


Deep neural network for the dielectric response of insulators

Linfeng Zhang *Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA*

Mohan Chen

*CAPT, HEDPS, College of Engineering, Peking University, Beijing 100871, China*Xifan Wu *Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, USA*Han Wang **Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Huayuan Road 6, Beijing 100088, China*

Weinan E

*Department of Mathematics and Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA
and Beijing Institute of Big Data Research, Beijing, 100871, China*Roberto Car †*Department of Chemistry, Department of Physics, Program in Applied and Computational Mathematics, Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, New Jersey 08544, USA*

(Received 2 July 2019; revised 9 June 2020; accepted 12 June 2020; published 22 July 2020)

We introduce a deep neural network to model in a symmetry preserving way the environmental dependence of the centers of the electronic charge. The model learns from *ab initio* density functional theory, wherein the electronic centers are uniquely assigned by the maximally localized Wannier functions. When combined with the deep potential model of the atomic potential energy surface, the scheme predicts the dielectric response of insulators for trajectories inaccessible to direct *ab initio* simulation. The scheme is nonperturbative and can capture the response of a mutating chemical environment. We demonstrate the approach by calculating the infrared spectra of liquid water at standard conditions, and of ice under extreme pressure, when it transforms from a molecular to an ionic crystal.

DOI: [10.1103/PhysRevB.102.041121](https://doi.org/10.1103/PhysRevB.102.041121)

Machine learning (ML) schemes introduced in the last decade [1–11] can model accurately the potential energy surface (PES) of a multiatomic system upon training with first-principles electronic density functional theory (DFT) data [12]. These approaches extend the size and time range of *ab initio* molecular dynamics (AIMD) [13,14], and make possible studies of rare events, such as crystal nucleation, with enhanced sampling methodologies in simulations of *ab initio* quality [15].

Most approaches so far have focused on representing the dependence of the PES, a scalar quantity, on the atomic coordinates. Recently, methods to fit the environmental dependence of electronic properties have been proposed [16–23]. In particular, kernel-based methods have been used to represent the polarization and its time derivatives [17,22,23], which are

needed in many studies of materials, including calculations of infrared (IR) [24], Raman [25,26], and sum frequency generation (SFG) [27] spectra, transport calculations in ionic liquids [28] and superionic crystals [29–31], and simulations of ferroelectric phase transformations [32,33]. These schemes learn from AIMD trajectories, but so far the calculated spectra of liquid water [23] do not match the quality of direct many-body expansions of the dipole moment [34].

Here, we propose an alternative approach based on deep neural networks (DNNs) and maximally localized Wannier functions (MLWFs) [35,36], i.e., electronic orbitals with minimal spatial spread obtained from a unitary transformation of the occupied orbitals in insulators. In spin saturated systems the MLWFs describe electron pairs. Thus, upon assigning charges of $2e^-$ to the Wannier centers (WCs), i.e., the centers of the MLWF distributions, the electric polarization is the dipole moment \mathbf{M} of the neutral system of point charges made by the WCs and the atomic nuclei. In extended periodic systems, \mathbf{M} includes an arbitrary quantum [37], but its

*wang_han@iapcm.ac.cn

†rcar@princeton.edu

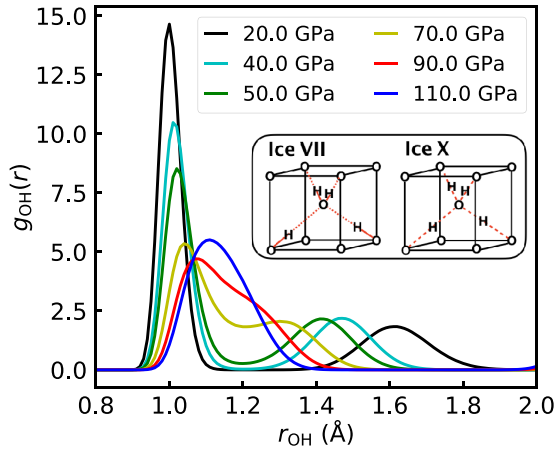


FIG. 1. Schematic illustration of the transitions from ice VII to ice VII' and from ice VII' to ice X by the O-H pair correlation functions $g_{OH}(r)$ calculated at $T = 300$ K and various pressures. Inset: Idealized structure of ice VII and ice X.

derivatives and correlation functions are well defined and describe observable properties.

In systems with different atomic species, the WCs fluctuate near the most electronegative atoms during molecular evolution. As a consequence of the nearsightedness of the electronic matter [38,39], the WCs only depend on the atoms in their local environment, and their positions can be accurately represented by the DNN model, called deep Wannier (DW), which is introduced here. It is an end-to-end scheme that does not use any *ad hoc* construction, in addition to the coordinate information and the network itself, to map input atomic coordinates into output WC coordinates. The model is size extensive, preserves translational, rotational, and permutational symmetry, and yields a polarization that varies continuously with the atomic coordinates. As such it can describe how the polarization responds to chemical bond changes in electronic insulators.

To demonstrate the methodology, we compute (a) the IR absorption spectrum of liquid water at standard temperature and pressure, and (b) the frequency-dependent imaginary dielectric function in the high-pressure transformation of ice VII to ice VII', and from ice VII' to ice X [40]. In ice VII there are two donor and two acceptor H bonds per oxygen, as stipulated by the ice rule [41,42], and the water molecules are always well defined. In ice VII' the hydrogens hop between two equivalent sites near each oxygen of an O-O bond, occasionally violating the ice rule. In ice X the hydrogens sit in the mid O-O bonds, the water molecules cannot be identified, and the system is better described in terms of O^{2-} and H^+ ions. This behavior is reflected in the O-H pair correlation functions in Fig. 1. For more sophisticated analyses and a phase diagram, see, e.g., Ref. [40]. The polarization change accompanying the above transformations is seamlessly described by DW. The scheme should also work at higher temperatures when ice X becomes superionic [43], or at even higher temperatures when the superionic crystal melts into an ionic liquid [28]. Similarly, the scheme should work for proton transfer events in water at standard conditions when neutral molecules interconvert with hydronium and hydroxide ion complexes (see, e.g., Ref. [44]).

In our implementation, DW is trained with valence-only pseudopotential electronic structure calculations. In the laboratory frame, the positions of the atoms and of the WCs are $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N$ and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i, \dots, \mathbf{w}_{N_w}$, respectively. We assume, for simplicity, that the WCs are only associated to one atomic species and consider water, in which there are four WCs per oxygen, as a concrete example [45]. It is easy to select with a cutoff distance the four WCs with coordinates $\mathbf{w}_{i=1, \dots, 4}$ that are closer to the oxygen i located at \mathbf{r}_i . Their centroid

$$\mathbf{w}_i = \frac{1}{4} \sum_{l=1}^4 \mathbf{w}_l \quad (1)$$

is well defined even when water molecules cannot be identified. Our aim is to construct a vector function \mathbf{f} that gives \mathbf{w}_i if the positions of the N_i atoms \mathbf{r}_k in the neighborhood \mathcal{N}_i of \mathbf{r}_i , defined by the radius r_c , are known, i.e.,

$$\mathbf{w}_i = \mathbf{f}(\{\mathbf{r}_k, k \in \mathcal{N}_i\}). \quad (2)$$

Here, i is an oxygen atom but the atoms k include oxygens and hydrogens. To ensure that \mathbf{f} preserves the translational, rotational, and permutational symmetry, we generalize the scheme for the deep potential (DP), the DNN representing the PES, introduced in Refs. [10,11].

First, we make a *local* frame transformation to the primed coordinates, which preserve translational symmetry,

$$\mathbf{r}'_{ki} \equiv \mathbf{r}_k - \mathbf{r}_i, \quad \mathbf{w}'_i \equiv \mathbf{w}_i - \mathbf{r}_i. \quad (3)$$

Then, we introduce a weight function $s(r'_{ki})$ equal to $1/r'_{ki}$ at a short distance $r'_{ki} = (\mathbf{r}'_{ki} \cdot \mathbf{r}'_{ki})^{1/2}$, and decaying smoothly to zero as r'_{ki} approaches r_c . Using $s(r'_{ki})$, we describe the atomic coordinates $\mathbf{r}'_{ki} = (x'_{ki}, y'_{ki}, z'_{ki})$ with the 4-vector $\mathbf{q}_{ki} = (q_{ki}^1, q_{ki}^2, q_{ki}^3, q_{ki}^4) = (s(r'_{ki}), s(r'_{ki})x'_{ki}, s(r'_{ki})y'_{ki}, s(r'_{ki})z'_{ki})$ to enforce continuous evolution when atoms enter/exit the neighborhood.

Next, we enforce permutational invariance and rotational covariance by introducing two DNNs, an *embedding* DNN and a *fitting* DNN. The number of hidden layers and outputs is refined in the training procedure. The embedding DNN is the matrix $\mathbf{E} = (E_{jk}) = (E_j(s(r'_{ki})))$ with M rows and N_i columns optimized by training, which maps the set $\{s(r'_{ki}), k \in \mathcal{N}_i\}$ onto M outputs.

The set of generalized coordinates $\{\mathbf{q}_{ki}\}$ in a neighborhood is represented by the matrix $\mathbf{Q} = (Q_{k\lambda}) = (q_{ki}^\lambda)$ with N_i rows and four columns. Multiplication of \mathbf{E} by \mathbf{Q} gives the matrix $\mathbf{T} = \mathbf{E}\mathbf{Q}$ with M rows and four columns, whose generic element is

$$T_{j\lambda} = \sum_{k=1}^{N_i} E_j(s(r'_{ki})) Q_{k\lambda}. \quad (4)$$

Let \mathbf{S} be the matrix formed by the first $M' (< M)$ rows of \mathbf{T} . Multiplication of \mathbf{T} by \mathbf{S}^T , the transpose of \mathbf{S} , gives the matrix \mathbf{D} of dimension $M \times M'$, called the *feature* matrix,

$$\mathbf{D} = \mathbf{T}\mathbf{S}^T. \quad (5)$$

\mathbf{D} is the argument of the fitting DNN, a row matrix $\mathbf{F}(\mathbf{D}) = \{F_j(\mathbf{D}), j = 1, \dots, M'\}$ that converts the atomic coordinate information encoded in \mathbf{D} onto M' outputs, which are mapped onto the centroid $\mathbf{w}'_i = (w'_i{}^1, w'_i{}^2, w'_i{}^3)$ upon multiplication

with with the last three columns of \mathbf{T} ,

$$w_i^{\lambda} = \sum_{j=1}^M F_j(\mathbf{D}) T_{j,\lambda+1}. \quad (6)$$

Finally, \mathbf{w}_i is retrieved from \mathbf{w}_i' using Eq. (3) and one obtains the desired representation of Eq. (2).

We notice that \mathbf{w}_i constructed in the way introduced above naturally preserves all the symmetry requirements. Translational symmetry is preserved by the adoption of a local frame and relative positions in Eq. (3). Permutational symmetry is preserved by the smooth sum over the neighboring atoms in Eq. (4). Finally, as shown in Eq. (4), the last three columns of \mathbf{T} ($\lambda = 2, 3, 4$) transform covariantly under rotation because (Q_{k2}, Q_{k3}, Q_{k4}) transforms as \mathbf{r}'_{ki} . Then it is straightforward to verify that the elements of \mathbf{D} , and hence $\mathbf{F}(\mathbf{D})$, are invariant under rotation. Therefore, w_i^{λ} in Eq. (6) transforms as $T_{j,\lambda+1}$, and is hence rotationally covariant. The values of M , M' , and of the number of layers of the DNNs are chosen empirically based on performance. In the applications discussed in this Rapid Communication we adopt $M = 100$ (of the same order of the number of atoms in a neighborhood), $M' = 6$, and use a three-layer representation for all the DNNs. The parameters $\boldsymbol{\gamma}$ of the embedding and fitting networks are determined by training, i.e., an optimization process that minimizes a loss function, here the mean-square difference between the DW prediction and the training data. The Adam stochastic gradient descent method [46] is adopted for the optimization. Generalization of this formalism to tensor properties (such as the polarizability) is introduced in Ref. [47].

DW should be combined with a DNN for the PES to study the evolution of the polarization along MD trajectories. For consistency, the two networks should be trained with electronic structure data at the same level of theory, as in the applications below, which used DW and the DP representation of the PES [10,11]. *Ab initio* electronic structure data are expensive and efficient learning strategies are crucial. We used the iterative learning scheme of Ref. [48]. In this approach, a DNN, initially trained with a limited pool of *ab initio* data, is used to explore inexpensively the configuration space. A small subset of the visited configurations is selected with a suitable error indicator and single-shot *ab initio* calculations are performed at these configurations. Training with the new data improves the model for further exploration and selection, followed by new data acquisition and learning. The protocol is repeated until all the explored configurations are described with satisfactory accuracy. The error indicator exploits the highly nonlinear dependence of the DNN models on the network parameters. As a consequence, different initializations of the parameters lead to different local minima in the landscape of the loss function, originating an ensemble of minimizing DNNs. The variance of the predictions within this ensemble is an intrinsic property of a DNN model and is often a reliable indicator of its accuracy [49,50]. In our experience, good DNN models constructed with the above procedure require significantly less *ab initio* data in the target thermodynamic range than learning approaches based on independent AIMD sampling data.

In the following we report calculations on liquid water at STP and on ice undergoing pressure-induced structural

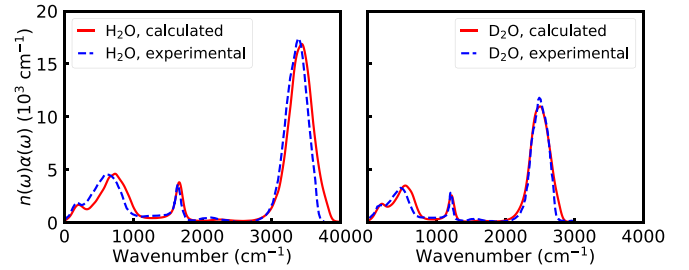


FIG. 2. IR absorption spectra of liquid H₂O (left) and D₂O (right). The continuous lines report DPMD+DW calculations at ~ 300 K. The dashed lines report experimental data for H₂O at 298 K [57] and for D₂O at 295 K [58].

phase transitions. For liquid water at STP, we did not use the incremental data generation scheme, since the training data were available and accessible online from previous work [51]. For high-pressure ice, electronic structure data were not available, and we constructed DP and DW from scratch using the incremental learning procedure outlined above [52]. The code for this work has been integrated into the open-source software package DEEPM-DW [53] and we used the DP-GEN package [54] for the iterative scheme.

We use DFT at the hybrid functional level [Perdew-Burke-Ernzerhof (PBE0) [55]] with dispersion corrections [56] for STP water. Using DW and DP [11] we calculate the IR absorption spectra of liquid H₂O and D₂O for a cell with 512 molecules under periodic boundary conditions. We use two microcanonical trajectories lasting 0.5 ns each, for H₂O and D₂O, at an average temperature of ~ 300 K at the equilibrium density of the simulation. The frequency-dependent absorption coefficient per unit length $\alpha(\omega)$ times the refractive index $n(\omega)$ is given by the Fourier transform of the time correlation function of the time derivative of the cell polarization $\dot{M} = +6e \sum_l \dot{\mathbf{r}}_{O_l} + e \sum_m \dot{\mathbf{r}}_{H_m} - 8e \sum_n \dot{\mathbf{w}}_n$ according to

$$\alpha(\omega)n(\omega) = \frac{2\pi\beta}{3cV} \int_{-\infty}^{+\infty} dt e^{-i\omega t} \langle \dot{M}(0) \cdot \dot{M}(t) \rangle, \quad (7)$$

where V is the volume, $\beta = 1/k_B T$ is the inverse temperature, and k_B is Boltzmann's constant. Figure 2 shows that the calculated spectra are in good agreement with the corresponding experimental observations.

Similarly accurate IR spectra of liquid water can be obtained from representations of the PES [59,60] and the dipole moment [34] based on many-body molecular expansions. These powerful approaches are limited to molecular liquids and crystals. By contrast, our nonperturbative method works also for nonmolecular systems, as we demonstrate by considering ice at $T = 300$ K in the pressure range from 20 to 110 GPa, wherein structural phase transitions from ice VII to ice VII' and to ice X occur. We adopted the PBE functional approximation of DFT as in Refs. [31,40]. We constructed DP/DW networks for ice in the temperature range from 240 to 330 K and pressure range from 20 to 120 GPa with the iterative learning approach set forth earlier in this Rapid Communication. This procedure required a total of 2248 single-shot DFT calculations with a 16 molecule cell and a total of 2400 single-shot DFT calculations with a 128 molecule cell.

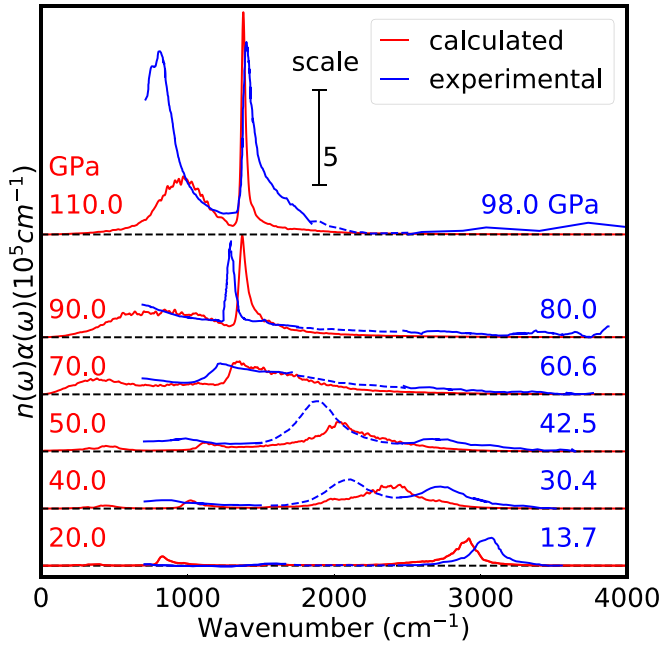


FIG. 3. Red: IR absorption spectra of H_2O ice at 300 K and pressures from 20 to 110 GPa (128 molecules, 0.5 ns trajectories). Blue: Experimental results converted from imaginary part of the dielectric function $\epsilon_2(\omega)$ of H_2O at 295 K and various pressures obtained from a Kramers-Kronig analysis of experimental reflectivity spectra [64], according to the relationship $n(\omega)\alpha(\omega) = 4\pi\omega\epsilon_2(\omega)$. Dashed lines denote the oscillator fit since there are no data in the range of strong diamond absorption (1800–2400 cm^{-1}). The curves are offset in the vertical direction for clarity.

For each cell size, the corresponding computational effort was less than the cost of a short (~ 5 ps) AIMD trajectory.

We sampled the ice configurations at 300 K in the pressure range 20–110 GPa with constant pressure deep potential molecular dynamics (DPMD) on a variable periodic cell with 128 water molecules, using a mild Nosé-Hoover thermostat [61,62] with damping time of 5 ps, much longer than the vibrational periods, to control the temperature. Using DP+DW, we calculated $\alpha(\omega)n(\omega)$ according to Eq. (7) with a set of 0.5-ns-long trajectories at various pressures. A direct AIMD study of the spectral changes in the closely related transformations from ice VIII (the proton ordered form of ice VII) to ice VII' and to ice X, reported in a pioneering paper by Bernasconi *et al.* [63], used a small 16 molecule cell and ~ 10 -ps-long trajectories.

Our results in Fig. 3 show a dramatic change with pressure of the product $\alpha(\omega)n(\omega)$. We also report the same quantity obtained from reflectance measurements of pressurized ice in a diamond anvil cell [64]. Experimental data are not available for $\omega \lesssim 800 \text{ cm}^{-1}$, and for $1800 \text{ cm}^{-1} < \omega < 2400 \text{ cm}^{-1}$. The theoretical curves are displayed at approximately 10 GPa higher pressure than the experimental curves to empirically correct for the missing nuclear quantum effects in the simulation and the inaccuracy of the adopted functional approximation.

Taking into account the limitations of theory and experiment, the experimental trend is reproduced nicely by the DP+DW simulations. At the lowest pressures, molecular

vibration features can be discerned, such as the H stretching band at about 3000 cm^{-1} . These modes soften dramatically and broaden with pressure, indicating a progressive weakening of the covalent O-H bonds. In the simulation, the spectral changes are quite abrupt at ~ 70 GPa, suggesting that this should be approximately the transition pressure to symmetric ice X. The same behavior is observed in experiment at ~ 60 GPa. By further increasing the pressure, two strong features emerge, characteristic of ice X, that harden and sharpen with pressure. The higher-frequency feature has dominant weight on H while the lower-frequency feature has more O weight. Interestingly, there is no close correspondence between the IR spectrum and the power spectrum of atomic dynamics reported in the Supplemental Material (SM) [52], suggesting that extreme anharmonicity affects the IR spectrum, as pointed out in Ref. [63].

Quantum effects in the dynamics were ignored in our calculations. These effects are small but non-negligible in liquid water and should be more pronounced near the ice VII to X transition in view of the relatively large isotope effect on the transition pressure, which is ~ 10 GPa higher for D_2O than for H_2O [64], suggesting that the transition is facilitated by proton tunneling. The calculation of dynamic quantum correlation functions is a major challenge for statistical simulations. Quantum IR correlations have been calculated recently for liquid water using approximate methods such as ring polymer [65] and centroid [65,66] MD, indicating that quantum corrections tend to redshift the classical spectral features. It would be extremely interesting to study how quantum corrections affect the dielectric properties in the ice VII, VII', and X transition sequence. Further studies of these issues will be facilitated by methods such as DPMD and DW, which improve significantly the statistical quality of *ab initio* simulations, as they are orders of magnitude faster than DFT methods and scale linearly with system size. Quantitative cost comparisons between direct AIMD and DP/DW simulations are reported in Figs. S2 and S3 in the SM [52].

In summary, DW is a useful tool to parametrize the dependence of the polarization on the atomic environment. The approach can be naturally extended [47] to the environmental dependence of the polarizability tensor $P_{\delta\sigma} = \frac{\partial M_\delta}{\partial E_\sigma}$ (M_δ and E_σ are Cartesian components of the polarization and of the electric field), allowing us to compute Raman [26,47] and sum frequency generation spectra [67,68]. Access to the concerted evolution of atomic coordinates and polarization in simulations of large systems over long timescales should also open the way to studies of ferroelectric phase transitions with MD simulations of *ab initio* quality, rather than relying on effective Hamiltonian models [32,69,70]. Finally, related symmetry preserving DNN schemes have been considered in Refs. [71–74] and we defer to future work a discussion of the mathematical and machine learning aspects of the DP/DW models.

The work of L.Z., X.W., W.E, and R.C. was conducted at the Center “Chemistry in Solution and at Interfaces” (CSI) funded by the DOE Award No. DE-SC0019394. The work of L.Z and W.E was partially supported by a gift from iFlytek to Princeton University and by ONR Grant No. N00014-13-1-0338. The work of H.W. was supported

by the NSFC under Grant No. 11871110, and the National Key Research and Development Program of China under Grants No. 2016YFB0201200 and No. 2016YFB0201203. We used resources of the National Energy Research Scientific

Computing Center (DOE Contract No. DE-AC02-05cH11231). We are also grateful for computing time at the Terascale Infrastructure for Groundbreaking Research in Science and Engineering (TIGRESS) of Princeton University.

-
- [1] J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- [2] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [3] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Phys. Rev. Lett.* **108**, 058301 (2012).
- [4] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *New J. Phys.* **15**, 095003 (2013).
- [5] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, *J. Phys. Chem. C* **121**, 511 (2016).
- [6] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, *Sci. Adv.* **3**, e1603015 (2017).
- [7] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengi, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Red Hook, NY, 2017), pp. 991–1001.
- [8] J. S. Smith, O. Isayev, and A. E. Roitberg, *Chem. Sci.* **8**, 3192 (2017).
- [9] J. Han, L. Zhang, R. Car, and Weinan E, *Commun. Comput. Phys.* **23**, 629 (2018).
- [10] L. Zhang, J. Han, H. Wang, R. Car, and Weinan E, *Phys. Rev. Lett.* **120**, 143001 (2018).
- [11] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and Weinan E, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Red Hook, NY, 2018), pp. 4441–4451.
- [12] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [13] R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [14] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge, U.K., 2009).
- [15] L. Bonati and M. Parrinello, *Phys. Rev. Lett.* **121**, 265701 (2018).
- [16] M. Gastegger, J. Behler, and P. Marquetand, *Chem. Sci.* **8**, 6924 (2017).
- [17] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, *Phys. Rev. Lett.* **120**, 036002 (2018).
- [18] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, *ACS Central Sci.* **5**, 57 (2018).
- [19] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, *Proc. Natl. Acad. Sci. USA* **116**, 3401 (2019).
- [20] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, *npj Comput. Mater.* **5**, 22 (2019).
- [21] L. Zepeda-Núñez, Y. Chen, J. Zhang, W. Jia, L. Zhang, and L. Lin, *arXiv:1912.00775*.
- [22] N. Raimbault, A. Grisafi, M. Ceriotti, and M. Rossi, *New J. Phys.* **21**, 105001 (2019).
- [23] V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, *J. Chem. Phys.* **152**, 124104 (2020).
- [24] M. Sharma, R. Resta, and R. Car, *Phys. Rev. Lett.* **95**, 187401 (2005).
- [25] A. Putrino, D. Sebastiani, and M. Parrinello, *J. Chem. Phys.* **113**, 7102 (2000).
- [26] Q. Wan, L. Spanu, G. A. Galli, and F. Gygi, *J. Chem. Theory Comput.* **9**, 4124 (2013).
- [27] Q. Wan and G. Galli, *Phys. Rev. Lett.* **115**, 246404 (2015).
- [28] V. Rozsa, D. Pan, F. Giberti, and G. Galli, *Proc. Natl. Acad. Sci. USA* **115**, 6952 (2018).
- [29] J. Sun, B. K. Clark, S. Torquato, and R. Car, *Nat. Commun.* **6**, 8156 (2015).
- [30] B. C. Wood and N. Marzari, *Phys. Rev. Lett.* **97**, 166401 (2006).
- [31] E. Schwegler, M. Sharma, F. Gygi, and G. Galli, *Proc. Natl. Acad. Sci. USA* **105**, 14779 (2008).
- [32] V. Srinivasan, R. Gebauer, R. Resta, and R. Car, in *Fundamental Physics of Ferroelectrics 2003*, edited by P. K. Davies and D. J. Singh, AIP Conf. Proc. Vol. 677 (AIP, Melville, NY, 2003), pp. 168–175.
- [33] A. Fluri, A. Marcolongo, V. Roddatis, A. Wokaun, D. Pergolesi, N. Marzari, and T. Lippert, *Adv. Sci.* **4**, 1700467 (2017).
- [34] H. Liu, Y. Wang, and J. M. Bowman, *J. Chem. Phys.* **142**, 194502 (2015).
- [35] N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997).
- [36] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [37] R. Resta, *Rev. Mod. Phys.* **66**, 899 (1994).
- [38] W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996).
- [39] E. Prodan and W. Kohn, *Proc. Natl. Acad. Sci. USA* **102**, 11635 (2005).
- [40] J.-A. Hernandez and R. Caracas, *J. Chem. Phys.* **148**, 214501 (2018).
- [41] J. D. Bernal and R. H. Fowler, *J. Chem. Phys.* **1**, 515 (1933).
- [42] L. Pauling, *J. Am. Chem. Soc.* **57**, 2680 (1935).
- [43] M. Millot, F. Coppari, J. R. Rygg, A. C. Barrios, S. Hamel, D. C. Swift, and J. H. Eggert, *Nature (London)* **569**, 251 (2019).
- [44] M. Chen, L. Zheng, B. Santra, H.-Y. Ko, R. A. DiStasio, Jr., M. L. Klein, R. Car, and X. Wu, *Nat. Chem.* **10**, 413 (2018).
- [45] Generalizations to more than one reference atom would be feasible to deal with situations in which the WC are shared among covalently bonded atoms of same electronegativity.
- [46] D. Kingma and J. Ba, in Proceedings of the International Conference on Learning Representations (ICLR), 2015 (unpublished).
- [47] G. M. Sommers, M. F. Calegari Andrade, L. Zhang, H. Wang, and R. Car, *Phys. Chem. Chem. Phys.* **22**, 10592 (2020).
- [48] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and Weinan E, *Phys. Rev. Mater.* **3**, 023804 (2019).
- [49] E. V. Podryabinkin and A. V. Shapeev, *Comput. Mater. Sci.* **140**, 171 (2017).

- [50] L. Zhang, H. Wang, and Weinan E, *J. Chem. Phys.* **148**, 124113 (2018).
- [51] H.-Y. Ko, L. Zhang, B. Santra, H. Wang, Weinan E, R. A. DiStasio, Jr., and R. Car, *Mol. Phys.* **117**, 3269 (2019).
- [52] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.102.041121> for full details on the implementation, training, and validation of the models (DP and DW).
- [53] H. Wang, L. Zhang, J. Han, and Weinan E, *Comput. Phys. Commun.* **228**, 178 (2018).
- [54] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and Weinan E, *Comput. Phys. Commun.* **253**, 107206 (2020).
- [55] C. Adamo and V. Barone, *J. Chem. Phys.* **110**, 6158 (1999).
- [56] A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.* **102**, 073005 (2009).
- [57] J. E. Bertie and Z. Lan, *Appl. Spectrosc.* **50**, 1047 (1996).
- [58] J. E. Bertie, M. K. Ahmed, and H. H. Eysel, *J. Phys. Chem.* **93**, 2210 (1989).
- [59] V. Babin, C. Leforestier, and F. Paesani, *J. Chem. Theory Comput.* **9**, 5395 (2013).
- [60] A. Shank, Y. Wang, A. Kaledin, B. J. Braams, and J. M. Bowman, *J. Chem. Phys.* **130**, 144314 (2009).
- [61] S. Nosé, *J. Chem. Phys.* **81**, 511 (1984).
- [62] W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- [63] M. Bernasconi, P. L. Silvestrelli, and M. Parrinello, *Phys. Rev. Lett.* **81**, 1235 (1998).
- [64] A. Goncharov, V. Struzhkin, M. Somayazulu, R. Hemley, and H. Mao, *Science* **273**, 218 (1996).
- [65] O. Marsalek and T. E. Markland, *J. Phys. Chem. Lett.* **8**, 1545 (2017).
- [66] G. R. Medders and F. Paesani, *J. Chem. Theory Comput.* **11**, 1145 (2015).
- [67] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, Oxford, U.K., 1999), pp. 591–592.
- [68] Y. Nagata and S. Mukamel, *J. Am. Chem. Soc.* **132**, 6434 (2010).
- [69] R. Resta and D. Vanderbilt, in *Physics of Ferroelectrics* (Springer, Berlin, 2007), pp. 31–68.
- [70] D. Vanderbilt and W. Zhong, *Ferroelectrics* **206**, 181 (1998).
- [71] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, in *Advances in Neural Information Processing Systems 30* (Ref. [7]), pp. 3391–3401.
- [72] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, [arXiv:1806.01261](https://arxiv.org/abs/1806.01261).
- [73] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, in *Proceedings of the European Conference on Computer Vision (ECCV)* (Springer, Berlin, 2018), pp. 52–68.
- [74] J. Han, Y. Li, L. Lin, J. Lu, J. Zhang, and L. Zhang, [arXiv:1912.01765](https://arxiv.org/abs/1912.01765).