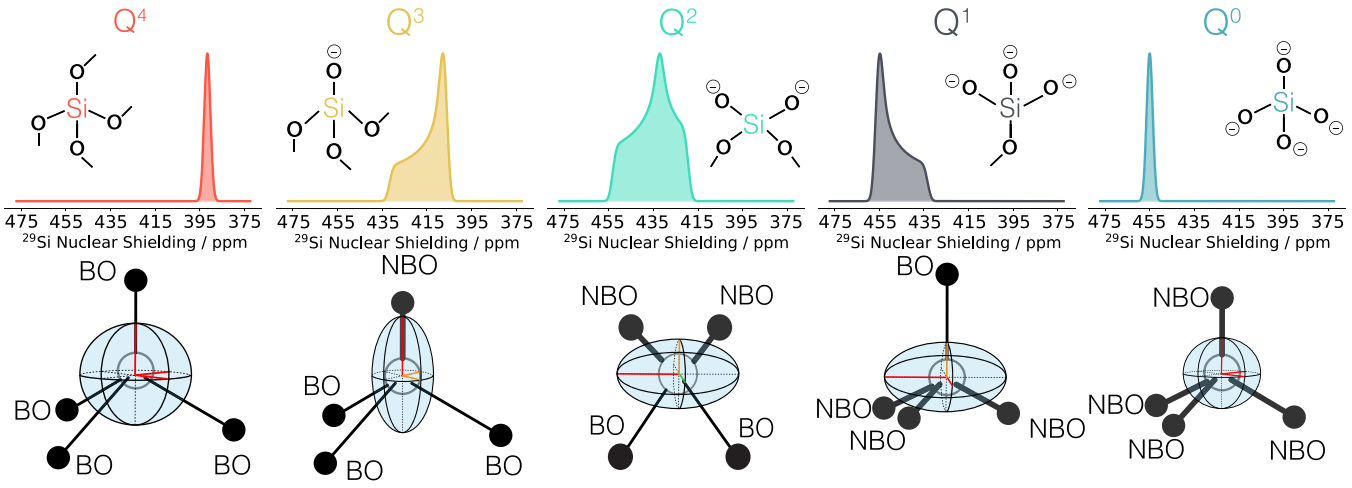


Figure 1. Structural and spectral representations of the NMR chemical shift tensors of different Q^n species. The structural and geometric variety due to the differences in bridging and nonbridging oxygen bonds results in different chemical environments for each Q^n and thus different chemical shift tensors.

ignoring the tensorial nature of the NMR measurement, a myriad of structural information is lost, which is an issue for many previous ML models, as they are only capable of predicting scalar quantities.

Recent advances in the field of geometric deep learning have allowed for prediction of tensorial targets.[39−41] These models seek to constrain the functions used internally in the model such that only functions that respect the symmetry of the target object can be learned. In the work herein, we present an equivariant GNN model capable of predicting full NMR shielding tensors and demonstrate its capability via ^{29}Si nuclear chemical shift tensors in silicate materials.

In the first part of this work we will assess the use of more traditional ML methods to attempt to learn NMR tensor parameters in order to demonstrate that a symmetry-invariant model is insufficient for NMR tensor parameter prediction. We show that all of the symmetry-invariant models trained herein fall short of predicting tensor elements and that instead a symmetry-equivariant model is needed to respect the symmetries of the tensor, as evidenced by a 53% (3.05 ppm vs 6.44 ppm) decrease in mean absolute error compared to invariant models. In the second part of this work we comprehensively assess the performance of our equivariant model. We show that it can learn the full NMR chemical shift tensor to a mean absolute error (MAE) of 1.05 ppm. When converted to the scalar isotropic chemical shift (which previous symmetry-invariant models can predict), our equivariant GNN model outperforms the state-of-the-art model by a large margin, with an MAE of 2.82 ppm versus 5.87 ppm.

We also assess the predicted tensors to show that the equivariant GNN model is capable of learning both tensor magnitude and shape as well as the tensor orientation in a diverse set of silicon local structures.

METHODS

Dataset. The ^{29}Si NMR dataset used in this study is a subset of *ab initio* NMR chemical shift tensors of relaxed structures calculated by Sun et al.[38] The dataset is composed of oxygen-coordinated silicon tetrahedral networks consisting of SiO_2 along with silicates containing group 1 and 2 cations (Li^+, Na^+, Mg^{2+}, etc.). It contains a wide variety of structures with different numbers of bridging oxygen atoms, n, commonly denoted as Q^n, as shown in Figure 1. Each Q^n species has a different chemical environment and local point-group symmetry due to the differing bond lengths to bridging oxygen (BO) and nonbridging oxygen (NBO), which in effect results in different chemical shift tensor symmetries. In total there are 421 unique silicate structures, consisting of 1387 unique silicon sites. The silicon sites consist of 874 Q^4 sites, 174 Q^3 sites, 172 Q^2 sites, 32 Q^1 sites, and 97 Q^0 sites. From each site, the raw calculated rank-2 asymmetric chemical shift tensor was extracted and processed in accordance with each tensor space or tensor convention used in the training of the ML models outlined below.

For each ML model, a train, validation, and test split of 8:1:1 was used. As the dataset consists of multiple types of sites (i.e., each Q^n species), an attempt was made to stratify the data such that there would be an approximately equal weighting of each Q^n species in each of the training, validation, and test sets. To remove any opportunity for data leakage, stratification was done at the structure level rather than the site level. Due to structures containing multiple sites, often with different Q^n, each structure was given a label based on whichever n was least common in the dataset. For example, in a structure with both a Q^1 and Q^2 site, the structure would be labeled as Q^1 because the dataset consists of a smaller number of Q^1 sites (32) than Q^2 sites (172). Structures were then randomly stratified to give roughly equal proportions of each n type in each set.

Chemical Shift Tensor Conventions. Often in the context of NMR, the quantity of interest is a scalar value, the *isotropic chemical shift*, δ^{iso}, but it is important to keep in mind that the chemical shift is a tensor quantity, formally an antisymmetric second-rank tensor. Typically, only the symmetric part of the tensor is used, as the symmetric tensor influences the line shapes seen in the NMR spectrum. The distinction between nuclear shielding and chemical shift should also be noted. Nuclear shielding describes the relative change in magnetic field about a nuclear position with respect to the external field and is the quantity calculated during an *ab initio* calculation. In NMR experiments, however, the shielding is not measured directly, and instead, the common practice is to measure the chemical shift as the difference in resonant frequencies between the nucleus of interest and a reference compound.

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

While nuclear shielding tensors are the typical quantities calculated by *ab initio* methods, we elect to instead use the *absolute chemical shift tensor*, which is the quantity calculated by VASP[42] and the form originally reported for the database from which the data used herein was obtained.[38] The formal relationship between the absolute chemical shift tensor, $\boldsymbol{\delta}$, and the nuclear shielding tensor, $\boldsymbol{\sigma}$, is given by[43]

$$\boldsymbol{\delta} = \mathbf{1}\sigma^{\text{iso}} - \boldsymbol{\sigma}$$

where σ^{iso} is the isotropic nuclear shielding, which is defined as the average of the trace of the nuclear shielding tensor, given as

$$\sigma^{\text{iso}} = \frac{1}{3}\,\text{Tr}(\boldsymbol{\sigma}) = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33})$$

or in the case of the isotropic chemical shift,

$$\delta^{\text{iso}} = \frac{1}{3}\,\text{Tr}(\boldsymbol{\delta}) = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33}) \tag{1}$$

To convert from the absolute chemical shift tensor to the more familiar referenced chemical shift tensor, $\boldsymbol{\delta}^{\text{referenced}}$, the nuclear shielding tensor of the reference compound, $\boldsymbol{\sigma}^{\text{ref}}$, must be calculated and added to the absolute chemical shift tensor:

$$\boldsymbol{\delta}^{\text{referenced}} = \boldsymbol{\sigma}^{\text{ref}} + \boldsymbol{\delta}$$

For more information on NMR conventions, the reader is directed to the numerous reviews and textbooks on the topic.[1,44,45]

In symmetry-invariant ML models, an assortment of tensor conventions are used as the training targets. The targets will, in each case, consist of three parameters: the isotropic chemical shift and two additional parameters used to describe the shape of the tensor. The two additional parameters broadly fall into two categories based on the ordering of the principal components of the chemical shift tensor in the principal axis system. The complete list of the conventions used herein is presented in Table 1.

Of the conventions used, the Maryland ($\Omega\kappa$) and Haeberlen ($\zeta\eta$) conventions are the most commonly found, as both are recommended for reporting by IUPAC.[46] It should be noted that Haeberlen is occasionally presented as ($\Delta\delta\eta$), but the ($\zeta\eta$) definition is far more common. In addition to Haeberlen ($\Delta\delta\eta$), another less practiced convention, the axiality/rhombicity (AxRh) convention,[47] is sometimes used in spin dynamics and spin relaxation theory, as the parameters come from the irreducible spherical tensor expansion of an interaction Hamiltonian. In addition to the four conventions listed in Table 1, we also investigate learning directly on the principal axes (i.e., δ_{11}, δ_{22}, and δ_{33}) in the standard convention to give a total of five conventions investigated for the invariant property predictions.

Table 1. Chemical Shift Tensor Conventions, Which Fall into Two Categories Based on the Principal Axis Labeling Scheme: Standard Ordering (δ_{11}, δ_{22}, δ_{33}) and Haeberlen Ordering (δ_{XX}, δ_{YY}, δ_{ZZ})

convention name	parameter 1	parameter 2	
Standard Ordering: $\delta_{11} \geq \delta_{22} \geq \delta_{33}$			
Maryland ($\Omega\kappa$)	$\Omega = \delta_{11} - \delta_{33}$	$\kappa = \frac{3}{2}\frac{\delta^{\text{iso}} - \delta_{22}}{\Omega}$	(2)
axiality/rhombicity (AxRh)	$\text{Ax} = 2\delta_{11} - (\delta_{33} + \delta_{22})$	$\text{Rh} = \delta_{22} - \delta_{33}$	(3)
Haeberlen Ordering: $\lvert\delta_{ZZ} - \delta^{\text{iso}}\rvert \geq \lvert\delta_{XX} - \delta^{\text{iso}}\rvert \geq \lvert\delta_{YY} - \delta^{\text{iso}}\rvert$			
convention name	parameter 1	parameter 2	
Haeberlen ($\Delta\delta\eta$)	$\Delta\delta = \delta_{ZZ} - \frac{1}{2}(\delta_{XX} + \delta_{YY})$	$\eta = \frac{3(\delta_{YY} - \delta_{XX})}{2\Delta\delta}$	(4)
Haeberlen ($\zeta\eta$)	$\zeta = \delta_{ZZ} - \delta^{\text{iso}}$	$\eta = \frac{\delta_{YY} - \delta_{XX}}{\zeta}$	(5)

For our new symmetry-equivariant GNN model, we use both symmetric and asymmetric tensor spaces as targets. In addition to the traditional Cartesian tensors, we also use irreducible representations (irreps) of the Cartesian tensor, $\boldsymbol{\delta}$, which are defined as[48]

$$\begin{aligned} E &= \frac{1}{3}\,\text{Tr}(\boldsymbol{\delta}) \\ A_{ik} &= \delta_{ik} - \delta_{ki} \\ S_{ik} &= \frac{1}{2}(\delta_{ik} + \delta_{ki}) - \frac{1}{3}\,\text{Tr}(\boldsymbol{\delta})\delta_{ik} \end{aligned} \tag{6}$$

where E is the isotropic part of the tensor, A_{ik} is the traceless-antisymmetric part of the tensor, S_{ik} is the traceless-symmetric part of the tensor, and δ_{ik} is the Kronecker delta.

Additionally, in the case of the symmetry-equivariant models, a decision must be made on which representation is to be used for interpreting the results. The MAE is satisfactory for a machine to come to an optimized solution but tells us nothing about how well we learned the tensors. A symmetric second-rank tensor has six independent parameters, each of which will need to be assessed. The spherical tensor elements are ideal but are difficult to interpret, as are the tensor indices themselves. The IUPAC-recommended conventions may not be optimal either. The Maryland convention is a descriptor of the line shape as a statistical distribution and is only applicable in specific cases and lacks generalization. The Haeberlen convention is based on the tensor itself and is generally true to the chemical shift tensor; however, the convention requires the ζ parameter to be sign-invariant at $\eta = 1$, which creates a degeneracy. Instead, we turn toward a convention recently proposed by Srivastava and Grandinetti[49] to improve on the issues of the Haeberlen convention. In this convention, ζ and η from the Haeberlen ($\zeta\eta$) convention are first mapped to a polar grid:

$$\begin{aligned} r_\zeta &= |\zeta| \\ \theta &= \begin{cases} \frac{\pi}{4}\eta & \zeta \leq 0 \\ \frac{\pi}{2}\left(1 - \frac{\eta}{2}\right) & \zeta > 0 \end{cases} \\ 0 \geq \theta &\geq \frac{\pi}{2} \end{aligned}$$

Then the polar grid is mapped onto the first quadrant of the Cartesian grid to give X and Y, where

$$\begin{aligned} X &= r_\zeta \cos\theta \\ Y &= r_\zeta \sin\theta \end{aligned} \tag{7}$$

For interpretation and discussion of the tensors, we use the isotropic chemical shift (magnitude of the tensor), X and Y

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

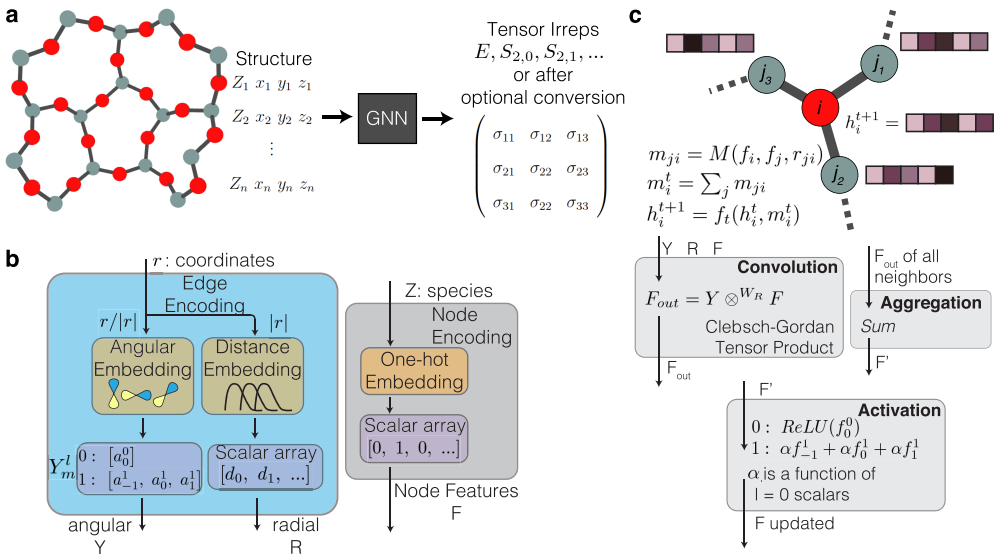


Figure 2. Schematic of an equivariant graph neural network. Each component of the model can have different choices, and the examples here are for the herein-used TFN-based model. (a) Black-box overview of the model, where a structure (represented by a set of atomic numbers, *Z*, and the associated *xyz* coordinates) is passed in as model input and a tensor (which may be represented as tensor irreps or converted to a Cartesian tensor) is obtained as model output. (b) The embedding procedure, in which edge encoding utilizes atomic coordinates to build angular embeddings from a set of spherical harmonics and distance embeddings from a set of learned distance functions. Node encoding utilizes one-hot encoding of the atomic number. (c) The message passing update procedure, in which the edge and node encodings from the neighboring atoms are passed through a convolution filter, aggregated via a sum, and passed through an activation function to update the feature vector on the atom of interest.

(shape/anisotropy of the tensor), and the Euler angles between the laboratory frame and the molecular frame using the *ZYZ* ordering (α, β, γ) (orientation of the tensor).

Machine Learning Models. *Graph Neural Networks.* In a chemical graph neural network (GNN), shown in Figure 2a, a crystalline structure can be represented as a graph in which each atom is represented by a node, v, and relationships between nodes are represented by edges, e, which are commonly considered to be chemical bonds, Coulombic interactions, etc. Often the notion of a chemical bond is ill-defined in a crystalline structure, so edges are frequently constructed as all pairwise node connections within some cutoff radius, r_{cut}, about each atom, taking into account the periodic boundary conditions of the system. All together, the set of nodes, $V = \{v_1, v_2, \ldots, v_N\}$, and the set of edges, $E = \{e_1, e_2, \ldots, e_M\}$, make up a graph, $G(V, E)$. To make the graph G amenable to machine learning, each node is assigned a feature vector (information of the atomic number and the Cartesian coordinates of the atom in this work). The node data may also be processed to create edge features, e_{ij}, which encode positional information between the two nodes i and j, as shown in Figure 2b.

The GNNs used in the present work follow the message passing neural network (MPNN) paradigm, in which node features are updated from neighboring nodes in a message passing phase and then the updated features on a node are mapped to a property of interest in a readout phase,[50−53] as shown in Figure 2c. The objective of the message passing phase is to learn an embedding for each node, h, such that unique structural fingerprints for the node are encoded. The message passing typically occurs over a certain number of iterations. During iteration t, pairwise interactions between atom i and neighboring atoms j are summed and processed to produce a message

$$m_i^t = \sum_{j \in N_i} M_t(h_i^t, h_j^t, e_{ij}) \tag{8}$$

where N_i is the neighborhood of all atoms surrounding atom i within a distance cutoff r_{cut}, M_t is a learnable function that takes as input the embeddings h_i^t and h_j^t of atoms i and j as well as their edge data e_{ij}. The embedding of atom i is then updated using the message from eq 8:

$$h_i^{t+1} = f_t(h_i^t, m_i^t) \tag{9}$$

where f_t is a learnable function. Once the embeddings have been satisfactorily learned, the embedding h_i of atom i in the last iteration is then passed to a readout function to produce the property of interest for this atom (NMR tensor of Si atoms in this work).

Invariant Graph Neural Networks. We explored rotation-invariant GNNs as example ML models designed for scalar properties. Specifically, the DGL (deep graph library) implementation[54] of the DimeNet++[55,56] invariant GNN is selected (we note that DimeNet++ is a strong baseline model in terms of accuracy). The model is customized to allow for the prediction of node properties. A hyperparameter grid sweep was performed to optimize the DimeNet++ model to predict the shift tensor eigenvalues. A total of six separate models are created, each trained on a different shift tensor convention (eqs 1−5) and the standard convention eigenvalues, using the three parameters of the convention as the target. See the Supporting Information for further implementation details and the optimal hyperparameters.

The DimeNet++ model itself is limited to a rotation-invariant mapping from the input structure to the target chemical shift tensor parameters. While the basis functions used in eqs 8 and 9 themselves are rotation-equivariant, the message passing framework is rotation-invariant. Specifically, the coordinate information on each node is only used to initialize distances and angles, which are invariant geometric

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 2. Mean Absolute Error (MAE) for Individual Eigenvalues and Their Averaged Total Prediction Error for Invariant-Target Models[a]

			MAE/ppm			
model	embedding symmetry	target symmetry	total	δ_{11}	δ_{22}	δ_{33}
LRR	invariant	invariant	7.66	8.89	5.86	8.24
DimeNet++	invariant	invariant	6.44	6.45	7.18	5.70
GNN	equivariant	invariant	7.82	9.50	5.23	8.73
GNN	equivariant	equivariant	3.05	3.08	2.84	3.22

[a]Models are categorized by the symmetry of the embedding functions used in training (training symmetry) and the symmetry of the target at the time the loss was calculated (target symmetry). It should be noted that the last row was not obtained by directly fitting the eigenvalues but rather by fitting the full tensor and then computing the eigenvalues.

properties. As a result, DimeNet++ and similar frameworks are limited to predictions of scalar targets, in our case multiple uncorrelated scalar values. The target shift tensor parameters, henceforth called *scalar NMR parameters*, fall under the assumption that the tensor parameters are independent of crystal orientation, as the powder pattern may be used to obtain the scalar NMR parameters. Therefore, the parameters are invariant to rotations.

Equivariant Graph Neural Networks. A rotation-equivariant model was created using the Tensor Field Network (TFN)[41] and e3nn[57] frameworks, as implemented in the MatTEN[58] package. In addition to the scalar NMR shift tensor parameters discussed above, this model can directly predict the full chemical shift tensors. An initial equivariant GNN model was implemented (details are given in the Supporting Information) to determine the optimal target on which to train (symmetric vs asymmetric and spherical vs Cartesian tensors). A symmetric spherical tensor target was found to yield the best loss, and a hyperparameter grid search was performed to optimize the equivariant GNN model for the symmetric spherical tensor. To yield a useful model, the symmetric spherical tensor is then converted to a Cartesian tensor, which may be processed as a shift tensor.

An additional rotation-invariant GNN model was similarly created, but this model was trained on the shift tensor eigenvalues. Internally, this model still does message passing using equivariant embeddings (as explained in the following paragraph), but in this case the target is set to the (scalar) eigenvalues instead of the full chemical shift tensor. A hyperparameter grid search was performed to optimize the invariant GNN model (implementation details may be found in the Supporting Information).

In the TFN framework, similarly to DimeNet++, the embedding functions in eq 9 are rotation-equivariant. The TFN framework, however, differs in that the message update function contains convolution filters constrained to the form

$$W(r) = R(r)Y_l^m(\hat{r}) \tag{10}$$

where $R(r)$ is a learnable function of the distance between the two nodes and $Y_l^m(\hat{r})$ are spherical harmonics taking in the orientation between the nodes. The **W** matrix has the form of a block-diagonal matrix where the blocks correspond to the irreps selected for the network. Additionally, the embedding vectors used in the TFN framework contain blocks corresponding to the irreps. The **W** matrix along with the embedding vectors may then be convolved according to Clebsch−Gordan tensor products to ensure that the symmetry of each irrep is preserved. Thus, the message passing phase of the TFN model uses equivariant embeddings.

Benchmarking. To the best of our knowledge, no previous model has been proposed to predict full shift tensors. Thus, benchmarking will take place in two steps. The current state-of-the-art model for ^{29}Si scalar NMR parameter prediction was introduced by Chaker et al.,[35] who used linear ridge regression (LRR) over the smooth overlap of atomic positions (SOAP)[59] features to predict the ^{29}Si isotropic chemical shift. We reimplemented this approach using the SOAP features generated by DScribe[60] and the LRR in scikit-learn.[61] During the invariant-target benchmark, an LRR-SOAP model will be trained to predict the three eigenvalues of the chemical shift tensor, and all models will be compared on their predictions of the eigenvalues. An LRR-SOAP model will also be trained to predict chemical shift and compared to the models as a benchmark of NMR property prediction. It should be noted that the SOAP formalism has been adapted to be symmetry-equivariant, which may allow SOAP-based models to better predict tensor components.[62] However, we opted to look only at the invariant SOAP kernel in our benchmarking, as that was the original kernel used by Chaker et al.

Additionally, wherever possible, historic models will be added in the benchmark. These models are neglected in the ML literature for NMR modeling; however, they are widely used in the NMR community, and therefore, benchmarking on such models can provide valuable information. One of these models is the Si−O−Si angle-based ρ model by Engelhardt and Radeglia:[63]

$$\delta = a\langle\rho\rangle + b \tag{11}$$

where a and b are fitting parameters and ρ is a function of the Si−O−Si bond angle ($\Omega_{\mathrm{Si-O-Si}}$) that approximates the oxygen s character:

$$\langle\rho(\Omega_{\mathrm{Si-O-Si}})\rangle = \frac{1}{4}\sum_{i=1}^{4}\frac{\cos\Omega_{\mathrm{Si-O-Si}}^{(i)}}{\cos\Omega_{\mathrm{Si-O-Si}}^{(i)} - 1}$$

An ^{29}Si ζ model introduced by Grimmer et al.[64,65] and later improved by Jardón Álvarez et al.[66] is also considered. This model correlates the anisotropy to the difference between the nonbridging-oxygen Si−O bond length and the average bridging-oxygen Si−O bond length in Q^3 species:

$$\zeta = m(\langle d_{\mathrm{Si-BO}}\rangle - d_{\mathrm{Si-NBO}}) \tag{12}$$

RESULTS AND DISCUSSION

Symmetry-Invariant Learning. The first model types investigated are those trained on rotation-invariant scalar targets. The models are categorized according to the symmetries of the internal embeddings and the final target, labeled as "embedding symmetry" and "target symmetry",

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 3. Performance of DimeNet++ Trained on Different Tensor Conventions[a]

training convention	eigenvalues (δ_{11}, δ_{22}, δ_{33})/ppm	Haeberlen ($\zeta\eta$) convention (ζ/ppm, η)	Maryland ($\Omega\kappa$) convention (Ω/ppm, κ)	isotropic shift/ppm
Haeberlen ($\zeta\eta$)	(7.53, 6.24, 7.54)	(10.88, 0.24)	(11.56, 0.36)	4.43
Haeberlen $\Delta\delta$	(7.83, 5.36, 7.85)	(10.91, 0.21)	(12.21, 0.38)	4.07
Maryland ($\Omega\kappa$)	(6.58, 5.83, 6.10)	(12.51, 0.30)	(7.92, 0.38)	4.12
Axiality/Rhombicity	(5.58, 4.62, 6.39)	(9.71, 0.12)	(7.96, 0.17)	4.21
principal axis system	(6.45, 7.18, 5.70)	(11.02, 0.26)	(7.18, 0.32)	4.23

[a]For each model, the convention used to train is specified, along with the evaulation errors upon conversion to the Haeberlen ($\zeta\eta$) and Maryland conventions as well as to the eigenvalues and isotropoic shifts. Mean absolute error (MAE) values are reported.

respectively, in Table 2. Of the invariant target models, DimeNet++ performs the best, and LRR-SOAP and the equivariant GNN are on par with each other. However, if a fully equivariant GNN model (equivarant embedding and equivariant target) is used to predict the full shift tensor and then diagonalized to yield the eignevalues, significant improvement over the invariant target models can be achieved. For example, the total MAE is reduced to 3.05 ppm, which is less than half of that from the DimeNet++ model (6.44 ppm). Comparing the equivariant GNN models trained using invariant versus equivariant target symmetry, it is clear that the boost in performance is due to the additional constraint afforded by learning a second-rank tensor rather than three independent scalars. The fully equivariant model will be further discussed in the next section, and here we focus on the invariant models.

The eigenvalues of the shift tensor are not the only invariant targets to consider; one can train ML models to directly predict the NMR parameters in different conventions outlined in Table 1. For this purpose, we selected the DimeNet++ model based on its superior performance in eigenvalue prediction. Five DimeNet++ models were trained (one for each convention), and for ease of comparison, their predictions were converted to the eigenvalues in the standard convention and the two IUPAC-recommended conventions (i.e., the Haeberlen ($\zeta\eta$) convention and the Maryland ($\Omega\kappa$) convention) using the equations in Table 1. The results are listed in Table 3. We observe that there is no single optimal model, only a "best-in-class" per NMR parameter. For example, the Axiality/Rhombicity convention has the best overall performance but still underperforms on isotropic shift and Maryland Ω values. Furthermore, tensor conventions should be interconvertible. While this is typically true for experimental spectra, we find that it is not the case for ML models trained on individual NMR parameters. For example, the model trained in the Maryland convention performs well when predicting Maryland convention values, but when converted to Haeberlen ($\zeta\eta$), the model ranks the lowest.

Indeed, it is worthwhile to note that the three NMR parameters predicted are not independent scalars, and when fitted as such, information and internal symmetry constraints are lost. It is therefore not surprising that the use of rotation-invariant models (e.g., DimeNet++) provide inferior results compared to a fully equivariant model. As noted earlier, some conventions are ill-defined and discontinuous for certain values. Additionally, all tensor conventions based on the Cartesian tensor have an issue of explicitly defined axes that can cause confusion when parameters are predicted outside their range, which results in a change of the order of the eigenvalues. A full discussion of the numerical issues that arise when fitting with the tensor conventions can be found in the Supporting Information.

Symmetry-Equivariant Learning. We now turn our attention toward the rotation-equivariant GNN model, which was shown to significantly outperform the invariant models (see Table 2). Similar to the case of invariant models, there are a variety of output targets available for the equivariant model, among which there is not an *a priori* optimal choice. We focus on using an asymmetric Cartesian tensor, symmetric Cartesian tensor, asymmetric irreps (E, A_{ik}, S_{ik}), or symmetric irreps (E, S_{ik}) as the target (refer to eq 6). Additionally, the question of which loss function is suitable for learning chemical shift tensors has not, to our knowledge, been investigated for training of ML models. However, there has been substantial work by the MRI diffusion tensor community on optimal tensor metrics for diffusion tensors.[67−72] The l_n norms offer a good balance of optimizing the shape, magnitude, and orientation of a tensor, and in our case we adopted the l_1 norm as the loss function. We found there is a small benefit to learning on a symmetric tensor versus an asymmetric tensor before symmetrizing. Additionally, there is a minor decrease in both epoch time and the loss when training on an irreps tensor versus a Cartesian tensor. Thus, the optimal space was chosen to be a symmetric irreps tensor using an l_1 norm loss function, and all subsequent results are obtained from models trained using this optimal space.

The best-performing equivariant GNN model exhibits an MAE of 1.05 ppm over the entire chemical shift Cartesian tensor. However, we admit that an MAE calculated for all of the tensor components is challenging to interpret because an asymmetric second-rank tensor has six independent parameters that must be assessed in order to evaluate the performance without loss of information. Therefore, we compare the predicted and DFT-calculated isotropic chemical shift, X, Y, and Euler angles (α, β, γ) (see Chemical Shift Tensor Conventions for their definition). These parameters were chosen because they provide an intuitive view of the magnitude, shape, and orientation of the tensor, as described above. Additionally, because the shift tensor is very closely linked to the structural point group,[73] the results are grouped by Q^n into three clusters reflecting the broad symmetry point group: T_d, C_3, and C_2. The results are summarized in Table 4.

The tetrahedral (T_d) sites exhibit low error in the prediction of the magnitude and shape of the tensor, and the tensors are well-predicted by the equivariant GNN model, as shown in Figure 3. While the errors for the Euler angles appear severe in the T_d case, the context of this prediction should be kept in mind. The tensor for a T_d site is highly spherical, showing little to no anisotropy. The axes of a sphere are not unique in these cases, and one would expect a random distribution of the Euler angles. Close inspection of the outlier Euler angles reveals that the high error predictions correspond to sites which are very nearly spherical (small X and Y). Even in these cases where the orientation of the tensor is less meaningful, the model still

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 4. Performance (Reported as Mean Absolute Error) of the Equivariant GNN Model for Each of the Relevant Six Tensor Parameters δ^{iso}, X, Y, α, β, and γ, Organized by Q^n

Q^n	δ^{iso}/ppm	X/ppm	Y/ppm	α/deg	β/deg	γ/deg
Q^4	1.52	0.52	0.60	60.6	12.2	42.4
Q^3	3.28	0.80	0.66	47.9	17.9	59.2
Q^2	6.95	1.71	1.80	39.2	9.0	27.4
Q^1	8.58	0.85	1.00	1.2	1.3	0.7
Q^0	1.46	1.27	1.10	49.4	39.9	107
total	2.82	0.85	0.85	53.6	13.9	45.1

provides an accurate prediction of the magnitude and shape of the tensor.

Moving to the lower-symmetry point group, the C_3-symmetric Q^1 and Q^3 sites are responsible for a cylindrically symmetric tensor. The equivariant GNN model shows low error in the anisotropy prediction, as expected due to the strong correlation to the C_3 axis in these sites, as shown in Figure 4. The isotropic shift, however, performs considerably worse than in the tetrahedral case, as is clearly seen in Table 4 (3.28 and 8.58 ppm in Q^3 and Q^1, respectively, vs 1.52 and 1.46 ppm in Q^4 and Q^0, respectively). Barring a cluster of three data points, the Q^3 isotropic shift tends to be overpredicted by the equivariant GNN. The sites with overpredicted isotropic shift also correspond to the sites with poorly predicted Euler angles in Figure 4d–f. We speculate that the relatively poor performance of the model is due to the lack of Q^3 sites in the training data combined with the significant increase in structural diversity compared to the Q^4 and Q^0 sites. Furthermore, the structures with anomalous Q^3 sites trend as outliers in the training set as well and are not well-sampled, which results in poor learning of their structural correlations. In some cases the formula of the material was not seen in the training set, nor were any similar formulas seen, which resulted in a poor extrapolation by the equivariant GNN model. In other cases, the same formula was seen but with minor structural variation, which led to an unfortunate case of the equivariant GNN memorizing the solution poorly.

The situation is made even more extreme in the Q^1 case due to the low number of samples. The Q^1 case is, however, fortunate in that despite the poor isotropic shift prediction, the remaining five parameters all show good correlation.

The C_2-symmetric sites exhibit the lowest-order symmetry group and represent a tensor with the shape of an asymmetric spheroid. These sites show the highest error for the isotropic shift, as shown in Figure 5, compared to the T_d and C_3 sites. This is likely a combination of fewer samples in the training data and also the high structural distortion, impairing learning of the isotropic shift (a parameter representing an average). However, the X and Y values still show that the equivariant GNN model is well-suited for predicting the shape of the tensor, as even in the C_2 case the tensor has rotational axes to which it can correlate. Similar to the C_3 case, the outliers observed in the C_2 case are likely due to the high degree of local structural variation possible. The outliers seen are all poorly sampled in the training set, resulting in the model learning these cases poorly, as was seen with the C_3 sites.

Overall, the equivariant GNN model is able to learn to predict the tensors of silicates, with the best performance being on sites with high symmetry, as summarized in Table 4. Even in cases where the model struggles with isotropic shift or tensor orientation, the shape of the tensor is well-predicted.

It is also instructive to benchmark our equivariant GNN model to historic models and previous state-of-the-art models to ensure that our model constitutes an advance in the field, especially for the domains where previous models were successful. For ^{29}Si NMR, the current state-of-the-art model is the LRR-SOAP model proposed by Chaker et al.[35] to predict isotropic shift.

We trained this model using our dataset to predict isotropic nuclear shifts and obtained an error of 5.87 ppm over the entire dataset (compared to 2.82 ppm with the equivariant GNN over our entire dataset), a Q^4 MAE of 4.78 ppm (compared to 1.52 ppm with the equivariant GNN), and a Q^3 MAE of 7.23 ppm (compared to 3.28 ppm with the equivariant GNN) for isotropic shifts. If we instead train the LRR-SOAP on the Q^4 and Q^3 chemical shifts separately rather than the entire set, the Q^4 trained model has an MAE of 4.88 ppm, and the Q^3 trained model has an MAE of 6.57 ppm, showing that the equivariant GNN improves significantly over state of the art. Additionally, it should be noted that one drawback of the SOAP descriptor is that its size scales with the number of species in the dataset, N, as $N(N-1)$. Thus, for our entire dataset the SOAP descriptor encoding has a size of 10 980, and while this is not an issue for LRR, the dimensionality may be an issue for other methods especially as the descriptor size is far greater than the dataset size. It should also be noted that the LRR model can only handle scalar values, whereas the

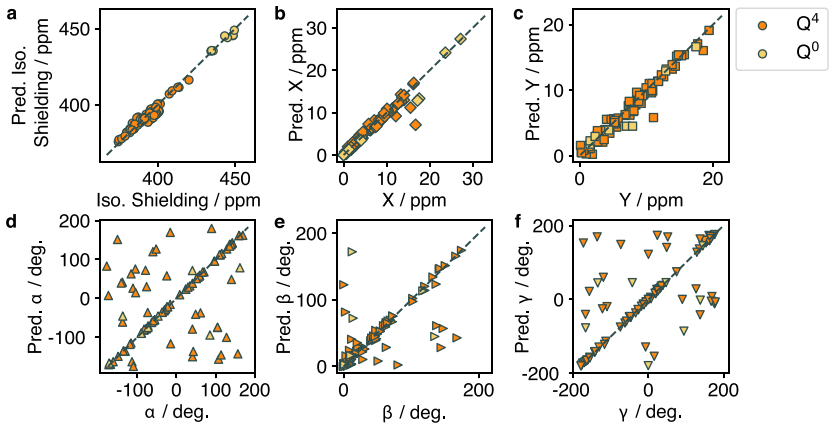


Figure 3. Predicted vs true values of (a) δ^{iso}, (b) X, (c) Y, (d) α, (e) β, and (f) γ for T_d-symmetric Q^4 and Q^0 sites.

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

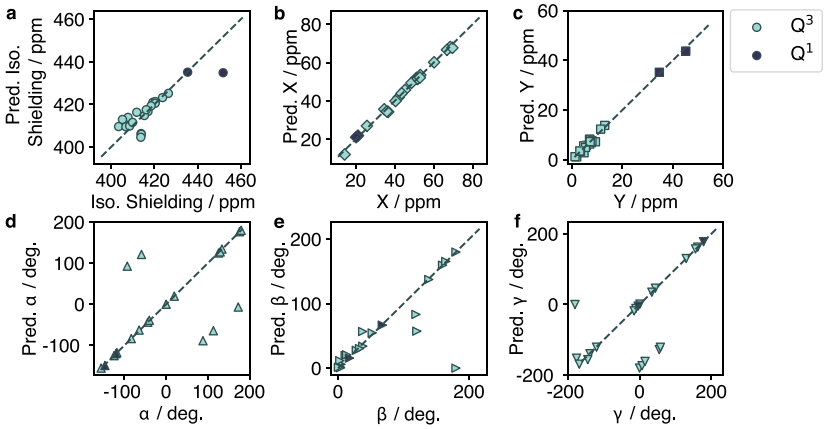


Figure 4. Predicted vs true values of (a) δ^{iso}, (b) X, (c) Y, (d) α, (e) β, and (f) γ for C_3-symmetric Q^3 and Q^1 sites.

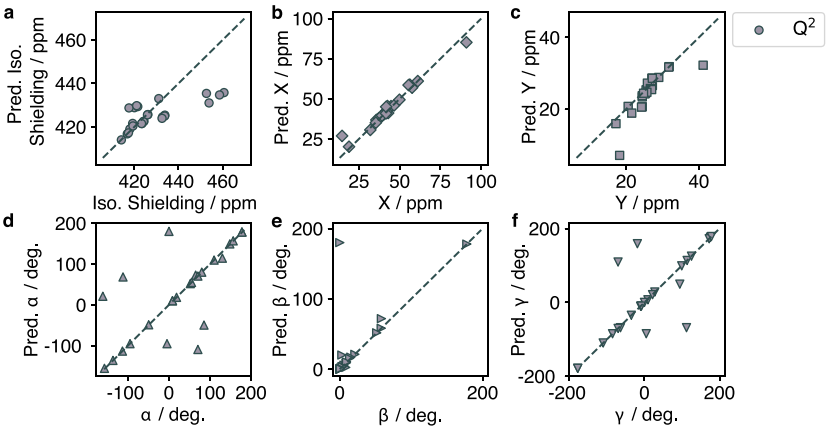


Figure 5. Predicted vs true values of (a) δ^{iso}, (b) X, (c) Y, (d) α, (e) β, and (f) γ for C_2-symmetric Q^2 sites.

main benefit of the equivariant GNN model is in providing the full shift tensor.

It is also important to consider historical analytical expressions, as we lose the descriptive power of analytical expressions when we select an ML method. Determining the coefficient and intercept from eq 11, we obtain $a = -186.3$ ppm and $b = 477.6$ ppm to yield a linear expression with an MAE of 3.60 ppm over the Q^4 subset. Performing the same analysis with eq 12, we obtain a coefficient of $m = -1009$ ppm Å^{-1} to yield a model with an MAE of 8.77 ppm over the Q^3 data subset. Compared to the equivariant GNN's performance with a Q^4 isotropic shift MAE of 1.52 ppm and a Q^3 anisotropy MAE of 0.78 ppm, the benefits of the equivariant GNN outweigh the loss of a simple functional form.

CONCLUSIONS

Machine learning approaches are increasingly employed to predict a variety of physical properties, accelerating and expanding access to material data. However, many of those physical properties adhere to inherent constraints, such as symmetry relationships or limits. In the case of tensorial properties, each eigenvalue may be predicted as an independent scalar, but such treatment effectively ignores the underlying symmetry information on the tensor. In this work, we explore the performance of machine learning models that rely on symmetry-invariant fitting procedures and contrast the results with a symmetry-equivariant approach. We find that the NMR tensor parameters cannot be easily learned via symmetry-invariant processes and often contain algebraic structure that makes the learning process more difficult, independent of the tensor convention. By imposing symmetry equivariance, our equivariant GNN model is able to outperform by 53% the symmetry invariant models, demonstrating that handling the tensorial nature of the target is the key to accurately modeling the system.

Examining the results of the equivariant GNN model, we observe that the model is able to accurately predict the tensor, not just in terms of shape and magnitude, but in most cases in the orientation as well. Closer inspection of the cases where the orientation seems to fail shows that these are often the cases of highly spherically symmetric tensors where an orientation is not meaningful. Most surprisingly, the model is able to capture the shape (anisotropy and asymmetry) of the tensor, even in cases where the tensor exhibits very little anisotropy, for example for Q^4 and Q^0 sites. Despite the successes of the model, there are still cases where it fails; however, these failures are likely associated with a lack of data in the training set. Future work will be focused toward expanding the dataset, particularly for the Q^n species, which are less well sampled.

Notably, through the demonstrated work on silicates, it is feasible to predict the full NMR tensor with reasonable accuracy in seconds rather than hours to days as required for *ab initio* calculations. This opens a realm of possibilities from high-throughput screening of materials via comparison to

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

experimental NMR spectra to expediting NMR crystallography refinement procedures.

■ ASSOCIATED CONTENT

Data Availability Statement

Structures and NMR tensors used in the study are hosted on https://contribs.materialsproject.org/projects/lsdi_nmr. The MatTEN code repository is hosted at https://github.com/mjwen/MatTEN.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c07530.

Theoretical, implementation, and hyperparameter sweep details for DimeNet++ and equivariant GNN, hyperparameter sweep details for LRR-SOAP, and a discussion on tensor conventions in NMR (PDF)

Jupyter notebook and checkpoint file for a pretrained equivariant GNN model to demonstrate the usage of the model (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Kristin A. Persson − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States; Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0003-2495-5509; Email: kristinpersson@berkeley.edu

Authors

Maxwell C. Venetos − *Department of Materials Science and Engineering, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0003-3468-2006

Mingjian Wen − *Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas 77204, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c07530

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation under Grant DIBBS OAC 1640899. M.W. was also supported by startup funds from the Presidential Frontier Faculty Program at the University of Houston. This work made use of computational resources and software infrastructure provided through the Materials Project, which is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract DE-AC02-05-CH11231 (Materials Project Program KC23MP).

■ REFERENCES

(1) Harris, R. K.; Wasylishen, R. E.; Duer, M. J. *NMR Crystallography*; Wiley: San Francisco, CA, 2009.

(2) Spearing, D. R.; Farnan, I.; Stebbins, J. F. Dynamics of the α-β phase transitions in quartz and cristobalite as observed by in-situ high temperature ^{29}Si and ^{17}O NMR. *Phys. Chem. Miner.* **1992**, *19*, 307−321.

(3) Kihara, K. An x-ray study of the temperature dependence of the quartz structure. *Eur. J. Mineral.* **1990**, *2*, 63−77.

(4) Chan, S.; Gladden, L.; Elliott, S. Magnetic resonance studies on neutron irradiated quartz and amorphous silica. *J. Non-Cryst. Solids* **1988**, *106*, 413−416.

(5) Wells, S. A.; Dove, M. T.; Tucker, M. G.; Trachenko, K. Real-space rigid-unit-mode analysis of dynamic disorder in quartz, cristobalite and amorphous silica. *J. Phys.: Condens. Matter* **2002**, *14*, 4645−4657.

(6) McMillan, P. F.; Akaogi, M.; Ohtani, E.; WIlliams, Q.; Nieman, R.; Sato, R. Cation Disorder in Garnets Along the $Mg_3Al_2Si_3O_{12}$ Join: An Infrared, Raman, and NMR Study. *Phys. Chem. Miner.* **1989**, *16*, 428−435.

(7) Phillips, J. C.; Kerner, R. Structure and function of window glass and Pyrex. *J. Chem. Phys.* **2008**, *128*, 174506.

(8) Prasad, S.; Clark, T. M.; Sefzik, T. H.; Kwak, H.-T.; Gan, Z. H.; Grandinetti, P. J. Solid-state Multinuclear Magnetic Resonance Investigation of Pyrex. *J. Non-Cryst. Solids* **2006**, *352*, 2834−2840.

(9) Hansen, M. R.; Jakobsen, H. J.; Skibsted, J. ^{29}Si Chemical Shift Anisotropies in Calcium Silicates from High-Field ^{29}Si MAS NMR Spectroscopy. *Inorg. Chem.* **2003**, *42*, 2368−2377.

(10) Brouwer, D. H.; Brouwer, C. C.; Mesa, S.; Semelhago, C. A.; Steckley, E. E.; Sun, M. P.; Mikolajewski, J. G.; Baerlocher, C. Solid-state ^{29}Si NMR spectra of pure silica zeolites for the International Zeolite Association Database of Zeolite Structures. *Microporous Mesoporous Mater.* **2020**, *297*, 110000.

(11) Pedone, A.; Pavone, M.; Menziani, M. C.; Barone, V. Accurate First-Principle Prediction of ^{29}Si and ^{17}O NMR Parameters in SiO_2 Polymorphs: The Cases of Zeolites Sigma-2 and Ferrierite. *J. Chem. Theory Comput.* **2008**, *4*, 2130−2140.

(12) Brouwer, D. H.; Van Huizen, J. NMR crystallography of zeolites: How far can we go without diffraction data? *Magn. Reson. Chem.* **2019**, *57*, 167−175.

(13) Brouwer, D.; Horvath, M. A simulated annealing approach for solving zeolite crystal structures from two-dimensional NMR correlation spectra. *Solid State Nucl. Magn. Reson.* **2015**, *65*, 89−98.

(14) Brouwer, D. H.; Cadars, S.; Eckert, J.; Liu, Z.; Terasaki, O.; Chmelka, B. F. A General Protocol for Determining the Structures of Molecularly Ordered but Noncrystalline Silicate Frameworks. *J. Am. Chem. Soc.* **2013**, *135*, 5641−5655.

(15) Brouwer, D. H.; Langendoen, K. P. A graph theory approach to structure solution of network materials from two-dimensional solid-state NMR data. *CrystEngComm* **2013**, *15*, 8748−8762.

(16) Brouwer, D. H. Structure solution of network materials by solid-state NMR without knowledge of the crystallographic space group. *Solid State NMR* **2013**, *51*−*52*, 37−45.

(17) Cadars, S.; Brouwer, D. H.; Chmelka, B. F. Probing local structures of siliceous zeolite frameworks by solid-state NMR and first-principles calculations of ^{29}Si-O-^{29}Si scalar couplings. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1825−1837.

(18) Brouwer, D. H. NMR Crystallography of Zeolites: Refinement of an NMR-Solved Crystal Structure Using ab Initio Calculations of ^{29}Si Chemical Shift Tensors. *J. Am. Chem. Soc.* **2008**, *130*, 6306−6307.

(19) Brouwer, D. H.; Enright, G. D. Probing Local Structure in Zeolite Frameworks: Ultrahigh-Field NMR Measurements and Accurate First-Principles Calculations of Zeolite ^{29}Si Magnetic Shielding Tensors. *J. Am. Chem. Soc.* **2008**, *130*, 3095−3105.

(20) Luchinat, C.; Parigi, G.; Ravera, E.; Rinaldelli, M. NMR Solid-State Crystallography through Paramagnetic Restraints. *J. Am. Chem. Soc.* **2012**, *134*, 5006−5009.

(21) Salager, E.; Stein, R. S.; Pickard, C. J.; Elena, B.; Emsley, L. Powder NMR crystallography of thymol. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2610−2621.

(22) Venetos, M. C.; Dwaraknath, S.; Persson, K. A. Effective Local Geometry Descriptor for ^{29}Si NMR Q^4 Anisotropy. *J. Phys. Chem. C* **2021**, *125*, 19481−19488.

(23) Helgaker, T.; Jaszuński, M.; Ruud, K. Ab Initio Methods for the Calculation of NMR Shielding and Indirect Spin−Spin Coupling Constants. *Chem. Rev.* **1999**, *99*, 293−352. PMID: 11848983.

(24) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140*, 9844−9853. PMID: 30010335.
(25) Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, *5*, 83.
(26) Wen, M.; Tadmor, E. B. Hybrid neural network potential for multilayer graphene. *Phys. Rev. B* **2019**, *100*, 195419.
(27) Wen, M.; Tadmor, E. B. Uncertainty quantification in molecular simulations with dropout neural network potentials. *npj Comput. Mater.* **2020**, *6*, 124.
(28) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J.; et al. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater.* **2022**, *8*, 59.
(29) Jonas, E.; Kuhn, S. Rapid prediction of NMR spectral properties with quantified uncertainty. *J. Cheminf.* **2019**, *11*, 50.
(30) Ito, K.; Xu, X.; Kikuchi, J. Improved Prediction of Carbonless NMR Spectra by the Machine Learning of Theoretical and Fragment Descriptors for Environmental Mixture Analysis. *Anal. Chem.* **2021**, *93*, 6901−6906. PMID: 33929838.
(31) Kuhn, S.; Egert, B.; Neumann, S.; Steinbeck, C. Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. *BMC Bioinf.* **2008**, *9*, 400.
(32) Laatikainen, R.; et al. Comprehensive Strategy for Proton Chemical Shift Prediction: Linear Prediction with Nonlinear Corrections. *J. Chem. Inf. Model.* **2014**, *54*, 419−430. PMID: 24455975.
(33) Blinov, K. A.; Smurnyy, Y. D.; Churanova, T. S.; Elyashberg, M. E.; Williams, A. J. Development of a fast and accurate method of ^{13}C NMR chemical shift prediction. *Chemom. Intell. Lab. Syst.* **2009**, *97*, 91−97. Selected papers presented at the Sixth Winter Symposium on Chemometrics, Kazan, Russia, February 16−22, 2008.
(34) Bremser, W. Hose—a novel substructure code. *Anal. Chim. Acta* **1978**, *103*, 355−365.
(35) Chaker, Z.; Salanne, M.; Delaye, J.-M.; Charpentier, T. NMR shifts in aluminosilicate glasses via machine learning. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21709−21725.
(36) Cuny, J.; Xie, Y.; Pickard, C. J.; Hassanali, A. A. Ab Initio Quality NMR Parameters in Solid-State Materials Using a High-Dimensional Neural-Network Representation. *J. Chem. Theory Comput.* **2016**, *12*, 765−773. PMID: 26730889.
(37) Marsmann, H.; Uhlig, F. *Chemical Shifts and Coupling Constants for Silicon-29*; Springer: Berlin, 2008.
(38) Sun, H.; Dwaraknath, S.; Ling, H.; Qu, X.; Huck, P.; Persson, K.; Hayes, S. Enabling materials informatics for ^{29}Si solid-state NMR of crystalline materials. *npj Comput. Mater.* **2020**, *6*, 53.
(39) Cohen, T. S.; Welling, M. Group Equivariant Convolutional Networks. *arXiv (SSSS)*, June 3, 2016, 1602.07576, ver. 3. https://arxiv.org/abs/1602.07576 (accessed 2022-05-20).
(40) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. *arXiv (Computer Science.Machine Learning)*, May 9, 2021, 2102.09844, ver. 2. https://arxiv.org/abs/2102.09844 (accessed 2022-05-20)..
(41) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv (Computer Science.Machine Learning)*, May 18, 2018, 1802.08219, ver. 3. https://arxiv.org/abs/1802.08219 (accessed 2022-05-20)..
(42) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758−1775.
(43) Facelli, J. C. Chemical shift tensors: theory and application to molecular structural problems. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 176−201.
(44) Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*; Wiley: New York, 2001.
(45) Mulder, F. A. A.; Filatov, M. NMR chemical shift data and ab initio shielding calculations: emerging tools for protein structure determination. *Chem. Soc. Rev.* **2010**, *39*, 578−590.
(46) Harris, R. K.; Becker, E. D.; Cabral De Menezes, S. M.; Granger, P.; Hoffman, R. E.; Zilm, K. W. Further conventions for NMR shielding and chemical shifts, IUPAC recommendations 2008. *Solid State Nucl. Magn. Reson.* **2008**, *33*, 41−56.
(47) Kuprov, I.; Wagner-Rundell, N.; Hore, P. Bloch-Redfield-Wangsness theory engine implementation using symbolic processing software. *J. Magn. Reson.* **2007**, *184*, 196−206.
(48) Grandinetti, P. J.; Ash, J. T.; Trease, N. M. Symmetry Pathways in Solid-State NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 121−196.
(49) Srivastava, D. J.; Grandinetti, P. J. Statistical learning of NMR tensors from 2D isotropic/anisotropic correlation nuclear magnetic resonance spectra. *J. Chem. Phys.* **2020**, *153*, 134201.
(50) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *Proc. Mach. Learn. Res.* **2017**, *70*, 1263−1272.
(51) Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R. et al. Relational inductive biases, deep learning, and graph networks. *arXiv (Computer Science.Machine Learning)*, October 17, 2018, 1806.01261, ver. 3. https://arxiv.org/abs/1806.01261 (accessed 2022-05-20).
(52) Wen, M.; Blau, S. M.; Spotte-Smith, E. W. C.; Dwaraknath, S.; Persson, K. A. BonDNet: a graph neural network for the prediction of bond dissociation energies for charged molecules. *Chemical Science* **2021**, *12*, 1858−1868.
(53) Wen, M.; Blau, S. M.; Xie, X.; Dwaraknath, S.; Persson, K. Improving machine learning performance on small chemical reaction data with unsupervised contrastive pretraining. *Chemical Science* **2022**, *13*, 1446−1458.
(54) *xnouhz/DimeNet-dgl: A DGL Implementation of DimeNet and DimeNet++*. https://github.com/xnuohz/DimeNet-dgl.git (accessed 2021-01-15).
(55) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *CoRR*, March 6, 2020, 2003.03123, ver. 1. https://arxiv.org/abs/2003.03123 (accessed 2022-05-20).
(56) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *CoRR*, December 1, 2020, 2011.14115, ver. 2. https://arxiv.org/abs/2011.14115 (accessed 2022-05-20).
(57) Geiger, M.; et al. *e3nn/e3nn: 2022-04-13*, 2022. DOI: 10.5281/zenodo.6459381.
(58) Wen, M.; Venetos, M. C. *MatTEN: Equivariant Invariant Graph Enabled Neural Network*. https://github.com/mjwen/matten.git (accessed 2022-05-20).
(59) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
(60) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(61) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(62) Grisafi, A.; Wilkins, D. M.; Csányi, G.; Ceriotti, M. Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.
(63) Engelhardt, G.; Radeglia, R. A semi-empirical quantum-chemical rationalization of the correlation between SiOSi angles and ^{29}Si NMR chemical shifts of silica polymorphs and framework aluminosilicates (zeolites). *Chem. Phys. Lett.* **1984**, *108*, 271−274.
(64) Grimmer, A.-R.; Peter, R.; Fechner, E.; Molgedey, G. High Resolution ^{29}Si NMR in Solid Silicates: Correlations between Shielding Tensor and Si-O Bond Length. *Chem. Phys. Lett.* **1981**, *77*, 331−335.

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX

(65) Grimmer, A.-R. Correlation between Individual Si-O Bond Lengths and the Principal Values of the ^{29}Si Chemical-Shift Tensor in Solid Silicates. *Chem. Phys. Lett.* **1985**, *119*, 416−420.
(66) Jardón-Álvarez, D.; Sanders, K. J.; Phyo, P.; Baltisberger, J. H.; Grandinetti, P. J. Cluster formation of network-modifier cations in cesium silicate glasses. *J. Chem. Phys.* **2018**, *148*, 094502.
(67) Pasternak, O.; Sochen, N.; Basser, P. J. The effect of metric selection on the analysis of diffusion tensor MRI data. *NeuroImage* **2010**, *49*, 2190−2204.
(68) Peeters, T. H. J. M.; Rodrigues, P. R.; Vilanova, A.; ter Haar Romeny, B. M. Analysis of Distance/Similarity Measures for Diffusion Tensor Imaging. In *Visualization and Processing of Tensor Fields: Advances and Perspectives*; Laidlaw, D., Weickert, J., Eds.; Springer: Berlin, 2009; pp 113−136.
(69) Vargas-Cardona, H. D.; Orozco, Á. A.; Álvarez, A. M.; Álvarez, M. A. Tensor decomposition processes for interpolation of diffusion magnetic resonance imaging. *Expert Syst. Appl.* **2019**, *118*, 92−108.
(70) Pennec, X.; Fillard, P.; Ayache, N. A Riemannian Framework for Tensor Computing. *Int. J. Comput. Vis.* **2006**, *66*, 41−66.
(71) Whitcher, B.; Wisco, J. J.; Hadjikhani, N.; Tuch, D. S. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn. Reson. Med.* **2007**, *57*, 1065−1074.
(72) Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005*; Springer: Berlin, 2005; pp 115−122.
(73) Buckingham, A.; Malm, S. Asymmetry in the nuclear magnetic shielding tensor. *Mol. Phys.* **1971**, *22*, 1127−1130.

https://doi.org/10.1021/acs.jpca.2c07530
J. Phys. Chem. A XXXX, XXX, XXX−XXX