

# DeePCG: Constructing coarse-grained models via deep neural networks

Cite as: J. Chem. Phys. **149**, 034101 (2018); <https://doi.org/10.1063/1.5027645>

Submitted: 05 March 2018 . Accepted: 27 June 2018 . Published Online: 16 July 2018

 Linfeng Zhang,  Jiequn Han,  Han Wang, Roberto Car, and Weinan E



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Perspective: Machine learning potentials for atomistic simulations](#)

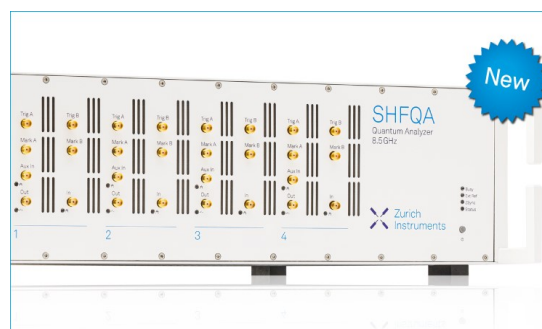
The Journal of Chemical Physics **145**, 170901 (2016); <https://doi.org/10.1063/1.4966192>

[SchNet – A deep learning architecture for molecules and materials](#)

The Journal of Chemical Physics **148**, 241722 (2018); <https://doi.org/10.1063/1.5019779>

[Reinforced dynamics for enhanced sampling in large atomic and molecular systems](#)

The Journal of Chemical Physics **148**, 124113 (2018); <https://doi.org/10.1063/1.5019675>



## Your Qubits. Measured.

Meet the next generation of quantum analyzers

- Readout for up to 64 qubits
- Operation at up to 8.5 GHz, mixer-calibration-free
- Signal optimization with minimal latency

Find out more



# DeePCG: Constructing coarse-grained models via deep neural networks

Linfeng Zhang,<sup>1</sup> Jiequn Han,<sup>1</sup> Han Wang,<sup>2,a)</sup> Roberto Car,<sup>3</sup> and Weinan E<sup>4,b)</sup>

<sup>1</sup>*Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA*

<sup>2</sup>*Institute of Applied Physics and Computational Mathematics, Fenghao East Road 2, Beijing 100094, People's Republic of China and CAEP Software Center for High Performance Numerical Simulation, Huayuan Road 6, Beijing 100088, People's Republic of China*

<sup>3</sup>*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*

<sup>4</sup>*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, People's Republic of China*

(Received 5 March 2018; accepted 27 June 2018; published online 16 July 2018)

We introduce a general framework for constructing coarse-grained potential models without *ad hoc* approximations such as limiting the potential to two- and/or three-body contributions. The scheme, called the Deep Coarse-Grained Potential (abbreviated DeePCG), exploits a carefully crafted neural network to construct a many-body coarse-grained potential. The network is trained with full atomistic data in a way that preserves the natural symmetries of the system. The resulting model is very accurate and can be used to sample the configurations of the coarse-grained variables in a much faster way than with the original atomistic model. As an application, we consider liquid water and use the oxygen coordinates as the coarse-grained variables, starting from a full atomistic simulation of this system at the *ab initio* molecular dynamics level. We find that the two-body, three-body, and higher-order oxygen correlation functions produced by the coarse-grained and full atomistic models agree very well with each other, illustrating the effectiveness of the DeePCG model on a rather challenging task. Published by AIP Publishing. <https://doi.org/10.1063/1.5027645>

## I. INTRODUCTION

In molecular dynamics (MD), we are often faced with two types of coarse-graining tasks. In a first set of applications, we are interested in evaluating the Landau free energy, which is a function of a small subset of coarse-grained (CG) variables. In this case, the CG variables are either scalar or low dimensional vector variables. In a second set of applications, we are interested in sampling with molecular dynamics (MD) or with Monte Carlo (MC) the configurations of an extensive set of CG variables. In this case, the dimensionality of the CG space is proportional to the size of the system but is reduced relative to the full space of atomistic coordinates. The first type of CG variables is typically adopted to study problems like phase transitions, where the objective is to perform detailed analyses of the Landau free energy surface by finding the metastable states, the free energy barriers between these states, the transition pathways, etc. Take the melting of a solid as an example; the Steinhardt order parameters<sup>1</sup> have been used as CG variables to differentiate solid (crystal) and liquid phases. The second type of CG variables is typically used to accelerate configurational sampling relative to full atomistic

simulations. For example, one may coarse-grain a polymer by replacing the monomers with point-like particles, or beads, connected by springs.

For a good description of the Landau free energy surface, one needs to find good order parameters acting as CG variables and address the issues associated with crossing high energy barriers. Typically these approaches are limited to a few CG variables, but recent work demonstrated that machine learning methods allow us to describe the functional dependence of the Landau free energy surface on several CG variables.<sup>2–8</sup>

When considering extensive CG variables, the difficulty is often associated with finding an accurate free energy function in the space of the CG variables. Such a free energy function usually depends on the CG variables in a complex and nonlinear way. Therefore, finding a good representation of this function often requires substantial physical/chemical intuition.<sup>9–25</sup> In principle, machine learning methods can address this problem more accurately and in an automated way,<sup>26–31</sup> but most machine learning approaches so far have focused on the representation of the potential energy surface in the space of the atomistic degrees of freedom rather than the representation of the free energy surface in the space of the CG variables. For example, the deep potential method,<sup>30</sup> a model based on deep neural networks, has made it possible to parameterize an atomistic potential energy function derived from

<sup>a)</sup>Electronic mail: wang\_han@iapcm.ac.cn

<sup>b)</sup>Electronic mail: weinan@math.princeton.edu

quantum mechanics without *ad hoc* approximations. A subsequent development of this approach, called Deep Potential Molecular Dynamics (DeePMD),<sup>31</sup> has allowed us to perform MD simulations of comparable quality to *ab initio* molecular dynamics (AIMD)<sup>32</sup> at the cost of classical empirical force fields.

The free energy surface, rather than the potential energy surface, is the key physical quantity that we need to represent when dealing with CG variables. In this work, we introduce the Deep Coarse-Grained Potential (DeePCG) scheme, an approach that generalizes the deep potential and DeePMD methods to representations of the free energy surface in the space of the CG variables, a quantity that will be called the CG potential in the following. A related method to represent the many-body character of the CG potential in molecules was recently reported in Ref. 33. In our approach, similar to the deep potential and DeePMD methods, no *ad hoc* approximations are required, in addition to the network model itself, to represent the CG potential. The scheme is very accurate as demonstrated by the almost perfect agreement of the many-body correlations extracted from CG simulations with the corresponding correlations extracted from the original atomistic model. In the present work, we use liquid water as an example to illustrate the approach. We choose AIMD as the underlying atomistic model and replace the individual water molecules with point-like particles located at the oxygen sites in the CG model. The excellent agreement of the second-, third-, and higher-order correlation functions between CG and atomistic models shows the promise of the DeePCG approach.

## II. METHODOLOGY

### A. Basic theory

We consider a  $d$ -dimensional system with  $N$  atoms in the constant-volume canonical ( $NVT$ ) ensemble. The coordinates of the atoms, in the laboratory frame, are  $\mathbf{q} = \{q_1, q_2, \dots, q_{dN}\} \in \mathbb{R}^{dN}$ . The configurational distribution function is defined by

$$p(\mathbf{q}) = \frac{1}{Z} e^{-\beta V(\mathbf{q})}, \quad (1)$$

where  $Z = \int e^{-\beta V(\mathbf{q})} d\mathbf{q}$  is the partition function.

The coarse-grained variables  $\boldsymbol{\xi}(\mathbf{q}) = \{\xi_1(\mathbf{q}), \xi_2(\mathbf{q}), \dots, \xi_M(\mathbf{q})\}$  are a reduced set of coordinates ( $M < dN$ ).  $M$  can be finite and independent of the system size or it can be extensive with the system size, as in the two cases discussed in the introduction. When  $M$  is finite, the CG variables are the so-called order parameters of the system. When  $M$  is extensive, the CG variables replace molecular objects with simpler sub-objects. The configurational distribution of the CG system is the projection of the configurational distribution of the microscopic (atomistic) system onto the space of the CG variables,

$$p(\boldsymbol{\xi}) = \frac{1}{Z} \int e^{-\beta V(\mathbf{q})} \delta(\boldsymbol{\xi}(\mathbf{q}) - \boldsymbol{\xi}) d\mathbf{q}. \quad (2)$$

The probability distribution in Eq. (2) allows us to define the CG potential and the forces acting on the CG variables as

$$U(\boldsymbol{\xi}) = -\frac{1}{\beta} \ln p(\boldsymbol{\xi}) \quad (3)$$

and

$$\mathbf{F}(\boldsymbol{\xi}) = -\nabla_{\boldsymbol{\xi}} U(\boldsymbol{\xi}), \quad (4)$$

respectively. Equation (3) tells us that a good CG potential should reproduce accurately the *full* configurational distribution of the CG variables in the atomistic model. Testing the quality of the full configurational distribution of the CG variables is difficult, and typically tests have been based only on two- and three-body correlation functions.<sup>15,16,19,20</sup>

$U(\boldsymbol{\xi})$  is uniquely specified by the underlying atomistic model and the definition of the CG degrees of freedoms. Therefore, constructing a CG model involves two steps: (1) the choice of an appropriate CG-potential representation and (2) the optimization of the parameters that define the potential representation. The way in which these two issues are addressed differentiates alternative schemes.

We notice that, even if we knew exactly  $U(\boldsymbol{\xi})$ , we would not have a closed deterministic form for the equation of motion of the CG variables due to the  $dN - M$  missing degrees of freedom in the CG potential. The issue of the dynamics of the CG variables has been addressed in the literature. See, e.g., Refs. 34 and 35. Further assumptions, like a time scale separation between the CG variables and the remaining degrees of freedom, are usually required to recover dynamical information of the atomistic model. In the following, we shall focus on the accurate construction of the CG potential. We will leave to future studies the investigation of CG dynamics.

### B. CG potential representation

We adopt a neural network representation  $U^w(\boldsymbol{\xi})$  for the CG potential  $U(\boldsymbol{\xi})$ . Here  $w$  are the parameters to be optimized by the training process.  $U^w(\boldsymbol{\xi})$  should be constructed using as input only the generalized coordinates  $\boldsymbol{\xi}$ , without any human intervention in the optimization process. The  $U^w(\boldsymbol{\xi})$  constructed in this way should preserve the symmetry properties of  $U(\boldsymbol{\xi})$ . In this manuscript, we limit ourselves to considering CG objects that behave as point particles and have only positional dependence. In this case, the  $\boldsymbol{\xi}$  variables are the coordinates of the CG particles. More general choices of the CG objects have been suggested in the literature<sup>36–39</sup> when dealing with, e.g., polymers, biological molecules, or colloidal particles. In these cases, it may be useful to consider rods, ellipsoids, particles connected by springs, etc., as the CG objects. In principle, all these cases could be treated with the present formalism. In the setup adopted here of point-like CG objects, the CG potential  $U^w$  is extensive, intrinsically many-body, and should preserve the translational and rotational invariance, as well as the permutational symmetry of the CG objects.

All the properties of the CG potential described above are preserved by the deep potential model.<sup>30</sup> To illustrate how it works, we use the example of liquid water. We write the CG potential as a sum of the local contributions of the CG particles, i.e.,  $U^w(\boldsymbol{\xi}) = \sum_i U_i^w(\boldsymbol{\xi})$ .  $U_i^w(\boldsymbol{\xi})$ , the potential contribution of the CG particle  $i$ , is constructed in two steps. First, the coordinates of the CG particle  $i$  and its neighbors within a cut-off

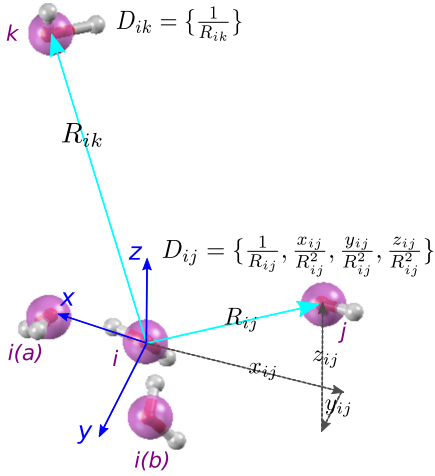


FIG. 1. Schematic plot of the neural network input for the environment of CG particle  $i$ , using water as an example. Red and white balls represent the oxygen and the hydrogen atoms of the microscopic system, respectively. Purple balls denote CG particles, which, in our example, are centered at the positions of the oxygens. We first sort all the CG particles within the cutoff radius  $R_c$  centered at  $i$ , according to their inverse distances from  $i$ . These particles constitute the neighbors of  $i$ .  $i(a)$  and  $i(b)$  are the first and the second nearest neighbor, respectively, of  $i$ .  $j$  and  $k$  are generic neighbors of  $i$ .  $i(a)$ , and  $i(b)$  define the local frame of  $i$ . In this local frame,  $i$  is the origin; the arrow from  $i$  to  $i(a)$  defines the  $x$  axis; the directional normal to the plane containing  $i$ ,  $i(a)$ , and  $i(b)$  defines the  $z$  axis; and the  $y$  axis is then assigned with the right-hand rule. Site  $j$  is close to  $i$  and is described with full radial and angular information by the descriptor  $\mathbf{D}_{ij} = \{1/R_{ij}, x_{ij}/R_{ij}^2, y_{ij}/R_{ij}^2, z_{ij}/R_{ij}^2\}$ , where  $(x_{ij}, y_{ij}, z_{ij})$  are Cartesian coordinates in the local frame of  $i$ . Site  $k$  is far from  $i$  and is described with radial information only by the descriptor  $\mathbf{D}_{ik} = \{1/R_{ik}\}$ .

radius  $R_c$  are transformed into the descriptors  $\{\mathbf{D}_{ij}\}$  of the local environment of the CG particle  $i$ . We call this procedure local frame transformation and refer to Fig. 1 for more details. In the following, we use the symbol  $\mathcal{D}_i$  to denote the entire set of descriptors for atom  $i$ . Next, as illustrated in Fig. 2, the

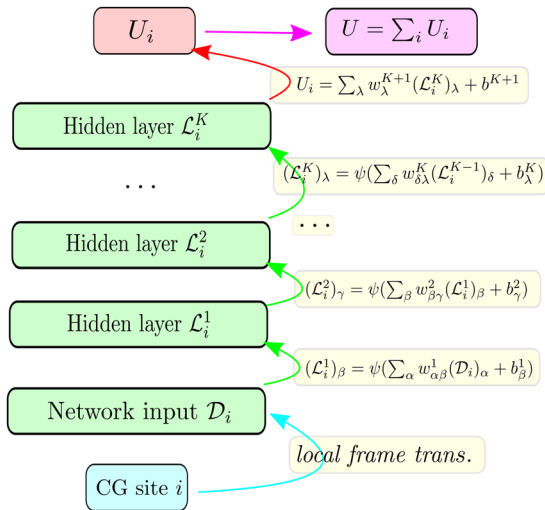


FIG. 2. Schematic plot of the sub-network structure for the CG particle  $i$ .  $\mathcal{D}_i$  (see definition in the text) is the input, and  $U_i$  is the output. In this sub-network, data flow from the input layer ( $\mathcal{D}_i$ ) to the output layer ( $U_i$ ) through  $K$  hidden layers, where each layer is a composition of a linear transformation and a piecewise nonlinear operation  $\psi(\dots)$ . We use the hyperbolic tangent for the nonlinear function  $\psi$ . This procedure is adopted for all the hidden layers. In the final step going from the last hidden layer to  $U_i$ , only the linear transformation is applied.

descriptors  $\mathcal{D}_i$  are given as input to a fully connected feedforward neural network to compute the potential contribution of the CG particle  $i$ . The mathematical formulation of the network structure is also presented in Fig. 2, where the operation of each layer of the network corresponds to a linear mapping of the output from the previous layer combined with a nonlinear mapping. The translational and rotational symmetries are preserved by the local frame transformation. The permutational symmetry is preserved because, (a) for each CG particle  $i$ , its descriptors  $\mathbf{D}_{ij}$  are sorted in ascending order according to the inverse distances between particles  $i$  and  $j$ ; (b) the subnetworks associated with the same type of particles share the same parameters  $\mathbf{w}$ ; and (c)  $U^w(\xi) = \sum_i U_i^w(\xi)$  is an additive relationship. More details on the deep potential method can be found in Refs. 30 and 31. Due to the adoption of a finite cutoff radius, the simulation cost of the DeePCG model scales linearly with the system size.

### C. CG potential optimization

The construction of the CG potential  $U^w(\xi)$ , introduced in Subsection II B, has many similarities with the construction of the potential energy  $V(\mathbf{q})$ , using the DeePMD method. There is, however, a very important difference in the two cases. In the DeePMD case, the potential energy  $V(\mathbf{q})$  is directly available from the underlying AIMD simulations. In the DeePCG case, the CG potential is a free energy and is not directly available. Therefore, the optimization for the CG potential requires a specific formulation. We adopt a force-matching scheme like the one in the multi-scale coarse-graining (MS-CG) method introduced by Voth *et al.*<sup>15</sup> In addition, we pay special attention to the fact that a neural network representation  $U^w(\xi)$  may have tens of thousands, or more, variational parameters, and a suitable optimization algorithm is needed.

A straightforward force-matching approach would consist in fitting accurate mean forces from atomistic simulations. There have been many efforts in this direction.<sup>40–42</sup> Of particular interest is a simple formula proposed by Ciccotti *et al.*,<sup>40</sup> in which a set of  $dN$ -dimensional vectors  $\mathbf{b}_j(\mathbf{q})$  that satisfy

$$\nabla_{\mathbf{q}} \xi_i(\mathbf{q}) \cdot \mathbf{b}_j(\mathbf{q}) = \delta_{ij}, \quad i, j = 1, \dots, M \quad (5)$$

is introduced. Then the mean force on  $\xi_i(\mathbf{q})$ , namely, the negative gradient of  $U(\xi)$  with respect to the position of the  $i$ th CG particle, can be expressed as

$$F_i(\xi) = -\partial_i U(\xi) = \langle \mathcal{F}_i(\mathbf{q}) \rangle_{\xi=\xi(\mathbf{q})}, \quad (6)$$

with an instantaneous force estimator

$$\mathcal{F}_i(\mathbf{q}) = -\mathbf{b}_i(\mathbf{q}) \cdot \nabla_{\mathbf{q}} V(\mathbf{q}) + \frac{1}{\beta} \nabla_{\mathbf{q}} \cdot \mathbf{b}_i(\mathbf{q}). \quad (7)$$

Here  $\langle \dots \rangle_{\xi=\xi(\mathbf{q})}$  denotes conditional expectation over the equilibrium distribution of the system restricted to the hypersurface  $\xi = \xi(\mathbf{q})$ . To train the DeePCG model, one needs to minimize the so-called loss function with respect to the model parameters  $\mathbf{w}$ . The most natural choice of loss function in terms of force-matching is



$$\hat{L}(\mathbf{w}) = \frac{1}{dDM} \sum_{n=1}^D \sum_{i=1}^{dM} |F_i(\xi_n) + \partial_i U^{\mathbf{w}}(\xi_n)|^2, \quad (8)$$

where  $D$  is the number of configurations of CG variables  $\xi_n$  in the dataset and the mean force  $F_i(\xi_n)$  is estimated with Eq. (6). We notice that the sample of CG configurations in Eq. (8) is very general in the sense that it does not need to be an equilibrium sample at the thermodynamic conditions of the atomistic simulation. For example, the sample could include, in an enhanced way, accessible CG configurations that have a small probability of occurrence at the thermodynamic conditions of interest, such as in the case of rare events. We stress, however, that the sampling of the microscopic degree of freedom orthogonal to  $\xi$  in Eq. (6) must be done at the appropriate thermodynamic conditions. In practice, the different configurations  $\xi_n$  in the dataset can be extracted from unconstrained MD or Monte Carlo (MC) simulations of the microscopic atomistic model at different temperatures.

The above straightforward approach is not convenient when the conditional expectation values in Eq. (6) require computationally expensive constrained/restrained simulations. In this situation, we find it more convenient to approximate the ensemble average  $\langle \cdots \rangle_q$  with the average  $(\frac{1}{D} \sum_{n=1}^D \cdots)$  over the configurations  $\xi_n$  [see Eq. (10) below]. The latter average does not require constrained/restrained simulations, but it requires  $\xi_n$  to be extracted from equilibrium atomistic simulations at the temperature selected in Eq. (1). Then the mean force  $F_i$  in the loss function (8) can be replaced by the instantaneous force  $\mathcal{F}_i$ . In other words, this corresponds to using an instantaneous version of the loss function,

$$\hat{L}^{\text{ins}}(\mathbf{w}) = \frac{1}{dDM} \sum_{n=1}^D \sum_{i=1}^{dM} |\mathcal{F}_i(\xi_n) + \partial_i U^{\mathbf{w}}(\xi_n)|^2. \quad (9)$$

With a sufficiently large representative dataset, we expect that the ensemble average of the difference between predicted and instantaneous forces should be approximated quite well by  $\hat{L}^{\text{ins}}(\mathbf{w})$ , i.e.,

$$L^{\text{ins}}(\mathbf{w}) := \frac{1}{dM} \sum_{i=1}^{dM} \langle |\mathcal{F}_i(\xi(\mathbf{q})) + \partial_i U^{\mathbf{w}}(\xi(\mathbf{q}))|^2 \rangle_q \approx \hat{L}^{\text{ins}}(\mathbf{w}). \quad (10)$$

This amounts to an ergodicity requirement for the atomistic system and is always valid if the system samples an equilibrium thermodynamic state.

By definition,  $\hat{L}^{\text{ins}}(\mathbf{w})$  is much easier to compute than  $\hat{L}(\mathbf{w})$ . Below we argue that  $\hat{L}^{\text{ins}}(\mathbf{w})$  is also a valid loss function to optimize the CG potential. To see this, note that the instantaneous force can be viewed as the mean force plus a random error  $R$ , which depends on the microscopic configuration  $\mathbf{q}$ , i.e.,

$$\mathcal{F}_i(\mathbf{q}) = F_i(\xi(\mathbf{q})) + R_i(\mathbf{q}). \quad (11)$$

By using Eq. (6), the average  $\langle R_i(\mathbf{q}) \rangle_{\xi=\xi(\mathbf{q})}$  in the constrained ensemble vanishes, so the average  $\langle R_i(\mathbf{q}) \rangle_q$  also vanishes. By inserting (11) into (10), the instantaneous loss function (10) becomes

$$L^{\text{ins}}(\mathbf{w}) = L(\mathbf{w}) + \frac{1}{dM} \sum_{i=1}^{dM} \langle R_i^2(\mathbf{q}) \rangle_q, \quad (12)$$

with

$$L(\mathbf{w}) := \frac{1}{dM} \sum_{i=1}^{dM} \langle |F_i(\xi(\mathbf{q})) + \partial_i U^{\mathbf{w}}(\xi(\mathbf{q}))|^2 \rangle_q. \quad (13)$$

Since the second term on the right-hand side of Eq. (12) is independent of  $\mathbf{w}$ ,  $L^{\text{ins}}(\mathbf{w})$  and  $L(\mathbf{w})$  have the same minimizer. This equivalence justifies our usage of  $\hat{L}^{\text{ins}}(\mathbf{w})$  as the loss function.

In the application example that we discuss in Sec. III, we use CG variables that depend linearly on the microscopic coordinates, similar to Ref. 15. However, the method that we have illustrated in Eqs. (6) and (7) can deal with nonlinear dependencies as well, as in the method discussed in Ref. 43.

In practice, we find that the stochastic gradient descent (SGD) method is very efficient to optimize the loss function (9), which is a highly non-convex function corresponding to a rugged landscape in the large parameter space due to the nonlinearity of the neural network interpolation. This ruggedness does not seem to constitute an essential difficulty since the different local minima found with the stochastic gradient descent (SGD) method approximate equally well the physics associated with the target function. We will discuss this issue in more detail later. Within our approach, the stochastic gradients  $\nabla_{\mathbf{w}} L(\mathbf{w})$  applied to update the parameters at each step are provided by the average over a small batch  $\mathcal{B}$ ,

$$l(\mathbf{w}) := \frac{1}{dM} \sum_{i=1}^{dM} \frac{1}{|\mathcal{B}|} \sum_{\alpha \in \mathcal{B}} |\mathcal{F}_i(\xi(\mathbf{q}_\alpha)) + \partial_i U^{\mathbf{w}}(\xi(\mathbf{q}_\alpha))|^2, \quad (14)$$

where  $\mathcal{B}$  is a subset of the whole dataset and  $|\mathcal{B}|$  denotes the batch size. The above procedure is different from the scheme adopted in Ref. 15, in which the full gradients  $\nabla_{\mathbf{w}} \hat{L}^{\text{ins}}(\mathbf{w})$  are applied to update the parameters at each step. We find that SGD greatly reduces the number of gradient evaluations that are required.

We note that other systematic procedures to optimize the parameters have been discussed in the literature. Of particular interest is the iterative Boltzmann inversion method,<sup>12</sup> which works by iteratively optimizing the CG interactions until the radial distribution functions (RDF) of the CG system match those of the target atomistic simulation. By construction, it provides accurate two-body correlations.

### III. COARSE-GRAINING OF LIQUID WATER

To show how we construct the CG potential for an extensive CG system, we use the coarse graining of a liquid water model from an *ab initio* density functional theory (DFT)<sup>44</sup> based simulation into effective “water particles” as an example. Because of its importance as a solvent in chemical and biological systems and its unique properties, the study of water is of wide interest. The DFT potential energy surface is intrinsically many-body. Developing an accurate CG model that represents a water molecule by a single particle is an ever-evolving and ongoing quest.<sup>9,12,15–20</sup>

Constructing effective interactions to achieve this goal has usually required a large amount of human effort combined

with substantial physical/chemical intuition. For example, in the mW monatomic potential,<sup>18</sup> which has been successfully used to study crystallization of water,<sup>45</sup> a specially designed Lennard-Jones-like form is used for two-body interactions, while three-body interactions are adapted from the Stillinger-Weber potential.<sup>46</sup> In principle, coarse graining approaches that do not require physical/chemical intuition are possible by exploiting general variational principles, such as the one adopted in the multi-scale coarse-graining (MS-CG)<sup>15</sup> or the iterative Boltzmann inversion<sup>12</sup> methods. However, even when using general variational principles to bridge atomistic and CG scales, the faithfulness of distribution of the CG variables may still depend on the CG representation. For example, in the applications of the MS-CG method to liquid water, the two- and three-body distribution functions of the CG variables still show non-negligible deviations from the corresponding target microscopic distributions.<sup>19,20</sup>

The DFT dataset in our example comes from Ref. 47. The electronic structure of the water system is modeled by DFT with the PBE0 exchange-correlation functional<sup>48</sup> and includes the long-range dispersion interactions self-consistently using the Tkatchenko-Scheffler model.<sup>49</sup> The corresponding AIMD simulation<sup>32</sup> adopts periodic boundary conditions and deuterons replace protons for a larger integration time step (0.5 fs). The simulation data consist of snapshots from a 20 ps-long trajectory in the NVT ensemble, where  $N = 192$  (64 H<sub>2</sub>O molecules),  $V = 1.9275 \text{ nm}^3$  (simple cubic periodic simulation cell), and  $T = 330 \text{ K}$ . In total 40 000 snapshots are recorded. The important difference between training DeePMD and DeePCG is that in DeePCG, one is attempting to estimate mean forces that correspond to conditioned averages of fluctuating atomic forces for fixed CG configurations, while in DeePMD one is attempting to estimate the deterministic atomic force in a fixed atomic configuration. Therefore, a short AIMD trajectory is not sufficient to train a DeePCG model with satisfactory accuracy, but this difficulty is circumvented by constructing a DeePMD model<sup>31</sup> from the AIMD data and sampling the configurations with a much longer DeePMD trajectory (15 ns). Figures 3–5 compare DeePMD and AIMD

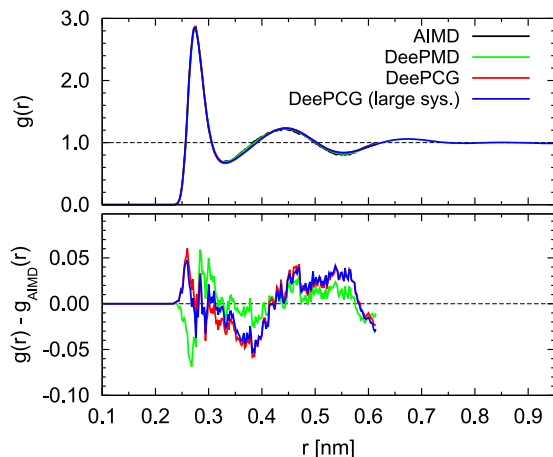


FIG. 3. Upper panel: the O—O RDFs of liquid water from AIMD and DeePMD for a system with 64 water molecules and from DeePCG simulation for systems with 64 and 512 CG particles; lower panel: the deviations of DeePMD and of two DeePCG models relative to the AIMD result.

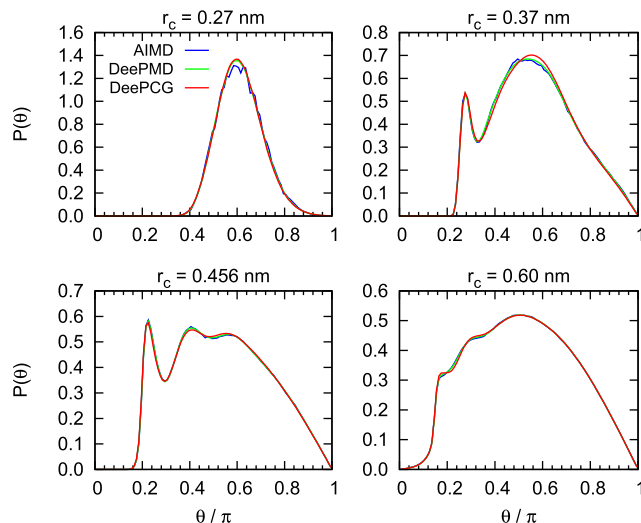


FIG. 4. The O—O—O ADFs of liquid water from AIMD, DeePMD, and DeePCG simulations. The results for four different cutoff radii are provided.

configurations in terms of the O—O radial distribution function (RDF), O—O—O angular distribution functions (ADFs), and the distributions of two averaged local Steinhardt parameters (defined in the Appendix),<sup>50</sup> respectively. It is observed that the configurations sampled by DeePMD are in almost perfect agreement with the AIMD data. Therefore, when considering the oxygen configurations, training with the data generated by DeePMD is essentially indistinguishable from that with data generated by AIMD.

Now we construct the DeePCG model. We use oxygen as the CG particle. We define the local environment of an O atom with the same cutoff radius adopted in the DeePMD model, i.e.,  $R_c = 6 \text{ Å}$ . We use the full radial and angular information for the 16 CG particles closest to the particle at the origin (see, e.g., particle  $j$  in Fig. 1), while retaining only radial information for all the other particles within  $R_c$  (see, e.g., particle  $k$  in Fig. 1). Next, the local environment of each CG particle defines a sub-network, and we use 4 hidden layers with decreasing

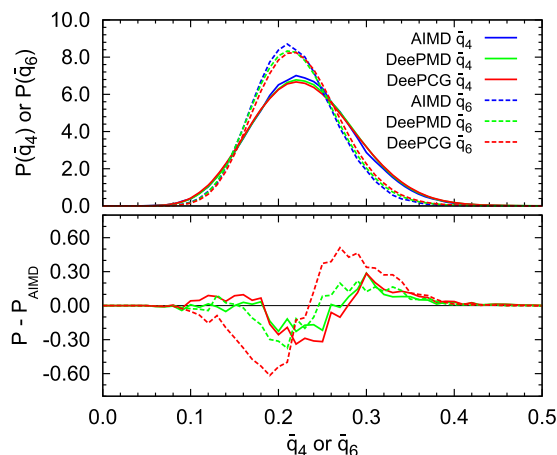


FIG. 5. Upper panel: the  $\bar{q}_4$  and  $\bar{q}_6$  distribution function of liquid water from AIMD, DeePMD, and DeePCG simulations; lower panel: deviations of DeePMD and of DeePCG from the AIMD results.

number of nodes per layer, i.e., 120, 60, 30, and 15 nodes from the innermost to the outermost layer, to construct the corresponding contribution to the CG potential.

The training process minimizes  $\hat{L}^{\text{ins}}(\mathbf{w})$  defined in Eq. (9). The force on each oxygen in the atomistic model serves as the instantaneous estimator  $\mathcal{F}_i$  in Eq. (7). We employ the stochastic gradient descent method with the Adam optimizer<sup>51</sup> to update the parameters of each layer, with a batch size of 4 and a learning rate that exponentially decays with the training step. In our current implementation, the training process requires 15 h on a ThinkPad p50 laptop computer with an Intel Core i7-6700HQ CPU and 32 GB memory. The DeePMD-kit<sup>52</sup> is used for optimizations and MD simulations of both the DeePMD and the DeePCG models.

After training, we perform an NVT simulation on the CG variables. The initial snapshot for this simulation is taken directly from a snapshot selected along the AIMD trajectory. The CG force is generated directly by the analytical gradient of the CG potential, the volume and the temperature are the same of the AIMD simulation, and the temperature is controlled using a Langevin thermostat with a damping time  $\tau = 0.1$  ps. In addition, using the same strategy, we perform an NVT simulation on 512 CG variables, where the only difference is that the number of CG variables and the size of the simulation region are 8 times larger than those of the AIMD simulation.

#### IV. DISCUSSION

In Figs. 3–5, we show that the DeePCG model reproduces very well the oxygen correlation functions of the atomistic DeePMD model and, by extension, of the underlying AIMD model. In addition to comparing 2- and 3-body correlations, as done in standard protocols,<sup>19,20</sup> we also perform tests on how well the DeePCG model preserves higher order distribution properties. In this regard, we calculate the sample averaged local Steinhardt bond order parameters  $\bar{q}_4$  and  $\bar{q}_6$  and find satisfactory agreement between the DeePCG and DeePMD models.

In the example that we discussed above, we use  $\hat{L}^{\text{ins}}(\mathbf{w})$  to optimize a CG model of water. We find that to base the optimization on  $\hat{L}(\mathbf{w})$  defined in Eq. (8) is significantly less efficient. This is because when the oxygens are the CG variables, very long constrained simulations using Eq. (6) are required to sample exhaustively the allowed configurations of the hydrogen bond network (HBN). Typically, when the oxygen positions are fixed, as in a constrained simulation, different HBN configurations are compatible with the fixed oxygen configurations, but it takes a very long time, typically of the order of a few nanoseconds, for the system to sample different HBN configurations. This is because of the long-range correlations imposed on the HBN by the Pauling ice rules (i.e., each oxygen has two nearer and two more distant hydrogen neighbors).<sup>53</sup> Thus, the scheme used here for matching the on-the-fly instantaneous forces is much more efficient.

It is well-known that neural network models are highly nonlinear functions of the parameters  $\mathbf{w}$ . Multiple local minima exist in the landscape of the loss functions  $L(\mathbf{w})$  or  $L^{\text{ins}}(\mathbf{w})$ .

Indeed, different initializations of the weights often lead to different local minimizers of the loss function. This, however, does not seem to be a serious problem as demonstrated by the test described below.

In this test, we prepare 1000 configurations randomly selected from the DeePMD data and pick up oxygen positions to define the CG particle configurations. For a CG particle  $i$  in each configuration, we define the model deviation  $\Sigma_i$  to be the standard deviation of the force on CG particle  $i$  predicted by CG models that only differ among themselves by the initialization of the simulation procedure, i.e.,

$$\Sigma_i = \sqrt{\langle \|\nabla_i U^{\mathbf{w}} - \langle \nabla_i U^{\mathbf{w}} \rangle_{\mathbf{w}}\|^2 \rangle_{\mathbf{w}}}, \quad (15)$$

where the ensemble average  $\langle \cdots \rangle_{\mathbf{w}}$  is taken with respect to models obtained from the same training process and the same training data set,<sup>56</sup> but different initialization of the parameters  $\mathbf{w}$ . In this way, 64 000 instances of the model deviation  $\Sigma_i$  are computed, and they are used to show the consistency of the predictions of different DeePCG models quantitatively. As shown by Fig. 6, with DeePMD data corresponding to 6 independent 2.5 ns-long trajectories, 99.3% of the model deviations, i.e., a large majority of them, are below 50 meV/Å. Moreover, the deviations do not become more significant when the magnitude of the CG force is large (inset in Fig. 6). Therefore, the differences of the CG forces predicted by different DeePCG models are generally consistent. Indeed the configurational distribution functions generated by DeePCG models that differ only in the initialization are indistinguishable. If we use shorter trajectories, the model deviations increase, as shown in Fig. 6 for DeePMD data corresponding to 2 independent 2.5 ns-long trajectories and for DeePMD data corresponding to a single 2 ns-long trajectory. This confirms that longer trajectories give better approximations of the ensemble average for  $L^{\text{ins}}(\mathbf{w})$ .

In terms of computational cost and scalability, in the current implementation, DeePCG accelerates DeePMD 7.5 times. Since all the physical quantities in DeePCG are sums of

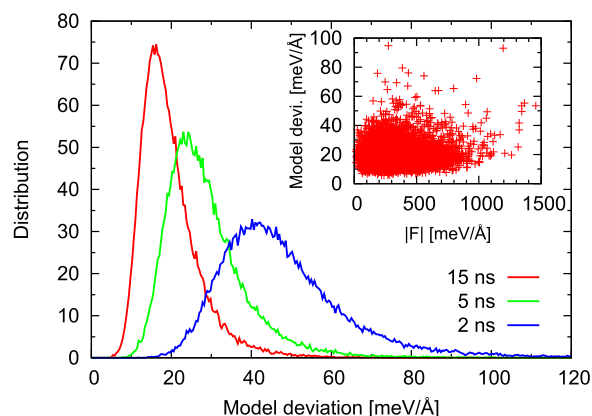


FIG. 6. The distributions of the model deviation  $\Sigma_i$  of the DeePCG model for liquid water, using training data from trajectories of total length 2 ns, 5 ns, and 15 ns. The 2 ns data are generated from a single trajectory; The 5 ns data are generated from 2 independent 2.5 ns-long trajectories; The 15 ns data are generated from 6 independent 2.5 ns-long trajectories. Inset: the correlation between the magnitude of the CG force and the model deviation, using 15 ns training data.

local contributions, upon training, the DeePCG model can be directly applied to much larger systems with linear scaling of cost. To test the reliability of DeePCG for larger systems, we perform a 1 ns-long NVT coarse-grained MD simulation on a system containing 512 water beads. This system is at the same temperature of the original DeePMD data, but is 8 times larger than the system used to construct the DeePCG model. The corresponding RDF, as shown in Fig. 3, is only very slightly less structured than the DeePCG result with 64 water beads, but tends to unity at large separation with a longer tail as we expect. This is consistent with the result in Ref. 54, which shows that the pair correlation function is almost converged in a 64-water fixed-cell system and larger cells only loosen the structure very slightly.

Finally, like in the case of the deep potential and DeePMD schemes,<sup>30,31</sup> in our implementation, discontinuities are present in the forces, due to adoption of a sharp cutoff radius, limitation of angular information to a fixed number of atoms, and abrupt changes in the atomic lists due to sorting. Upon training, these discontinuities become much smaller than the thermal fluctuations and can be subsumed in the thermal noise applied by the stochastic thermostat used to sample the canonical ensemble. The accuracy of the canonical distributions of the CG variables reported in Figs. 3–5, relative to the corresponding canonical distributions in the underlying atomistic simulation, validates our approach. While irrelevant for canonical sampling, the discontinuities make the CG potential a piece-wise continuous function of the CG coordinates, whereas in principle it should be a fully continuous function. We have recently generalized the present neural network representation in order to construct potentials that are fully continuous both in the space of the microscopic variables and in that of the CG degrees of freedom.<sup>55</sup> We leave to future work a discussion of applications using the fully continuous (smooth) version of our approach.

## V. CONCLUSION AND FUTURE WORK

In summary, DeePCG is a promising tool for parameterizing the CG potential and sampling CG configurations via MD. Due to the generality of the procedure adopted to construct the CG potential function, we expect DeePCG to be useful for a wide variety of tasks. In the case of water, we note that one reason for the great success of the mW potential is that it allows us to accelerate ice nucleation by several orders of magnitude because the absence of the hydrogen coordinates in the CG coordinate set eliminates the constraint imposed by the Pauling ice rules.<sup>18</sup> It would be interesting to investigate whether the CG water model introduced in this paper could describe not only the liquid but also the crystalline ice phase and whether the freezing temperature of the CG model could approximate closely that of the underlying microscopic model. Direct ice nucleation studies would be greatly facilitated by the CG model.

Coarse grained models are often used to describe the conformations of polymers, represented, for example, by a sequence of beads and springs. Until now, these models are typically constructed phenomenologically by requiring that a small set of force constants match experimental

and/or molecular simulation data. The DeePCG model presented here has the potential to completely eliminate phenomenological assumptions such as the restriction to harmonic spring interactions by systematically constructing a many-body potential for the beads depending on their configurations. We leave these studies and a more rigorous investigation of the dynamical properties of the CG models to future work.

## ACKNOWLEDGMENTS

The work of L. Zhang, J. Han, and W. E is supported in part by ONR Grant No. N00014-13-1-0338, DOE Grant Nos. DE-SC0008626 and DE-SC0009248, and NSFC Grant Nos. U1430237 and 91530322. The work of R. Car is supported in part by DOE-SciDAC Grant No. DE-SC0008626. The work of H. Wang is supported by the National Science Foundation of China under Grant Nos. 11501039 and 91530322, the National Key Research and Development Program of China under Grant Nos. 2016YFB0201200 and 2016YFB0201203, and the Science Challenge Project No. JCKY2016212A502.

## APPENDIX: DEFINITION OF THE LOCAL AVERAGED STEINHARDT PARAMETERS

The bond orientational order of particle  $i$  (atom or molecule) in a condensed environment is often described by a local Steinhardt parameter  $q_l(i)$ ,<sup>1</sup> defined as

$$q_l(i) = \left[ \frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} |q_{lm}(i)|^2 \right]^{1/2}, \quad (\text{A1})$$

with

$$q_{lm}(i) = \frac{\sum_{j \in N_b(i)} s(r_{ij}) Y_{lm}(\hat{\mathbf{r}}_{ij})}{\sum_{j \in N_b(i)} s(r_{ij})}. \quad (\text{A2})$$

Here  $N_b(i)$  denotes the set of neighbors of particle  $i$ ,  $Y_{lm}(\hat{\mathbf{r}}_{ij})$  are spherical harmonics, and  $s(r_{ij})$  is a switching function defined by

$$s(r) = \begin{cases} 1, & r < r_{\min}, \\ \frac{1}{2} + \frac{1}{2} \cos\left(\pi \frac{r - r_{\min}}{r_{\max} - r_{\min}}\right), & r_{\min} \leq r < r_{\max}, \\ 0, & r \geq r_{\max}. \end{cases} \quad (\text{A3})$$

In this work, we take  $r_{\min} = 0.31$  nm and  $r_{\max} = 0.36$  nm and adopt the modification of the local Steinhardt parameter proposed by Lechner and Dellago,<sup>50</sup> which is more sensitive than the original bond order parameter in distinguishing different crystal structures. The modified Steinhardt parameter is defined by

$$\bar{q}_l(i) = \left[ \frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} |\bar{q}_{lm}(i)|^2 \right]^{1/2}, \quad (\text{A4})$$

with

$$\bar{q}_{lm}(i) = \frac{\sum_{j \in \tilde{N}_b(i)} s(r_{ij}) q_{lm}(j)}{\sum_{j \in \tilde{N}_b(i)} s(r_{ij})}, \quad (\text{A5})$$

where  $\tilde{N}_b(i)$  includes  $N_b(i)$  and the tagged particle  $i$ . In the full expansion of the local averaged Steinhardt parameters,



4-body terms like  $Y_{lm}(\hat{r}_{ik}) \cdot Y_{lm}(\hat{r}_{jl})$ ,  $i \neq j \neq k \neq l$  are found. Therefore, the distribution of the value of the local averaged Steihardt parameters includes the effect of 4-body angular correlations.

- <sup>1</sup>P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Phys. Rev. B* **28**(2), 784 (1983).
- <sup>2</sup>T. Stecher, N. Bernstein, and G. Csányi, "Free energy surface reconstruction from umbrella samples using Gaussian process regression," *J. Chem. Theory Comput.* **10**(9), 4079 (2014).
- <sup>3</sup>L. Mones, N. Bernstein, and G. Csányi, "Exploration, sampling, and reconstruction of free energy surfaces with Gaussian process regression," *J. Chem. Theory Comput.* **12**, 5100–5110 (2016).
- <sup>4</sup>T. Lemke and C. Peter, "Neural network based prediction of conformational free energies—A new route toward coarse-grained simulation models," *J. Chem. Theory Comput.* **13**(12), 6213–6221 (2017).
- <sup>5</sup>R. Galvelis and Y. Sugita, "Neural network and nearest neighbor algorithms for enhancing sampling of molecular dynamics," *J. Chem. Theory Comput.* **13**(6), 2489–2500 (2017).
- <sup>6</sup>E. Schneider, L. Dai, R. Q. Topper, C. Drechsel-Grau, and M. E. Tuckerman, "Stochastic neural network approach for learning high-dimensional free energy surfaces," *Phys. Rev. Lett.* **119**(15), 150601 (2017).
- <sup>7</sup>L. Zhang, H. Wang, and W. E, "Reinforced dynamics for enhanced sampling in large atomic and molecular systems," *J. Chem. Phys.* **148**(12), 124113 (2018).
- <sup>8</sup>J. Zavdav, S. J. Marrink, and M. Praprotnik, "Multiscale simulation of protein hydration using the swinger dynamical clustering algorithm," *J. Chem. Theory Comput.* **14**(3), 1754–1761 (2018).
- <sup>9</sup>A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Phys. Rev. E* **52**(4), 3730 (1995).
- <sup>10</sup>R. E. Rudd and J. Q. Broughton, "Coarse-grained molecular dynamics and the atomic limit of finite elements," *Phys. Rev. B* **58**(10), R5893 (1998).
- <sup>11</sup>I. Pagonabarraga and D. Frenkel, "Dissipative particle dynamics for interacting systems," *J. Chem. Phys.* **115**(11), 5015–5026 (2001).
- <sup>12</sup>D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," *J. Comput. Chem.* **24**(13), 1624–1636 (2003).
- <sup>13</sup>S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, "Coarse grain models and the computer simulation of soft materials," *J. Phys.: Condens. Matter* **16**(15), R481 (2004).
- <sup>14</sup>W. Shinoda, R. DeVane, and M. L. Klein, "Coarse-grained molecular modeling of non-ionic surfactant self-assembly," *Soft Matter* **4**(12), 2454–2462 (2008).
- <sup>15</sup>W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *J. Chem. Phys.* **128**(24), 244114 (2008).
- <sup>16</sup>W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, "The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models," *J. Chem. Phys.* **128**(24), 244115 (2008).
- <sup>17</sup>M. Scott Shell, "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," *J. Chem. Phys.* **129**(14), 144108 (2008).
- <sup>18</sup>V. Molinero, E. B. Moore *et al.*, "Water modeled as an intermediate element between carbon and silicon," *J. Phys. Chem. B* **113**(13), 4008–4016 (2009).
- <sup>19</sup>L. Larini, L. Lu, and G. A. Voth, "The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials," *J. Chem. Phys.* **132**(16), 164107 (2010).
- <sup>20</sup>A. Das and H. C. Andersen, "The multiscale coarse-graining method. IX. A general method for construction of three body coarse-grained force fields," *J. Chem. Phys.* **136**(19), 194114 (2012).
- <sup>21</sup>M. Dinpajooh and M. G. Guenza, "On the density dependence of the integral equation coarse-graining effective potential," *J. Phys. Chem. B* **122**, 3426 (2017).
- <sup>22</sup>M. R. DeLyser and W. G. Noid, "Extending pressure-matching to inhomogeneous systems via local-density potentials," *J. Chem. Phys.* **147**(13), 134111 (2017).
- <sup>23</sup>J. W. Wagner, T. Dannenhoffer-Lafage, J. Jin, and G. A. Voth, "Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions," *J. Chem. Phys.* **147**(4), 044113 (2017).
- <sup>24</sup>T. Sanyal and M. Scott Shell, "Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation," *J. Chem. Phys.* **145**(3), 034109 (2016).
- <sup>25</sup>J. D. Moore, B. C. Barnes, S. Izvekov, M. Lísál, M. S. Sellers, D. E. Taylor, and J. K. Brennan, "A coarse-grain force field for RDX: Density dependent and energy conserving," *J. Chem. Phys.* **144**(10), 104501 (2016).
- <sup>26</sup>J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.* **98**(14), 146401 (2007).
- <sup>27</sup>A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.* **104**(13), 136403 (2010).
- <sup>28</sup>K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.* **8**, 13890 (2017).
- <sup>29</sup>S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," *Sci. Adv.* **3**(5), e1603015 (2017).
- <sup>30</sup>J. Han, L. Zhang, R. Car, and W. E, "Deep potential: A general representation of a many-body potential energy surface," *Commun. Comput. Phys.* **23**(3), 629–639 (2018).
- <sup>31</sup>L. Zhang, J. Han, H. Wang, R. Car, and W. E, "Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics," *Phys. Rev. Lett.* **120**, 143001 (2018).
- <sup>32</sup>R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Phys. Rev. Lett.* **55**(22), 2471 (1985).
- <sup>33</sup>S. T. John and G. Csányi, "Many-body coarse-grained interactions using Gaussian approximation potentials," *J. Phys. Chem. B* **121**(48), 10934–10949 (2017).
- <sup>34</sup>C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni, "Mori–Zwanzig formalism as a practical computational tool," *Faraday Discuss.* **144**, 301–322 (2010).
- <sup>35</sup>J. Lu and E. Vanden-Eijnden, "Exact dynamical coarse-graining without time-scale separation," *J. Chem. Phys.* **141**(4), 044109 (2014).
- <sup>36</sup>G. A. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, 2008).
- <sup>37</sup>P. T. Underhill and P. S. Doyle, "On the coarse-graining of polymers into bead-spring chains," *J. Non-Newtonian Fluid Mech.* **122**(1), 3–31 (2004).
- <sup>38</sup>I. Goodchild, L. Collier, S. L. Millar, I. Prokeš, J. C. D. Lord, C. P. Butts, J. Bowers, J. R. P. Webster, and R. K. Heenan, "Structural studies of the phase, aggregation and surface behaviour of 1-alkyl-3-methylimidazolium halide + water mixtures," *J. Colloid Interface Sci.* **307**(2), 455–468 (2007).
- <sup>39</sup>B. L. Bhargava and M. L. Klein, "Formation of micelles in aqueous solutions of a room temperature ionic liquid: A study using coarse grained molecular dynamics," *Mol. Phys.* **107**(4–6), 393–401 (2009).
- <sup>40</sup>G. Cicotti, R. Kapral, and E. Vanden-Eijnden, "Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics," *ChemPhysChem* **6**(9), 1809–1814 (2005).
- <sup>41</sup>L. Maragliano and E. Vanden-Eijnden, "A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations," *Chem. Phys. Lett.* **426**(1), 168–175 (2006).
- <sup>42</sup>J. B. Abrams and M. E. Tuckerman, "Efficient and direct generation of multi-dimensional free energy surfaces via adiabatic dynamics without coordinate transformations," *J. Phys. Chem. B* **112**(49), 15742–15757 (2008).
- <sup>43</sup>E. Kalliannaki, V. Harmandaris, M. A. Katsoulakis, and P. Plecháč, "The geometry of generalized force matching and related information metrics in coarse-graining of molecular systems," *J. Chem. Phys.* **143**(8), 084105 (2015).
- <sup>44</sup>W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.* **140**(4A), A1133 (1965).
- <sup>45</sup>B. Emily Moore and V. Molinero, "Structural transformation in supercooled water controls the crystallization rate of ice," *Nature* **479**(7374), 506–508 (2011).
- <sup>46</sup>F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," *Phys. Rev. B* **31**(8), 5262 (1985).
- <sup>47</sup>R. A. DiStasio, Jr., B. Santra, Z. Li, X. Wu, and R. Car, "The individual and collective effects of exact exchange and dispersion interactions on the *ab initio* structure of liquid water," *J. Chem. Phys.* **141**(8), 084502 (2014).
- <sup>48</sup>C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," *J. Chem. Phys.* **110**(13), 6158–6170 (1999).

- <sup>49</sup>A. Tkatchenko and M. Scheffler, “Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data,” *Phys. Rev. Lett.* **102**, 073005 (2009).
- <sup>50</sup>W. Lechner and C. Dellago, “Accurate determination of crystal structures based on averaged local bond order parameters,” *J. Chem. Phys.* **129**(11), 114707 (2008).
- <sup>51</sup>D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in 3rd International Conference on Learning Representations (ICLR); preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2015).
- <sup>52</sup>H. Wang, L. Zhang, J. Han, and W. E, “DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics,” *Comput. Phys. Commun.* **228**, 178–184 (2018).
- <sup>53</sup>L. Pauling, “The shared-electron chemical bond,” *Proc. Natl. Acad. Sci. U. S. A.* **14**(4), 359–362 (1928).
- <sup>54</sup>T. D. Kühne, M. Krack, and M. Parrinello, “Static and dynamical properties of liquid water from first principles by a novel Car-Parrinello-like approach,” *J. Chem. Theory Comput.* **5**(2), 235–241 (2009).
- <sup>55</sup>L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, and W. E, “End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems,” preprint [arXiv:1805.09003](https://arxiv.org/abs/1805.09003) (2018).
- <sup>56</sup>The training data set is the same, but the batches used at each step of the Adam iteration are picked from the data set randomly and independently for different models.