

MIT Open Access Articles

Coarse-graining auto-encoders for molecular dynamics

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Wang, Wujie and Rafael Gómez-Bombarelli. "Coarse-graining auto-encoders for molecular dynamics." npj Computational Materials, 5, 1 (December 2019): 125 © 2019 The Author(s)

As Published: 10.1038/S41524-019-0261-5

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/127224>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



ARTICLE OPEN

Coarse-graining auto-encoders for molecular dynamics

Wujie Wang¹ and Rafael Gómez-Bombarelli^{1*}

Molecular dynamics simulations provide theoretical insight into the microscopic behavior of condensed-phase materials and, as a predictive tool, enable computational design of new compounds. However, because of the large spatial and temporal scales of thermodynamic and kinetic phenomena in materials, atomistic simulations are often computationally infeasible. Coarse-graining methods allow larger systems to be simulated by reducing their dimensionality, propagating longer timesteps, and averaging out fast motions. Coarse-graining involves two coupled learning problems: defining the mapping from an all-atom representation to a reduced representation, and parameterizing a Hamiltonian over coarse-grained coordinates. We propose a generative modeling framework based on variational auto-encoders to unify the tasks of learning discrete coarse-grained variables, decoding back to atomistic detail, and parameterizing coarse-grained force fields. The framework is tested on a number of model systems including single molecules and bulk-phase periodic simulations.

npj Computational Materials (2019)5:125; <https://doi.org/10.1038/s41524-019-0261-5>

INTRODUCTION

Coarse-grained (CG) molecular modeling has been used extensively to simulate complex molecular processes with lower computational cost than all-atom simulations.^{1,2} By compressing the full atomistic model into a reduced number of pseudoatoms, CG methods focus on slow collective atomic motions while averaging out fast local motions. Current approaches generally focus on parameterizing coarse-grained potentials from atomistic simulations³ (bottom-up) or experimental statistics (top-down).^{4,5} The use of structure-based coarse-grained strategies has enabled important theoretical insights into polymer dynamics^{6–9} and lipid membranes¹⁰ at length scales that are otherwise inaccessible. Beyond efforts to parameterize CG potentials given a pre-defined all-atom to CG mapping, the selection of an appropriate map plays an important role in recovering consistent CG dynamics, structural correlation, and thermodynamics.^{11,12} A poor choice can lead to information loss in the description of slow collective interactions that are important for glass formation and transport. Systematic approaches to creating low-resolution protein models based on essential dynamics have been proposed,¹³ but a systematic bottom-up approach is missing for organic molecules of various sizes, resolutions, and functionalities. In general, the criteria for selecting CG mappings are usually based on a priori considerations and chemical intuition. Moreover, although there have been efforts in developing back-mapping algorithms,^{14–18} the statistical connections are missing to reversibly bridge resolutions across scales. We aim to address such multi-scale gaps in molecular dynamics using machine learning.

Recently, machine learning tools have facilitated the development of CG force fields^{19–23} and graph-based CG representations.^{24,25} Here we propose to use machine learning to optimize CG representations and deep neural networks to fit coarse-grained potentials from atomistic simulations. One of the central themes in learning theory is finding optimal hidden representations that capture complex statistical distributions to the highest possible fidelity using the fewest variables. We propose that finding coarse-grained variables can be formulated as a problem of learning latent variables of atomistic distributions. Recent work in unsupervised learning has shown great potential in uncovering

the hidden structure of complex data.^{26–29} As a powerful unsupervised learning technique, variational auto-encoders (VAEs) compress data through an information bottleneck³⁰ that continuously maps an otherwise complex data set into a low-dimensional space and can probabilistically infer the real data distribution via a generating process. VAEs have been applied successfully to a variety of tasks, from image de-noising³¹ to learning compressed representations for text,³² celebrity faces,³³ arbitrary grammars^{29,34}, and molecular structures.^{35,36} Recent studies have used VAE-like structures to learn collective molecular motions by reconstructing time-lagged configurations³⁷ and Markov state models.³⁸ For the examples mentioned, compression to a continuous latent space is usually parameterized using neural networks. However, coarse-grained coordinates are latent variables in 3D space, and need specially designed computational parameterization to maintain the Hamiltonian structure for discrete particle dynamics.

Motivated by statistical learning theory and advances in discrete optimization, we propose an auto-encoder-based generative modeling framework that (1) learns discrete coarse-grained variables in 3D space and decodes back to atomistic detail via geometric back-mapping; (2) uses a reconstruction loss to help capture salient collective features from all-atom data; (3) regularizes the coarse-grained space with a semi-supervised mean instantaneous force minimization to obtain a smooth coarse-grained free-energy landscape; and (4) variationally finds the highly complex coarse-grained potential that matches the instantaneous mean force acting on the all-atom training data.

RESULTS

Figure 1 shows the general schematics of the proposed framework, which is based on learning a discrete latent encoding by assigning atoms to coarse-grained particles. In Fig. 1b, we illustrate the computational graph of Gumbel-softmax reparameterization,^{39,40} which continuously relaxes categorical distributions for learning discrete variables. We first apply the coarse-grained auto-encoders to trajectories of individual gas-phase molecules. By variationally optimizing encoder and decoder networks to

¹Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02319, USA. *email: rafagb@mit.edu

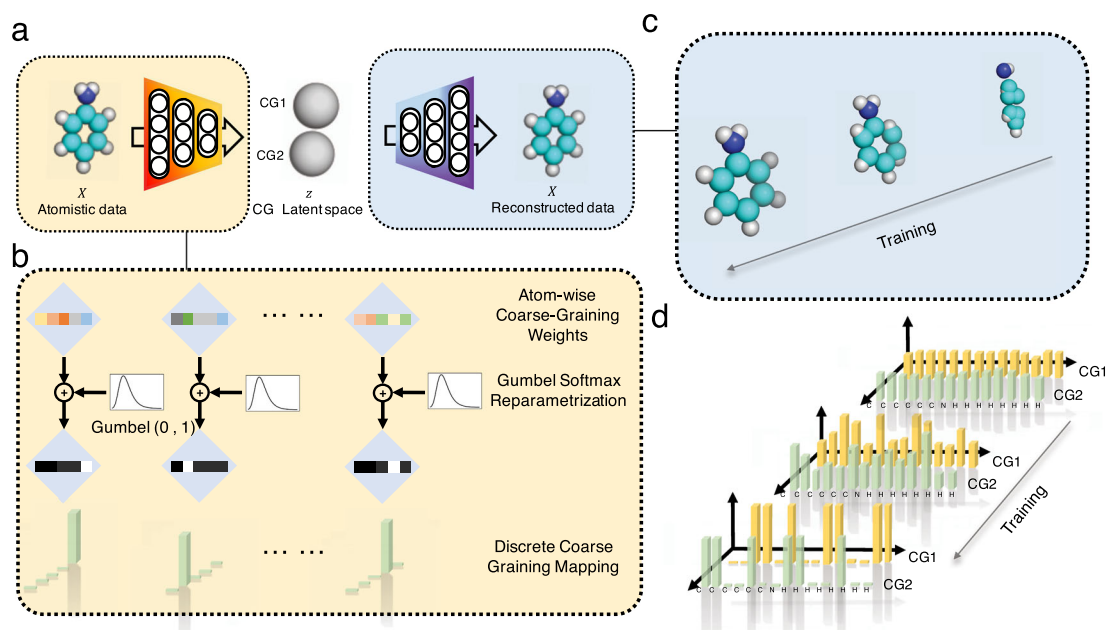


Fig. 1 Coarse-graining auto-encoding framework. **a** The model consists of an encoder and decoder, and is trained by reconstructing the original all-atom data by encoding atomistic trajectories through a low-dimensional bottleneck. **b** The computational graph to parameterize the CG mapping. The discrete optimization is done using the Gumbel-softmax reparametrization.^{39,40} **c** The learning task of reconstruction molecules conditioned on the CG variables in training time. The decoder is initialized with random weights, and these are variationally optimized to back-map atomistic coordinates with high accuracy. **d** Demonstration of continuously relaxation of CG mapping as in Eq. (6). In this demonstration, the x -axis represents individual atoms, the y -axis represents the two CG atoms. Each atom-wise CG assignment parameters is a vector of size 2 corresponding to the coarse-graining decision between the two coarse-grained beads. The discrete mapping operator is parametrized using the Gumbel-softmax reparametrization with a fictitious temperature τ . As τ approaches 0, the coarse-graining mapping operators effectively sample from a one-hot categorical distribution.

minimize the reconstruction loss as in Eq. (1), the auto-encoder picks up salient coarse-grained variables that minimize the fluctuation of encoded atomistic motions conditioned on a linear back-mapping function. We adopt an instantaneous-force regularizer (described in the Methods section), to minimize the force fluctuations of the encoded space. This facilitates the learning of a coarse-grained mapping that corresponds to a smoother coarse-grained free-energy landscape. For the unsupervised learning task, we minimize the following loss function.

$$L_{ae} = \frac{1}{N} \mathbb{E}_{x \sim P(x)} [(D(E(x)) - x)^2 + \rho F_{\text{inst}}(E(x))^2] \quad (1)$$

The first term on the right-hand side of Eq. (1) represents the atom-wise reconstruction loss and the second term represents the average instantaneous mean force regularization. The relative weight ρ is a hyperparameter describing the relative importance of the force regularization term. The force regularization loss is discussed in the Methods section and training details in the Supplementary Information.

We show the unsupervised auto-encoding process for gas-phase ortho-terphenyl (OTP) and aniline (C_6H_7N) in Fig. 2. The results show that the optimized reconstruction loss decreases with increasing coarse-grained resolution and that a small number of coarse-grained atoms have the potential to capture the overall collective motions of the underlying atomistic process. The reconstruction loss represents the information loss of coarse-grained particles to represent collective atomistic motions conditioned on a deterministic back-mapping. In the case of OTP, an intuitive 3-bead mapping is learned that partitions each of the phenyl rings. However, such an encoding loses the configuration information describing the relative rotation of the two side rings, resulting in decoded structures that yield higher error. When the number of coarse-grained degrees of freedom increases to 4, the additional beads are able to encode more configurational

information than three-bead models and therefore can decode back into atomistic coordinates with high accuracy. We further apply the auto-encoding framework to a small peptide molecule to examine, as a function of CG resolution, the capacity of the coarse-grained representation to capture the critical collective variables of the underlying atomistic states. Although it is not able to recover the arrangement of hydrogen atoms (Fig. 3), the coarse-grained latent variables of 8 CG atoms can faithfully recover heavy atom positions and represent different collective states in the Ramachandran map as the coarse-grained resolution is increased (Fig. 3).

The regularization term (second term in Eq. (1)) addresses the instantaneous mean forces that arise from transforming the all-atom forces. Inspired by gradient domain regularization in deep learning^{41–43} and the role of fluctuations in the generalized Langevin framework,⁴⁴ we minimize the average instantaneous force as a regularization term to facilitate the learning of a smooth coarse-grained free-energy surface and to average out fast dynamics. The factor ρ is a hyperparameter that controls the interplay between reconstruction loss and force regularization and is typically set to the highest value for which the CG encoding still uses all allotted dimensions.

In Figs 4, 5, 6, and 7, we demonstrate the applicability of the proposed framework to bulk simulations of liquids for small- (C_2H_6 , C_3H_8) and long-chain ($C_{24}H_{50}$) alkanes. Coarse-grained resolutions of 2 and 3 are used for ethane and propane, respectively, while two coarse-grained resolutions of 8 and 12 are used for the $C_{24}H_{50}$ alkane melt. We first train an auto-encoder to obtain the latent coarse-grained variables for ethane, propane, and $C_{24}H_{50}$, and subsequently train a neural network-based coarse-grained force field with additional excluded volume interactions using force matching to minimize Eq. (15) (in the case of $C_{24}H_{50}$, only the backbone carbon atoms are represented). Coarse-grained simulations are then carried out at the same

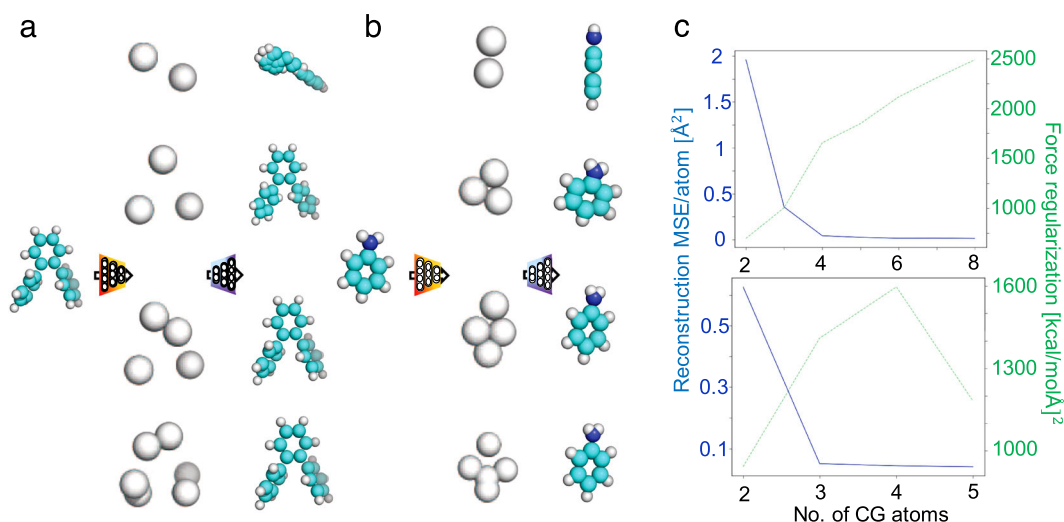


Fig. 2 Quantifying reconstruction and mean force losses in Auto-Encoder training of gas-phase molecules with different resolutions.

Coarse-graining encoding and decoding for OTP (**a**) and aniline (**b**) with different resolutions. As the coarse-grained resolution increases, the auto-encoder reconstructs molecules with higher accuracy. For the coarse-graining of OTP into 3 pseudo atoms, it automatically makes the coarse-graining decision of grouping each of the phenyl rings into one of the three pseudo-atoms and this mapping also yields a lower value for the mean force. For coarse-graining aniline into two atoms, the coarse-graining decision learned is to group the NH_2 moiety along with the two carbons and group the rest of the molecules into another pseudo-atom. However, we observe the coarse-graining mapping decision depends on the value of ρ which controls the force regularization in Eq. (1). When we choose a larger value of ρ , the mapping favors the coarse-graining decision of grouping the NH_2 and the phenyl group independently and this mapping choice yields a smaller average instantaneous mean force. **c** average instantaneous force residue and reconstruction loss of trained model. Although reconstruction loss decreases with higher resolutions, the average mean force increases with the coarse-graining resolutions because the coarse-graining mapping has increasingly rough underlying free-energy landscape that involves fast motions like bond stretching.

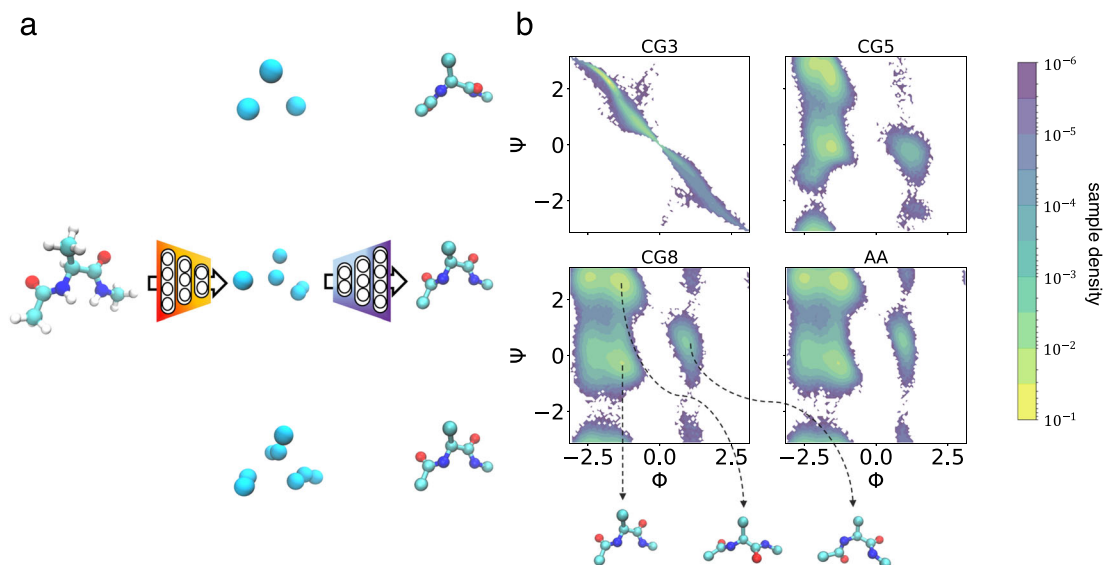


Fig. 3 Coarse-graining encoding and decoding for alanine dipeptide. **a** demonstrates the auto-encoding process for alanine dipeptide molecules at three different resolutions. Although the hydrogen atoms cannot be reconstructed accurately because of their relatively faster motions, the critical back-bone structures can be inferred with high accuracy by using a resolution of 3 CG atoms or greater. **b** comparison of dihedral correlation (Ramachandran map) between decoded atomistic distributions and atomistic data.

density and temperature as the atomistic simulation. We include the training details and model hyperparameters in the Supplementary Information. Coarse-grained forces are evaluated using PyTorch⁴⁵ and an MD integrator based on ASE (Atomistic Simulations Environment).⁴⁶

By minimizing the instantaneous force-matching loss term according to Eq. (15) in the Methods section, the neural network shows sufficient flexibility to reproduce a reasonably accurate structural correlation function. In the case of $\text{C}_{24}\text{H}_{50}$ (Figs 6 and 7), the neural network captures the bimodal bond length distribution

for the coarse-grained $\text{C}_{24}\text{H}_{50}$ chains and reproduces the end-to-end distance distribution and mapped monomer pair distribution function accurately. The mean squared displacement plots for all systems demonstrate faster dynamics than the atomistic ground truth due to loss of atomistic friction in the coarse-grained space. For $\text{C}_{24}\text{H}_{50}$, we also investigate the decoded C-C structural correlations shown in Fig. 8. The inter-chain structural correlation shows good agreement with the underlying atomistic ground truth, while the C-C bond distances are predicted to be shorter because the coarse-grained super-atoms can only infer average

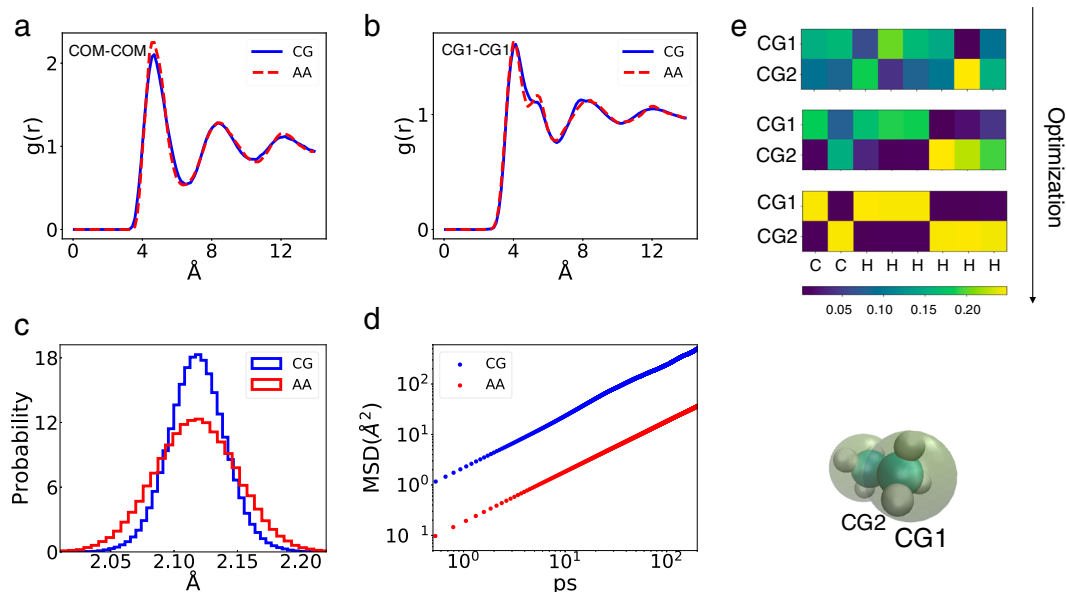


Fig. 4 Comparison between atomistic and CG simulation statistics for liquid ethane using a CG resolution of 2 per molecule. **a–c** are the pair correlation functions and bond length distributions of CG trajectories and mapped atomistic trajectories. **d** shows the comparison between Mean Squared Displacement for CG and mapped atomistic coordinates, and indicates that CG shows faster dynamics than the atomistic trajectory. **e** shows the learning of a discrete CG mapping during training of the auto-encoder. The rectangular matrix is a colored representation of matrix E_{ij} , the colors showing relative values of the matrix elements.

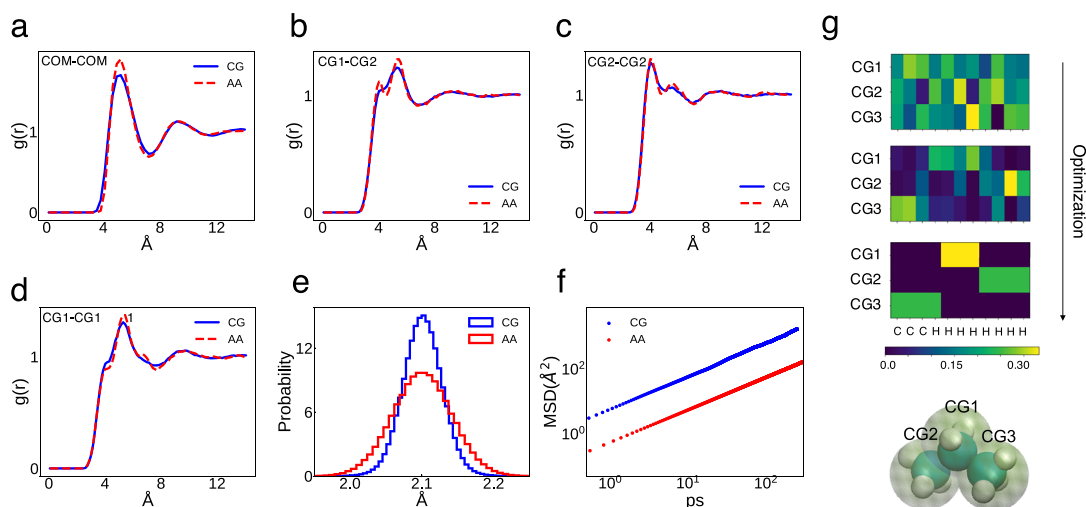


Fig. 5 Comparison between atomistic and CG simulation statistics for liquid propane with a CG resolution of 3 per molecule. **a–e** structural correlation for propane coarse-grained dynamics compared to the mapped atomistic trajectory. **f** shows the comparison between Mean Squared Displacement between CG and mapped atomistic coordinates. CGMD shows faster dynamics compared to the atomistic ground truth. **g** shows the learning of a discrete CG mapping during training of the auto-encoder. The rectangular matrix is a colored representation of matrix E_{ij} , the colors showing relative values of the matrix elements.

carbon poses based on the the deterministic inference framework using a linear back-mapping. The prospect of stochastic decoding functions to capture statistical up-scaling is discussed below.

DISCUSSION

Within the current framework, there are several possibilities for future research directions regarding both the supervised and unsupervised parts.

Here, we have presented a choice of deterministic encoder and decoder. However, such a deterministic CG mapping results, by construction, in an irreversible loss of information.

This is reflected in the reconstruction of average all-atom structures instead of the reference instantaneous configurations. To infer the underlying atomistic distributions, past methods have used random structure generation followed by equilibration.^{14–17} By combining this with predictive inference for atomistic back-mapping,¹⁸ a probabilistic auto-encoder can learn a reconstruction probability distribution that reflects the thermodynamics of the degrees of freedom that were averaged out by the coarse-graining. Using this framework as a bridge between different scales of simulation, generative models can help build better hierarchical understanding of multi-scale simulations.

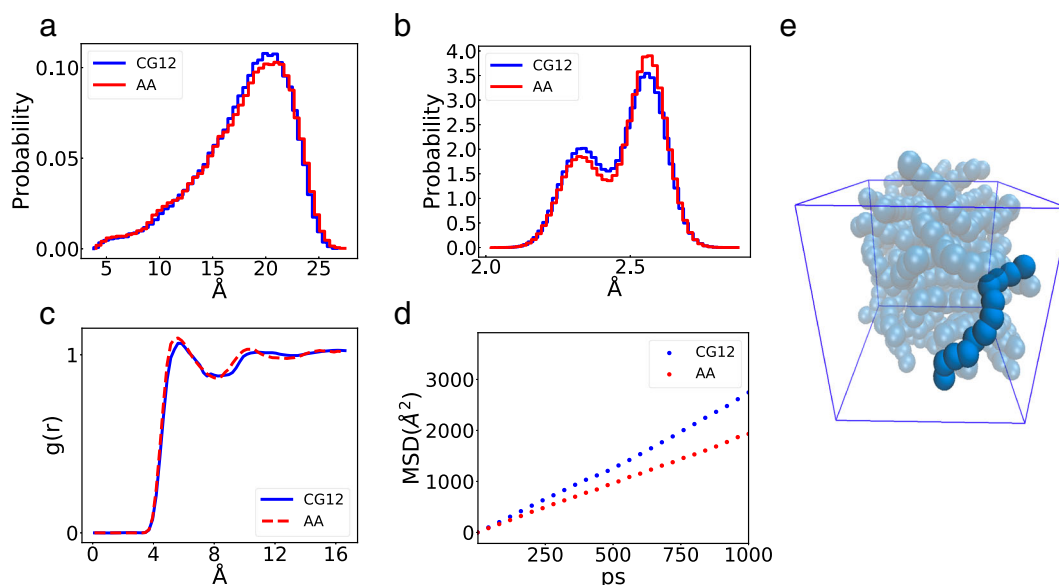


Fig. 6 Comparison between atomistic and CG simulation statistics for $C_{24}H_{50}$ with a coarse-grained resolution of 12. **a–c** are chain end-to-end distance, bond distance, and inter-chain radial distribution functions, respectively. All of these CG simulation statistics show good agreement with mapped atomistic ground truth. **d** The mean-squared displacement of the center of mass of the molecule for the CG and mapped atomistic trajectories. CG simulations show faster dynamics than the mapped atomistic kinetics. **e** Snapshot of the coarse-grained simulation box.

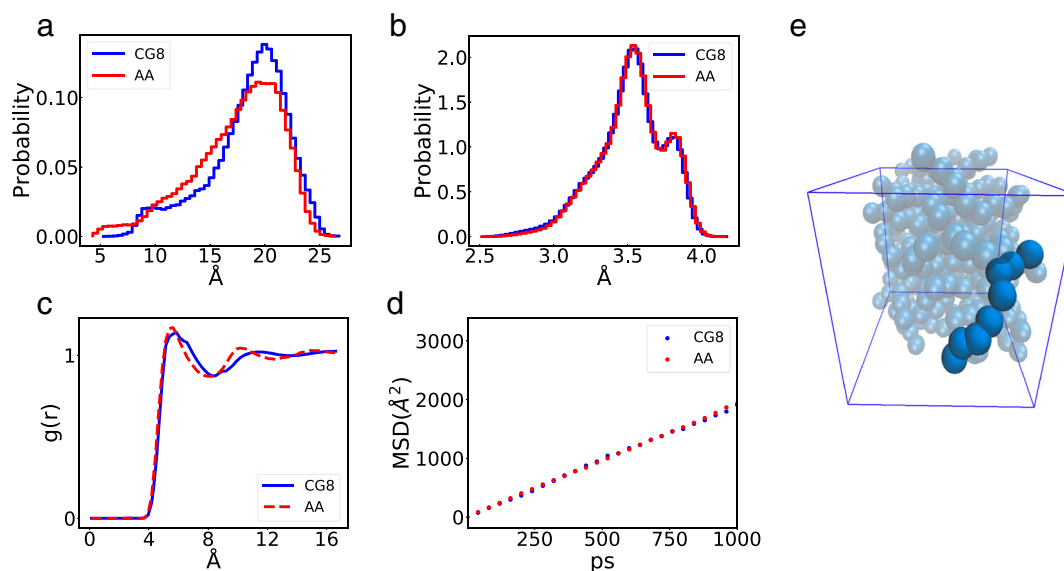


Fig. 7 Comparison between atomistic and CG simulation statistics for $C_{24}H_{50}$ with a coarse-grained resolution of 8. **a–c** are chain end-to-end distance, bond distance, and inter-chain radial distribution functions, respectively. All of these CG simulation statistics show good agreement with mapped atomistic ground truth. **d** The mean-squared displacement of the center of mass of the molecule for CG and mapped atomistic trajectories. CG simulations show comparable dynamics to the mapped atomistic kinetics. **e** Snapshot of the coarse-grained simulation box.

Furthermore, neural network potentials provide a powerful fitting framework to capture many-body correlations. The choice of force-matching approach does not guarantee the recovery of individual pair correlation functions derived from full atomistic trajectories^{12,47} because the cross-correlations among coarse-grained degrees of freedom are not explicitly incorporated. More advanced fitting methods can be incorporated in the current neural network framework to address the learning of structural cross-correlation, including iterative force matching⁴⁷ and relative entropy method.⁴⁸

Methods based on force-matching, like other bottom-up approaches such as relative entropy method, attempt to reproduce structural correlation functions at one point in the thermodynamic space. As such, they are not guaranteed to capture non-equilibrium transport properties^{12,49} and are not necessarily transferable among different thermodynamic conditions.^{12,50–53} The data-driven approach we propose enables learning over different thermodynamic conditions. In addition, this framework opens new routes to understanding how the coarse-grained representation influences transport properties by

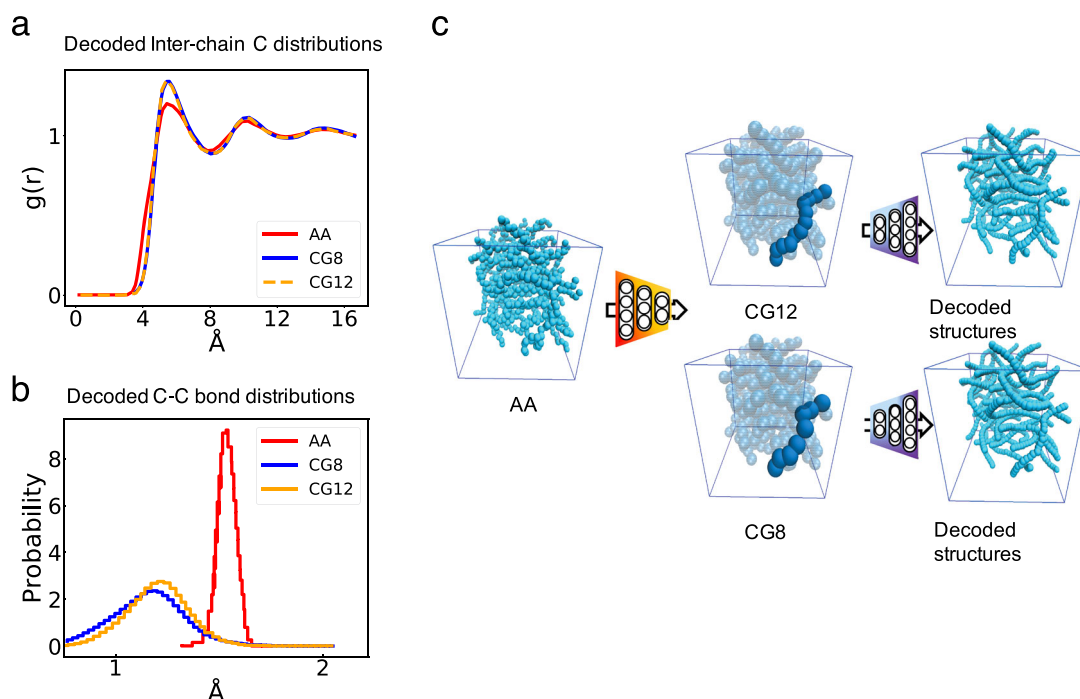


Fig. 8 Comparison of the decoded backbone carbon atoms with the atomistic ground truth. **a** the inter-chain C–C radial distributions shows reasonable agreement between the decoded and original distributions. **b** The decoded carbon backbones show shorter predicted bond length because the decoded structures represent the mean reconstruction of an ensemble of carbon chain poses. **c** demonstrates the auto-encoding of carbon backbones for $C_{24}H_{50}$ molecules. As a result of loss of mapping entropy, the decoded structures show a straight backbone compared to the atomistic ground truth.

training on time-series data. A related example in the literature is to use a time-lagged auto-encoder³⁷ to learn a latent representation that best captures molecular kinetics.

In summary, we propose to treat the coarse-grained coordinates as latent variables which can be sampled with coarse-grained molecular dynamics. By regularizing the latent space with force regularization, we train the encoding mapping, a deterministic decoding, and a coarse-grained potential that can be used to simulate larger systems for longer times and thus accelerate molecular dynamics simulations. Our work also enables the use of statistical learning as a basis to bridge across multi-scale coarse-grained simulations.

METHODS

Here we introduce the auto-encoding framework from the generative modeling point of view. The essential idea is to treat coarse-grained coordinates as a set of latent variables that are the most predictive of the atomistic distribution while having a smooth underlying free-energy landscape. We show that this is achieved by minimizing the reconstruction loss and the instantaneous force regularization term. Moreover, under the variational auto-encoding framework, we can understand the force matching as the minimization of the relative entropy between coarse-grained and atomistic distributions in the gradient domain.

Coarse-graining auto-encoding

The essential idea in generative modeling is to maximize the likelihood of the data under the generative process:

$$P(x) = \int P(x|z)P(z)dx \quad (2)$$

where z are the latent variables that carry the essential information of the distributions and x represents the samples observed in the data. Variational auto-encoders maximize the likelihood of the observed samples by

maximizing the evidence lower bound (ELBO):

$$L = \mathbb{E}_{Q_\phi(z|x)} \log P_D(x|z) - \mathbb{E}_{Q_\phi(z|x)} \log \frac{Q_\phi(z|x)}{P(z)} \quad (3)$$

where $Q_\phi(z|x)$ encodes the data into latent variables, $P_D(x|z)$ is the generative process parameterized by D , and $P(z)$ is the prior distribution (usually a multivariate Gaussian with a diagonal covariance matrix) which imposes a statistical structure over the latent variables. Minimizing the ELBO by propagating gradients through the probability distributions provides a parameterizable way of inferring complicated distributions of molecular dynamics.

Similar to variational auto-encoders with constraint on the latent space, a coarse-grained latent space should preserve the structure of the molecular mechanics phase space. Noid et al.³ have studied the general requirements for a physically rigorous encoding function. In order to address those requirements, the auto-encoder is trained to optimize the reconstruction of atomistic configurations by propagating them through a low-dimensional bottleneck in Cartesian coordinates. Unlike most instances of VAEs, the dimensions of the CG latent space have physical meaning. Since the CG space needs to represent the system in position and momentum space, latent dimensions need to correspond to real-space Cartesian coordinates and maintain the essential structural information of molecules.

We make our encoding function a linear projection in Cartesian space $E(x) : \mathbb{R}^{3n} \rightarrow \mathbb{R}^{3N}$ where n is the number of atoms and N is the desired number of coarse-grained particles.

Let x be the atomistic coordinates and z be the coarse-grained coordinates. The encoding function should satisfy the following requirements:^{3,54}

1. $z_{ik} = E(x) = \sum_{j=1}^n E_{ij} x_{jk} \in \mathbb{R}^3, i = 1 \dots N, j = 1 \dots n,$
2. $\sum_j E_{ij} = 1$ and $E_{ij} \geq 0$
3. Each atom contributes to at most one coarse-grained variable z

where E_{ij} defines the assignment matrix to coarse-grained variables, j is the atomic index, i is the coarse-grained atom index, and k represents the Cartesian coordinate index. Requirement (2) defines the coarse-grained variables to be a weighted geometric average of the Cartesian coordinates of the contributing atoms. In order to maintain consistency in momentum space after the coarse-grained mapping, the coarse-grained masses are

redefined as $M_i = (\sum_j \frac{E_j^2}{m_j})^{-1}$ ^{3,54} (m_j is the mass of atom j). This definition of mass is a corollary of requirement (3).

The encoder function parameters are initialized randomly as atom-wise vectors $\vec{\phi}_j$, with its elements ϕ_{ij} representing the parameter of assigning individual atom j to coarse-grained atom i ; ϕ_{ij} is further reparameterized to obtain coarse-graining encoding weights satisfying the requirements shown above. The goal of the parameterizable coarse-graining encoding function is to learn a one-hot assignment from each atom to a coarse-grained variable. Its weights are obtained by normalizing over the total number of contributing atoms per coarse-grained atom (an alternative is to normalize based on atomic masses or charges), thus satisfying requirement (2):

$$E_{ij} = \frac{C_{ij}}{\sum_j C_{ij}} \quad (4)$$

$$\vec{C}_j = \text{one_hot}(\arg \max_i \log \phi_{ij}) \quad (5)$$

\vec{C}_j is the coarse-graining one-hot assignment vector for atom j using Gumbel-softmax reparameterization with each vector element C_{ij} representing the assignment of atom j to coarse-grained atom i , so that requirement (3) is automatically satisfied. Gumbel-softmax reparameterization is a continuous relaxation of Gumbel-max reparameterization for differentiable approximation using the softmax function.⁴⁰ Similar parameterization techniques include concrete distributions,³⁹ REBAR⁵⁵ and RELAX.⁵⁶ The Gumbel-softmax reparameterization has been applied in various machine learning scenarios involving learning discrete structures,⁵⁷ propagating discrete policy gradient in reinforcement learning⁵⁸ and generating context-free grammar.³⁴ The continuously relaxed version of Eq. (5) is:

$$C_{ij} = \frac{e^{(\log \phi_{ij} + g_{ij})/\tau}}{\sum_j e^{(\log \phi_{ij} + g_{ij})/\tau}} \quad (6)$$

where g_{ij} is sampled from the Gumbel distribution via the inverse transformation $g_{ij} = -\log(-\log(u_{ij}))$ where u_{ij} is sampled from a uniform distribution from 0 to 1. During training, τ is gradually decreased with the training epoch and the one-hot categorical encoding is achieved in the limit of small τ . Therefore, the encoding distribution $Q(z|x)$ is a linear projection operator parameterized by discrete atom-wise categorical variables.

For the generation of atomistic coordinates conditioned on coarse-grained coordinates, we opt for a simple decoding approach via geometrical projection using a matrix \mathbf{D} of dimension n by N that maps coarse-grained variables back to the original space so that $\hat{x} = D(z) = \sum_{i=1}^N \mathbf{D}_{ji} z_{ik}$ where \hat{x} are the reconstructed atomistic coordinates. Hence, both the encoding and decoding mappings are deterministic. However, deterministic reconstruction via a low-dimensional space leads to irreversible information loss that is analogous to the mapping entropy introduced in Shell et al.⁴⁸ In our experiments, by assuming $P_D(x|z)$ is Gaussian, the reconstruction loss yields the term-by-term mean-squared error and is understood as a Gaussian approximation to the mapping entropy (scaled by the variance) defined by Shell et al.:

$$S_{\text{map}} = \mathbb{E}_{x \sim P(x)} \log \frac{P(x)\Omega(E(x))}{P(z)} \quad (7)$$

where $\Omega(E(x))$ is the configuration space volume that is mapped to the atomistic coordinates. The latent variable framework provides a clear parameterizable objective whose optimization minimizes the information loss due to coarse-graining by using the following objective as reconstruction loss.

$$\min_{D, \phi} L_{AE} = \min_{D, \phi} \mathbb{E}_{x \sim P(x)} \mathbb{E}_{g \sim \text{Gumbel}(0, 1)} (D(E(x, g, \tau)) - x)^2 \quad (8)$$

Hence, we present an analogous interpretation of the reconstruction loss in Eq. (3) but in the Cartesian space of coarse-grained pseudo-atoms in molecular dynamics. This loss can be optimized by algorithm 1. A regularized version is introduced in section C.

Algorithm 1 Variational Coarse-graining Auto-Encoding

$\phi_{ij}, D_{ji}, \tau, \Delta\tau \leftarrow$ initialize parameters

repeat

$x \leftarrow$ random mini-batch molecular dynamics frames $x \sim P(x)$

$g_{ij} \leftarrow \text{Gumbel}(0, 1)$

$C_{ij} \leftarrow \frac{e^{(\log \phi_{ij} + g_{ij})/\tau}}{\sum_j e^{(\log \phi_{ij} + g_{ij})/\tau}}$

$E_{ij} \leftarrow \frac{C_{ij}}{\sum_j C_{ij}}$

$g \leftarrow \nabla_{\phi_{ij}, D_{ji}} L_{AE}(\phi_{ij}, D_{ji}; \tau, g_{ij})$

$\phi_{ij}, D_{ji} \leftarrow$ update parameters using gradients g

$\tau \leftarrow \tau - \Delta\tau$

until convergence of L_{AE}

Variational force matching

The physical meaning of the regularization term has a natural analogy to the minimization of Kullback–Leibler divergence (KL divergence for short, also called relative entropy) in coarse-grained modeling to reduce the discrepancy between mapped atomistic distributions and coarse-grained distributions conditioned on a Boltzmann prior. The distribution function of coarse-grained variables $p(z)$ and the corresponding many-body potential of mean force $A(z)$ are:

$$P_{CG}(z) = \frac{1}{Z} \int e^{-\beta V(x)} \delta(E(x) - z) dx \quad (9)$$

where $V(x)$ is the atomistic potential energy function and $E(x)$ is the encoding function defined by requirement (2). Unlike the VAE, which assumes a prior Gaussian structure in the latent space, the coarse-grained latent prior (1) is variationally determined by fitting the coarse-grained energy function, and (2) has no closed-form expression for the KL loss. To recover the true $P_{CG}(z)$ requires constrained sampling to obtain the coarse-grained free-energy. To bypass such difficulties, we parameterize the latent distributions by matching the instantaneous mean forces. In order to learn the coarse-grained potential energy V_{CG} as a function of also-learned coarse-grained coordinates, we propose an instantaneous force-matching functional that is conditioned on the encoder. Unlike regularizing KL loss in the context of training a VAE, which is straightforward to evaluate, the underlying coarse-grained distributions are intractable. However, matching the gradient of the log likelihood of mapped coarse-grained distributions (the mean force) is more computationally feasible. Training potentials from forces has a series of advantages: (i) the explicit contribution on every atom is available, rather than just pooled contributions to the energy, (ii) it is easier to learn smooth potential energy surfaces and energy-conserving potentials⁵⁹, and (iii) instantaneous dynamics, which represent a trade-off in coarse-graining, can be better captured. Forces are always available if the training data comes from molecular dynamics simulations, and for common electronic structure methods based on density functional theory, forces can be calculated at nearly the same cost as self-consistent energies.

The force-matching approach builds on the idea that the average force generated by the coarse-grained potential V_{CG} should reproduce the coarse-grained atomistic forces from thermodynamic ensembles.^{19,60,61}

Given an atomistic potential energy function $V(x)$ with the partition function Z , the probabilistic distribution of atomistic configurations is:

$$P(x) = \frac{1}{Z} e^{-\beta V(x)} \quad (10)$$

The distribution function of coarse-grained variables $P_{CG}(z)$ and the corresponding many-body potential of mean force $A(z)$ are:

$$A(z) = -\frac{1}{\beta} \ln P_{CG}(z) \quad (11)$$

The mean force of the coarse-grained variables is the average of the instantaneous forces conditioned on $E(x) = z$ ^{54,62}, assuming the coarse

grained mapping is linear:

$$-\frac{dA}{dz} = F(z) = \langle -\mathbf{b}\nabla V(x) \rangle_{E(x)=z} \quad (12)$$

$$\mathbf{b} = \frac{\mathbf{w}}{\mathbf{w}^\top \nabla E(x)} \quad (13)$$

where $F(z)$ is the mean force and \mathbf{b} represents a family of possible vectors such that $\mathbf{w}^\top \nabla E(x) \neq 0$. We further define $F_{\text{inst}}(z) = -\mathbf{b}\nabla V(x)$ to be the instantaneous force and its conditional expectation is equal to the mean force $F(z)$. It is important to note that $F_{\text{inst}}(z)$ is not unique and depends on the specific choice of \mathbf{w} ,^{61–63} but their conditional averages return the same mean force. For possible \mathbf{b} , we further choose $\mathbf{w} = \nabla E(x)$ which is a well-studied choice,^{61,63} so that:

$$\mathbf{b} = \frac{\nabla E(x)}{\nabla E(x)^\top \nabla E(x)} = \mathbf{C} \quad (14)$$

where \mathbf{b} is a function of $\nabla E(x)$. In the case of coarse-graining encodings, $\mathbf{b} = \mathbf{C}$ where \mathbf{C} is the encoding matrix formed by concatenating atom-wise one-hot vectors as defined in Eq. (6). We adopt the force-matching scheme introduced by Izvekov et al.^{60,64}, in which the mean-squared error is used to match the mean force and the “coarse-grained force” is the negative gradient of the coarse-grained potential. The optimizing functional, developed based on Izvekov et al., is

$$\min_{\theta} L = \min_{\theta} \mathbb{E}[(F(z) + \nabla_z V_{CG}(z))^2] \quad (15)$$

where θ are the parameters in V_{CG} and ∇V_{CG} represents the “coarse grained forces” which can be obtained from automatic differentiation as implemented in open-source packages like PyTorch.⁴⁵ However, to compute the mean force F would require constrained dynamics⁶¹ to obtain the average of the fluctuating microscopic forces. According to Zhang et al.¹⁹, the force-matching functional can be alternatively formulated by treating the instantaneous mean force as an instantaneous observable with a well-defined average being the mean force $F(z)$:

$$F_{\text{inst}}(z) = F(z) + \epsilon(z) \quad (16)$$

on the condition that $\mathbb{E}_z[F_{\text{inst}}] = F(z)$. The original variational functional becomes instantaneous in nature and can be reformulated as the following minimization target:

$$\min_{\theta} L_{\text{inst}} = \min_{\theta} \mathbb{E}[F_{\text{inst}}(z) + \nabla V_{CG}(z)]^2 \quad (17)$$

Instead of matching mean forces that need to be obtained from constrained dynamics, our model minimizes L_{inst} with respect to $V_{CG}(z)$ and $E(x)$. L_{inst} can be shown to be related to L with some algebra: $L_{\text{inst}} = L + \mathbb{E}[e(E(x))^2]$.¹⁹ This functional provides a variational way to find a CG mapping and its associated force field functions.

Instantaneous mean force regularization

Here we introduce the gradient regularization term that is designed to minimize the fluctuation in the mean forces. Similar methods involving gradient regularization have been applied in supervised learning computer vision tasks to smoothen the loss landscape for improved model generalization.^{41–43} In coarse-grained modeling, minimizing the forces is important for learning the slow degrees of freedom and a smoother free-energy surface.

Based on the generalized Langevin equation, the difference between the true mean force and instantaneous mean force $\epsilon(E(x))$ can be approximated as:^{44,65}

$$\epsilon(E(x), t) = \gamma \frac{dE(x(t))}{dt} - \int_0^t \beta(\tau) \frac{dE(x(t-\tau))}{dt} d\tau + \tilde{\eta}(t) + \sum_i C_{ij} \eta_j \quad (18)$$

where γ is the friction coefficient, $\beta(\tau)$ is the memory kernel, $\tilde{\eta}(t)$ is the colored Gaussian noise, and $\sum_i C_{ij} \eta_j$ is the mapped atomistic white noise. To avoid the need for special dynamics when running ensemble calculations, it is desirable to minimize the memory and fluctuation term to yield dynamics with fewer fluctuation terms. A related example in the work by Guttenberg et al.⁴⁴ who compare the memory heuristics among coarse-grained mapping function. The objective we propose can be optimized by gradient descent to continuously explore the coarse-grained mapping space without iterating over the combinatorial spaces. We perform this regularization by minimizing the mean-squared instantaneous forces over mini-batches of atomistic trajectories to optimize the CG

mappings.

$$\min_{\phi, D} \mathbb{E}_{x \sim P(x)} (F_{\text{inst}}(z))^2 = \min_{\phi, D} \mathbb{E}_{x \sim P(x)} F(z)^2 + \epsilon^2(E(x)) \quad (19)$$

In practice, this regularization loss is combined with L_{ae} to obtain a coarse-grained mapping with a certain weight ρ that is added onto the reconstruction loss. We discuss the practical effect of including the regularization term in the Supplementary Information.

DATA AVAILABILITY

Data for training the model is available upon request. An implementation of the algorithm described in the paper is available at <https://github.com/learningmattermit/Coarse-Graining-Auto-encoders>.

Received: 27 March 2019; Accepted: 19 November 2019;

Published online: 18 December 2019

REFERENCES

- Agostino, M. D., Risselada, H. J., Lürick, A., Ungermann, C. & Mayer, A. A tethering complex drives the terminal stage of SNARE-dependent membrane fusion. *Nature* **551**, 634–638 (2017).
- Huang, D. M. et al. Coarse-grained computer simulations of polymer / fullerene bulk heterojunctions for organic photovoltaic applications. *J. Chem. Theory Comput.* **6**, 1–11 (2010).
- Noid, W. G. et al. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **128**, 243116 (2008).
- Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
- Periole, X., Cavalli, M., Marrink, S.-J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).
- Wijesinghe, S., Perahia, D. & Grest, G. S. Polymer topology effects on dynamics of comb polymer melts. *Macromolecules* **51**, 7621–7628 (2018).
- Salerno, K. M., Agrawal, A., Peters, B. L., Perahia, D. & Grest, G. S. Dynamics in entangled polyethylene melts. *Eur. Phys. J. Spec. Topics* **225**, 1707–1722 (2016).
- Salerno, K. M., Agrawal, A., Perahia, D. & Grest, G. S. Resolving dynamic properties of polymers through coarse-grained computational studies. *Phys. Rev. Lett.* **116**, 058302 (2016).
- Xia, W. et al. Energy renormalization for coarse-graining polymers having different segmental structures. *Sci. Adv.* **5**, eaav4683 (2019).
- Vögele, M., Köfinger, J. & Hummer, G. Hydrodynamics of diffusion in lipid membrane simulations. *Phys. Rev. Lett.* **120**, 268104 (2018).
- Rudzinski, J. F. & Noid, W. G. Investigation of coarse-grained mappings via an iterative generalized Yvon-Born-Green method. *J. Phys. Chem. B* **118**, 8295–8312 (2014).
- Noid, W. G. Perspective: coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 90901 (2013).
- Zhang, Z. et al. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys. J.* **95**, 5073–5083 (2008).
- Peng, J., Yuan, C., Ma, R. & Zhang, Z. Backmapping from multiresolution coarse-grained models to atomic structures of large biomolecules by restrained molecular dynamics simulations using bayesian inference. *J. Chem. Theory Comput.* **15**, 3344–3353 (2019).
- Chen, L. J., Qian, H. J., Lu, Z. Y., Li, Z. S. & Sun, C. C. An automatic coarse-graining and fine-graining simulation method: application on polyethylene. *J. Phys. Chem. B* **110**, 24093–24100 (2006).
- Lombardi, L. E., Marti, M. A. & Capece, L. CG2AA: backmapping protein coarse-grained structures. *Bioinformatics* **32**, 1235–1237 (2016).
- Machado, M. R. & Pantano, S. SIRAH tools: mapping, backmapping and visualization of coarse-grained models. *Bioinformatics* **32**, 1568–1570 (2016).
- Schöberl, M., Zabarar, N. & Koutsourelakis, P.-S. Predictive coarse-graining. *J. Comput. Phys.* **333**, 49–77 (2017).
- Zhang, L., Han, J., Wang, H., Car, R. & W, E. W. DeePCG: constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **149**, 034101 (2018).
- Bejagam, K. K., Singh, S., An, Y. & Deshmukh, S. A. Machine-learned coarse-grained models. *J. Phys. Chem. Lett.* **9**, 4667–4672 (2018).
- Lemke, T. & Peter, C. Neural network based prediction of conformational free energies - a new route toward coarse-grained simulation models. *J. Chem. Theory Comput.* **13**, 6213–6221 (2017).

22. Wang, J. et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **5**, 755–767 (2019).
23. Boninsegna, L., Gobbo, G., Noé, F. & Clementi, C. Investigating molecular kinetics by variationally optimized diffusion maps. *J. Chem. Theory Comput.* **11**, 5947–5960 (2015).
24. Webb, M. A., Delannoy, J.-Y. & de Pablo, J. J. Graph-based approach to systematic molecular coarse-graining. *J. Chem. Theory Comput.* **15**, 1199–1208 (2018).
25. Chakraborty, M., Xu, C. & White, A. D. Encoding and selecting coarse-grain mapping operators with hierarchical graphs. *J. Chem. Phys.* **149**, 134106 (2018).
26. Tolstikhin, I., Bousquet, O., Gelly, S., Schölkopf, B. & Schoelkopf, B. Wasserstein Auto-Encoders. In *Proc. International Conference on Learning Representations* (2018).
27. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *Proc. International Conference on Learning Representations* (2015).
28. Goodfellow, I. J. et al. Generative adversarial networks. In *Proc. Advances in Neural Information Processing Systems* (2014).
29. Kusner, M. J., Paige, B. & Hernández-Lobato, J. M. Grammar Variational Auto-encoder. In *Proc. International Conference on Machine Learning* (2017).
30. Tishby, N. & Zaslavsky, N. Deep Learning and the Information Bottleneck Principle. <https://arxiv.org/abs/1503.02406> (2015).
31. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371–3408 (2010).
32. Bowman, S. R. et al. Generating sentences from a continuous space. In *Proc. SIGLL Conference on Computational Natural Language Learning (CONLL)* (2016).
33. Liu, Z., Luo, P., Wang, X. & Tang, X. Deep learning face attributes in the wild. In *Proc. International Conference on Computer Vision (ICCV)* (2015).
34. Kusner, M. J. & Hernández-Lobato, J. M. GANS for sequences of discrete elements with the gumbel-softmax distribution. Preprint at <https://arxiv.org/abs/1611.04051> (2016).
35. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
36. Jin, W., Barzilay, R. & Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In *Proc. International Conference on Machine Learning* (2018).
37. Wehmeyer, C. & Noé, F. Time-lagged autoencoders: deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **148**, 241703 (2018).
38. Mardt, A., Pasquali, L., Wu, H. & Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **9**, 5 (2018).
39. Maddison, C. J., Mnih, A. & Teh, Y. W. The concrete distribution: a continuous relaxation of discrete random variables. In *Proc. International Conference on Learning Representations* (2016).
40. Jang, E., Gu, S. & Poole, B. Categorical reparameterization with gumbel-softmax. In *Proc. International Conference on Learning Representations* (2017).
41. Drucker, H. & LeCun, Y. Improving generalization performance using double backpropagation. *IEEE Trans. Neural Netw.* **3**, 991–997 (1992).
42. Varga, D., Csiszárík, A. & Zombori, Z. Gradient regularization improves accuracy of discriminative models. *arXiv* <https://arxiv.org/abs/1712.09936> (2017).
43. Hoffman, J., Roberts, D. A. & Yaida, S. Robust learning with jacobian regularization. *arXiv* <https://arxiv.org/abs/1908.02729> (2019).
44. Guttenberg, N. et al. Minimizing memory as an objective for coarse-graining. *J. Chem. Phys.* **138**, 094111 (2013).
45. Paszke, A. et al. Automatic differentiation in pytorch. In *NIPS-Workshop* (2017).
46. HjorthLarsen, A. et al. The atomic simulation environment - A Python library for working with atoms. *Matter* **29**, 273002 (2017).
47. Lu, L., Dama, J. F. & Voth, G. A. Fitting coarse-grained distribution functions through an iterative force-matching method. *J. Chem. Phys.* **139**, 121906 (2013).
48. Shell, M. S. Coarse-Graining With The Relative Entropy. In *Advances in Chemical Physics*, Vol. 161, p. 395–441 (Wiley-Blackwell, 2016).
49. Davtyan, A., Dama, J. F., Voth, G. A. & Andersen, H. C. Dynamic force matching: a method for constructing dynamical coarse-grained models with realistic time dependence. *J. Chem. Phys.* **142**, 154104 (2015).
50. Carbone, P., Varzaneh, H. A. K., Chen, X. & Müller-Plathe, F. Transferability of coarse-grained force fields: the polymer case. *J. Chem. Phys.* **128**, 64904 (2008).
51. Krishna, V., Noid, W. G. & Voth, G. A. The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *J. Chem. Phys.* **131**, 24103 (2009).
52. Xia, W. et al. Energy renormalization for coarse-graining the dynamics of a model glass-forming liquid. *J. Phys. Chem. B* **122**, 2040–2045 (2018).
53. Xia, W. et al. Energy-renormalization for achieving temperature transferable coarse-graining of polymer dynamics. *Macromolecules* **50**, 8787–8796 (2017).
54. Darve, E. Numerical methods for calculating the potential of mean force. In *New Algorithms for Macromolecular Simulation*, p. 213–249 (Springer-Verlag, Berlin/Heidelberg, 2006).
55. Tucker, G., Mnih, A., Maddison, C. J., Lawson, D. & Sohl-Dickstein, J. REBAR low-variance, unbiased gradient estimates for discrete latent variable models. In *Advances in Neural Information Processing Systems*. Vol. 2017, p. 2628–2637 (2017).
56. Grathwohl, W., Choi, D., Wu, Y., Roeder, G. & Duvenaud, D. Backpropagation through the Void: optimizing control variates for black-box gradient estimation. In *Proc. International Conference on Learning Representations* (2017).
57. Van Den Oord, A., Vinyals, O. & Kavukcuoglu, K. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*. Vol. 2017, p. 6307–6316 (2017).
58. Wu, Y., Wu, Y., Gkioxari, G. & Tian, Y. Building generalizable agents with a realistic and rich 3D environment. <https://arxiv.org/abs/1801.02209> (2018).
59. Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
60. Izvekov, S. & Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **109**, 2469–2473 (2005).
61. Ciccotti, G., Kapral, R. & Vanden-Eijnden, E. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem* **6**, 1809–1814 (2005).
62. Kalligiannaki, E., Harmandaris, V., Katsoulakis, M. A. & Plecháč, P. The geometry of generalized force matching and related information metrics in coarse-graining of molecular systems. *J. Chem. Phys.* **143**, 84105 (2015).
63. DenOtter, W. K. Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates. *J. Chem. Phys.* **112**, 7283–7292 (2000).
64. Izvekov, S. & Voth, G. A. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *J. Chem. Theory Comput.* **2**, 637–648 (2006).
65. Lange, O. F. & Grubmüller, H. Collective Langevin dynamics of conformational motions in proteins. *J. Chem. Phys.* **124**, 214903 (2006).

ACKNOWLEDGEMENTS

W.W. thanks Toyota Research Institute for financial support. R.G.B. thanks MIT DMSE and Toyota Faculty Chair for support. W.W. and R.G.B. thank Prof. Adam P. Willard (Massachusetts Institute of Technology) and Prof. Salvador Leon Cabanillas (Universidad Politécnica de Madrid) for helpful discussions. W.W. thanks Mr. William H. Harris for proofreading the manuscript and helpful discussions.

AUTHOR CONTRIBUTIONS

R.G.B. conceived the project, W.W. wrote the computer software and carried out simulations with contributions from R.G.B.; both authors wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41524-019-0261-5>.

Correspondence and requests for materials should be addressed to R.G.-B.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019