Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach FREE

Special Collection: Machine Learning Meets Chemical Physics

Wei Li ✉ 🆔 ; Craig Burkhart 🆔 ; Patrycja Polińska 🆔 ; Vagelis Harmandaris 🆔 ; Manolis Doxastakis 🆔

Check for updates

View Online   Export Citation

Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach

Wei Li,[1,a] [ID] Craig Burkhart,[2] [ID] Patrycja Polińska,[3] [ID] Vagelis Harmandaris,[4,5] [ID] and Manolis Doxastakis[1,b] [ID]

AFFILIATIONS

[1] Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA
[2] The Goodyear Tire and Rubber Company, Akron, Ohio 44305, USA
[3] Goodyear Innovation Center Luxembourg, Avenue Gordon Smith, L-7750 Colmar-Berg, Luxembourg
[4] Department of Applied Mathematics, University of Crete, and IACM FORTH, GR-71110 Heraklion, Greece
[5] Computation-Based Science and Technology Research Center, The Cyprus Institute, Nicosia 2121, Cyprus

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.
[a] Author to whom correspondence should be addressed: weili@utk.edu
[b] E-mail: edoxasta@utk.edu

ABSTRACT

Multiscale modeling of polymers exchanges information between coarse and fine representations of molecules to capture material properties over a wide range of spatial and temporal scales. Restoring details at a finer scale requires us to generate information following embedded physics and statistics of the models at two different levels of description. Techniques designed to address this persistent challenge balance among accuracy, efficiency, and general applicability. In this work, we present an image-based approach for structural backmapping from coarse-grained to atomistic models with *cis*-1,4 polyisoprene melts as an illustrative example. Through machine learning, we train conditional generative adversarial networks on the correspondence between configurations at the levels considered. The trained model is subsequently applied to provide predictions of atomistic structures from the input coarse-grained configurations. The effect of different data representation schemes on training and prediction quality is examined. Our proposed backmapping approach shows remarkable efficiency and transferability over different molecular weights in the melt based on training sets constructed from oligomeric compounds. We anticipate that this versatile backmapping approach can be readily extended to other complex systems to provide high-fidelity initial configurations with minimal human intervention.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0012320

Multiscale modeling is critical in diverse fields of research, including physics, chemistry, biology, and materials science and engineering.[1–4] Developing multiscale techniques in a hierarchical and consistent manner provides a comprehensive description of materials from bottom-up and, as a predictive tool, permits rational design of materials *in silico*. Coarse-graining of an atomistic/microscopic model simplifies the representation of a physical system by reducing the degrees of freedom and hence drastically accelerates calculations. In consequence, coarse-graining allows simulations to reach larger temporal and spatial scales that are inaccessible to detailed models retaining a fine level of description.

Systematic methods to construct coarse-grained (CG) models by passing information from fine to coarse levels continue to develop and improve over recent years.[5–9] This is particularly critical for macromolecular systems, whose structural properties and dynamical behavior span over a vast range of length and time scales.[10,11] Commercial grade polymeric materials consist of molecules with hundreds to thousands of repeat units that necessitate the simulation studies of large-scale systems, which evolve in exceedingly slow rates. As a result, systematic coarse-graining, based on atomistic simulations of small molecules, facilitates the quantitative interrogation of conformational and thermodynamic properties of industrial

polymers, such as polystyrene,[12,13] polyethylene,[14] polyisoprene,[15–17] and polybutadiene,[18–21] at molecular weights where atomistic modeling presents insurmountable demands even for state-of-the-art high performance computing resources.

For many applications, there is a need to study large systems with long correlation times and also fine details; examples include the studies of local structure and dynamics of long macromolecules at the interface with solid surfaces[15,22] and cross-linking reactions that depend on the atomistic-level configuration of specific repeat units.[23] In those cases, high-fidelity backmapping is essential for closing the loop starting from the atomistic model (AT) to a derived CG model and then going back to the detailed AT description. Although the accuracy of such a scheme is intimately coupled with the capability of the CG model to propose realistic structures, it is important to note that equally essential attributes are efficiency, flexibility, and broad applicability with minimal human intervention. Given the multitude of the already existent chemically specific CG models, a general backmapping technique, readily applicable with little modification that is not necessarily tied to the technique used to develop the CG force fields in the first place, is highly desirable. Several studies have focused on addressing this need with different strategies, based on random mapping, geometrical and mechanical considerations, Bayesian inference, hybrid scheme (using a time-dependent switching function), and subsequent position-restrained molecular dynamics (MD) or Monte Carlo relaxations.[13,15,24–34] However, those strategies more or less rely on maintaining libraries of molecular structures or force fields that are system specific, while often facing a delicate trade-off between efficiency and accuracy.

For polymers, CG models commonly preserve chemical specificity by promoting local configurations dictated by bond, bending angle, and dihedral angle potentials that are designed to match specific order parameters with atomistic structures. Through backmapping, these local configurations can be translated to fine structures absent in the CG configuration via sequential building of successive repeat units. Nevertheless, subsequent procedures are often needed to increase the quality of the resulting structures that are specific to the system at hand (as, for example, in our previous work with polyisoprene[15] and similar studies with polystyrene,[26] where dihedral angles in the retrieved atomistic models required further refinement). Such efforts take significant insight and algorithmic design, which are not generally applicable. In this work, we present a general method for backmapping polymer CG models, based on machine learning. While machine learning is currently being exploited in forward coarse-graining studies,[35–39] we reveal its versatility in the direct application of structural backmapping. Recently, a study on backmapping polystyrene oligomers constructed deep convolutional neural networks embedded with the force field description and utilized machine learning to predict atom placement sequentially through recurrent training.[40] By contrast, the proposed method in this work separates particle positions from the molecular topology and force field and hence avoids tedious bookkeeping of molecular details during the reconstruction while making predictions in groups. We expect that this method can be easily modified to apply to a broad class of polymeric materials given existing CG models. Noting that any backmapping procedure would result in "imperfections," our effort herein is to develop a systematic, model agnostic procedure that excels in flexibility and efficiency.

We formulated the backmapping from CG to atomistic as an image-to-image translation problem that translates "low-resolution" CG pictures to "high-resolution" atomistic pictures with fine details. This problem is essentially analogous to super-resolution,[41] while learning-based generative methods should be sought.[42] The rationale behind our approach is to use machine learning to capture complex statistical relationships between a CG model and the atomistic counterpart, in terms of probability distributions, through training with substantial samples. In the class of generative models, generative adversarial networks (GANs)[43] present a powerful deep learning approach, which simultaneously trains two types of artificial neural networks, namely, generator and discriminator, in a synergistic and opposing manner. Specifically, a generator network is trained to produce synthetic data following the same statistics as the training dataset, while a discriminator network learns to evaluate the authenticity of generated data as if they belong to the real dataset or not. Ideally, the quality of generated prediction improves through training without human intervention.

We employed modeling of *cis*-1,4 polyisoprene (PI) melts as an illustrative example and adopted a general-purpose conditional GAN model, "Pix2Pix,"[44] for the image-to-image translation. Pix2Pix captures the mapping between paired images through training on a corresponding dataset, predicts pixels from pixels, and synthesizes output images with conditioning on the input ones. Specifically, it utilizes a U-Net architecture[45] as a generator and a convolutional "patchGAN" classifier as a discriminator, which skips tweaking the loss function in machine learning and is found to be effective on a wide variety of image prediction problems.[44] In this study, the training sets were prepared based on atomistic configurations from MD simulation trajectories. The mapped CG configurations were employed as input, while the atomistic counterpart served as ground truth. After training the generative model, it was applied to data outside of the training sets to make backmapping predictions via the generator. The efficiency of this machine learning approach was examined, and the prediction quality was evaluated. We further compared the results between machine learning and a commonly used "random mapping" method, which generates configurations using random rotation of repeat units around their center-of-mass (COM) locations. The outcome demonstrates that the proposed image-based machine learning approach presents significant advantages by introducing correlations absent in the random mapping method, without sacrificing efficiency and general applicability.

A united-atom (UA) representation was selected for the atomistic model of PI, where each carbon with directly bonded hydrogens is lumped into a single interaction site. In this simplified atomistic representation, every PI monomer is defined by the positions of five particles [Fig. 1(a)]. The implemented UA force field has been extensively described in past studies, where it was shown to faithfully reproduce material properties.[46–49] Our CG mapping herein places a CG bead at the COM of each monomer [Fig. 1(a)]. The CG force field was developed using the Iterative Boltzmann Inversion (IBI) method,[15,50–52] a technique that has been widely used for systematic coarse-graining of atomistic polymer models. Detailed results and validation of the CG mapping are presented in Figs. S1 and S2 in the supplementary material.

A CG configuration comprises of coordinates of CG beads representing monomers for each polymer chain in the system
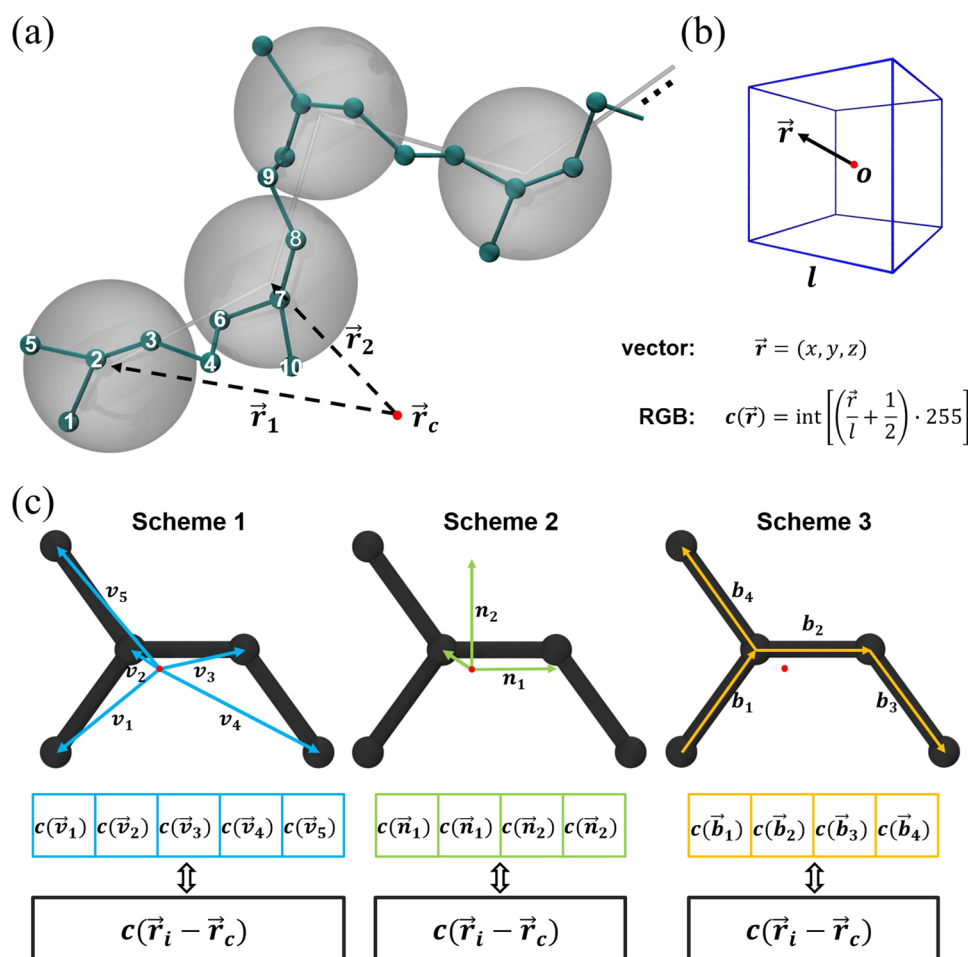
FIG. 1. (a) Schematic illustration of a tetramer section of the PI polymer in the two representations considered. In the UA model, each monomer contains five particles. Atom indices are included for reference to the schemes described in (c) and the potential employed in the supplementary material. In the CG model, beads are placed at the center-of-mass of each monomer. Successive CG position vectors describe the polymer chain contour. (b) Conversion to RGB color is performed via $c(\vec{r}) = \lfloor (\vec{r}/l + 0.5) \cdot 255 \rfloor$, normalizing each vector component (with its origin at the center of a reference unit cell) and rescaling values to integers in [0, 255] (cutoffs at 0 and 255 are applied). (c) Three types of monomer-based data representations used in machine learning: Scheme 1 takes relative positions of five united atoms with respect to the monomer COM (indicated by the red dot); Scheme 2 takes two orientation unit vectors of the reference plane defined by the monomer COM, atoms 2 and 3; Scheme 3 takes four bond vectors among connected pairs of atoms. The associated protocols of image translations are also illustrated. $\vec{r}_i$ and $\vec{r}_c$ refer to the COM of the ith monomer and the associated polymer chain, as shown in (a).

[Fig. 1(a)]. Backmapping involves generating atom coordinates, while molecular topology (describing the architecture and connectivity among atoms) can be extracted separately and treated as metadata that is not required for the coordinate generation. To achieve high-fidelity backmapping, the statistical relationships between CG and UA configurations need to be captured. For instance, the arrangement of atom positions in a polymer chain should follow the correlations among successive CG beads along the chain contour, as expressed by bond distances, angles, dihedrals, etc. Nevertheless, these relationships are mostly ignored in random mapping and can only be partially traced by the sequential building of polymers (via stochastic sampling and local optimization in a system-specific manner).[15,34] By contrast, machine learning, in principle, parses hidden representations of the full relationships and casts them into neural networks, regardless of the chemical species and force fields.

To highlight the correlations embedded in spatial configurations, we exploited an adaptive reference strategy for establishing the correspondence between CG and UA representations. Specifically, CG beads of monomers along the chain contour were referenced by vectors from the COM of a polymer fragment or entire chain [Fig. 1(a)], instead of the absolute positions in space. Meanwhile, UA particles were referenced by vectors that depict relative positions within an individual CG bead. As illustrated in Fig. 1(c), three UA reference schemes were considered: Scheme 1 takes five position vectors from the monomer COM to the united atoms; Scheme 2 treats the monomer as rigid body, where the relative positions of atoms are fixed, and the rotation of the rigid body can be described by two unit orientation vectors; Scheme 3 takes four bond vectors among pairs of atoms.

In Scheme 1, four position vectors are sufficient for determining particle coordinates inside the monomer, given the position constraint of the COM. However, all the five vectors were adopted to maximize the degrees of freedom for machine learning. For Scheme 2, a fixed unit monomer is required to construct approximate configurations using rotations; we selected the one prescribed by the preferred bonds and angles from the force field [as shown in Fig. 1(c), where all five particles lie in the same plane]. From the setup, there is no extra positional constraint among five atoms in Scheme 1 besides the COM position. Scheme 2 constrains both bonds and angles

inside the monomer, and the same bond constraints can be applied in Scheme 3 (i.e., bond lengths are fixed to 0.150 nm, 0.134 nm, 0.150 nm, and 0.150 nm for 1–2, 2–3, 3–4, and 2–5 bonds, respectively), which is the case in our UA simulations (constraining bonds allows for a larger time step in MD simulations). In general, these three schemes provide different data representations of the UA configuration. Their effect on the training and prediction of machine learning will be examined later.

To structure UA and CG data, spatial configurations were mapped into images. The translation was executed by converting XYZ components of vectors to RGB values through normalization with respect to adaptive reference cells and rescaling [Fig. 1(b)]. Each color channel, corresponding to an axis, takes integer values from 0 to 255. The spatial resolution of this mapping is thus dictated by $l/255$, with l being the side length of the reference cells, which is optimized for performance. Following the mapping schemes in Fig. 1(c), UA images, carrying different features of individual monomers, were obtained from mapping characteristic vectors (i.e., $\vec{v}_i$, $\vec{n}_i$, and $\vec{b}_i$) to pixels or color blocks. In particular, the relative position vectors in Scheme 1 were normalized to a reference box of $l = 0.46$ nm centered at the monomer COM (i.e., the CG bead location); the orientation and bond vectors in Schemes 2 and 3 were treated as unit vectors. CG color blocks, containing positional information of monomers, were designed to take the total size of the UA color blocks from the member vectors, which composed "low-resolution" CG images.

For a case study, we selected a PI melt system consisting of 100 chains of 30 monomers each (noted as 30-mer). MD simulations were performed at temperature $T = 413$ K and pressure $P = 1$ bar to generate system configurations (using Gromacs,[53,54] version 2019.3, see simulation details and Fig. S3 in the supplementary material). Typical chain conformations in UA and mapped CG representations are shown in Figs. 2(a) and 2(b), respectively. Image translation was performed per polymer chain. A reference

cell of $l = 7.5$ nm was adopted at the chain COM for the color mapping of CG beads (the average 30-mer end-to-end distance $R_{EE} \approx 3.28$ nm, see the supplementary material and Fig. S4 for detailed discussion). Figure 2(c) presents three images translated from the conformations via the mapping schemes in Fig. 1(c). Each image represents single training data, consisting of two concatenated 256×256 pixel matrices of RGB values from the UA and CG representations. The CG image is used as the input for machine learning, while the UA image is the target. The particular data arrangements displayed in those images were employed to maximize the use of information in training. The training sets were constructed accordingly from an atomistic MD trajectory (each frame corresponds to 100 images).

The implementation of Pix2Pix[44] was carried out in the open source TensorFlow 2 platform, and a single NVIDIA GeForce GTX 1080 GPU was employed for calculations. The efficiency of this machine learning mechanism in training was optimized, and the adjustment of relevant control parameters is presented in Figs. S5–S10 in the supplementary material. As the machine learning is based on images, the acquired parameter set should also exhibit good performance in other systems with similar data representations. A typical training process using Scheme 3 mapping is shown in Fig. 3(a), which is characterized by the evolution of defined loss functions along with the number of epochs. Specifically, generator loss contains cross entropy between generated images (namely, predictions) and a matrix of the same shape with value ones, plus a weighted L1 difference between predictions and targets (the effect of weight coefficient was tested, cf. Fig. S10 in the supplementary material); discriminator loss contains cross entropy between predictions and a matrix of zeros, plus cross entropy between targets and a matrix of ones.[44] The decay of total loss (i.e., sum of the generator and discriminator losses) as learning progresses indicates that the training improves the quality of predicted images from the generator (cf. Fig. S8).
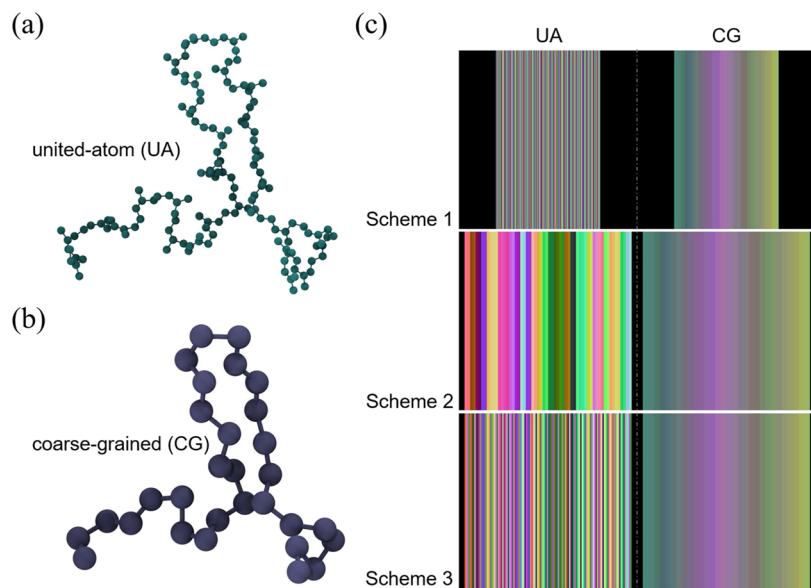

FIG. 2. (a) A 30-mer PI (i.e., 150 united atoms) UA chain configuration. (b) The same chain in the CG representation. (c) Corresponding images used in training for Schemes 1–3. Each image contains concatenated 256×256 pixel matrices of RGB values converted from UA and CG configurations, respectively. Zero-padding is applied in the conversion (as black margins), while the data body is centered. In Scheme 1, each position vector of a united atom $\vec{v}_i$ is depicted by a 256×1 unicolor stripe. In Scheme 2, each unit orientation vector $\hat{n}_i$ is depicted by a 256×4 unicolor stripe. In Scheme 3, each unit bond vector $\hat{b}_i$ is depicted by a 256×2 unicolor stripe. The color of the corresponding 256×5, 256×8, and 256×8 stripes for each CG bead is converted from the relative position of the monomer COM with respect to the chain COM. These data representations maximize the use of information in training.
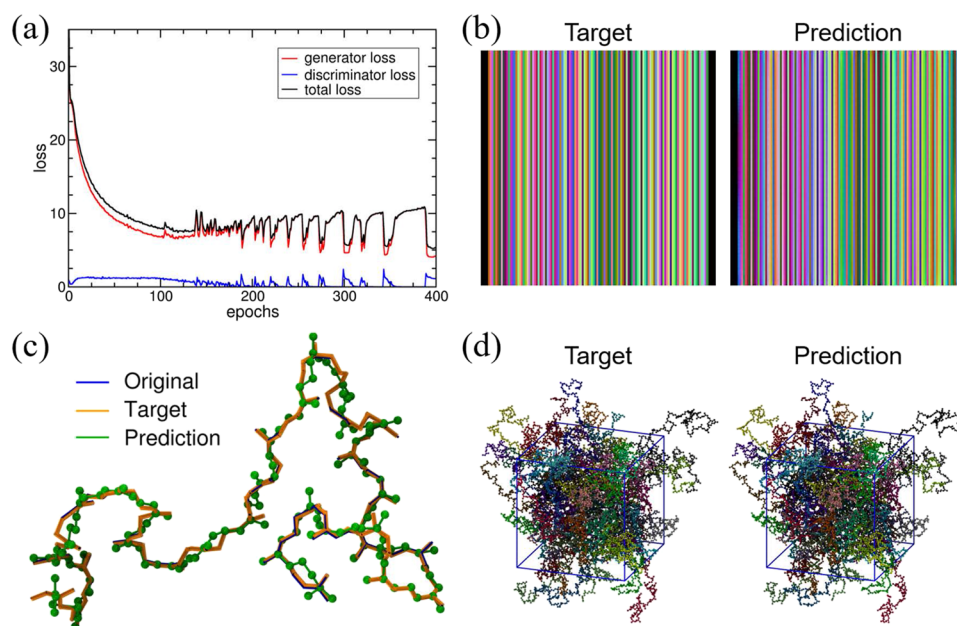
FIG. 3. (a) The evolution of loss functions along with epochs in the training of Scheme 3. Each epoch refers to a cycle through all the training data (and takes ~21.2 s in real time). (b) Comparison between a target image (i.e., ground truth) and predicted image from the trained model after 400 epochs. (c) Polymer chain configurations translated from the images in (b). The original chain configuration is shown for comparison. The marginal difference between the original and the target is caused by the rounding error of float to integer conversion in image translation. (d) Simulation cells of the UA model converted from target and predicted images (polymer chains are unwrapped in the periodic simulation box and colored differently for visualization).

Our machine learning approach was validated by testing configurations not included in the training set as input. Figure 3(b) displays an example of a predicted image from the trained model after 400 epochs, alongside the corresponding target image. Besides mild blurry near the margins, the prediction in general exhibits an outstanding visual similarity with the target. The averaged RGB values, $\vec{c}$, over individual color stripes are translated to bond vectors (i.e., $\vec{b}_i = \vec{c}_i/255 - 0.5$). Taking the CG bead coordinates from the pairing CG image, the atom coordinates with respect to the chain COM can thus be determined. Figure 3(c) shows the chain conformations converted from images in Fig. 3(b), with the original chain conformation also presented for comparison (the image translation results in a marginal difference between the target and original conformations due to float to integer conversion). One finds a good consistency between the prediction and the target or original. Meanwhile, the deviation of the prediction from the original contour exposes the inherent feature of the one-to-many correspondence from CG to UA configurations, as such a backmapping process generates information and is not deterministic in solutions. With the information of chain COM (recorded during the image translation), a UA configuration of the entire system can be handily constructed from assembling chains in space [see Fig. 3(d)].

Obtaining the trained models, the backmapping from CG to UA proceeds in three steps: convert a CG configuration to CG images following the image translation protocols used in the training set, input the images into machine learning for a prediction, and translate the predicted images to a UA configuration. The CG configuration can be mapped from atomistic trajectories or taken directly from CG simulations with no original atomistic counterpart. Due to adaptive reference, the predictions are made simultaneously for all monomers in the image (i.e., 30 under the current

setting as a typical choice). This offers a significant improvement over sequential building in efficiency. In addition, as the data arrangement in image translation is flexible, the trained models can directly apply to other PI melt systems of different size or chain length. Specifically, shorter chains can be treated by extra zero-padding in the input CG images; longer chains can be processed on the basis of a 30-mer or shorter fragments. We note that although the random mapping method (introducing repeat units based on the COM location and random orientation) inherently has the same feature (i.e., not limited to systems), it requires a system-specific tuning in practice and fails to capture intra-chain correlations, in contrast to the machine learning approach. As an example, we present a CG chain conformation of a 500-mer chain alongside a prediction from random mapping in Fig. 4(a). For comparison, backmapped UA configurations from machine learning are shown in Fig. 4(b) using different data representation schemes.

To quantify the prediction quality, we examined the distributions of all five types of bonds, five types of angles, and four types of dihedrals along the chain contour and computed the probability densities (p_i, i indicates individual distribution) for original, target, and prediction UA configurations, respectively (cf. Figs. S7 and S11–S13 in the supplementary material, where radial distribution functions are also shown). A comparison between the original and predicted UA configurations for a dihedral distribution, illustrating a typical intra-chain correlation, is summarized in Fig. 4(c). Random mapping does not carry any correlation among particles in different monomers. By contrast, certain degrees of correlations were captured through machine learning on the chain connectivity depicted by the color block sequence inside the images. As no constraint was applied for the connection between neighboring monomers, machine learning yielded a spread distribution of bond length for the connecting bonds and smeared distributions of the associated
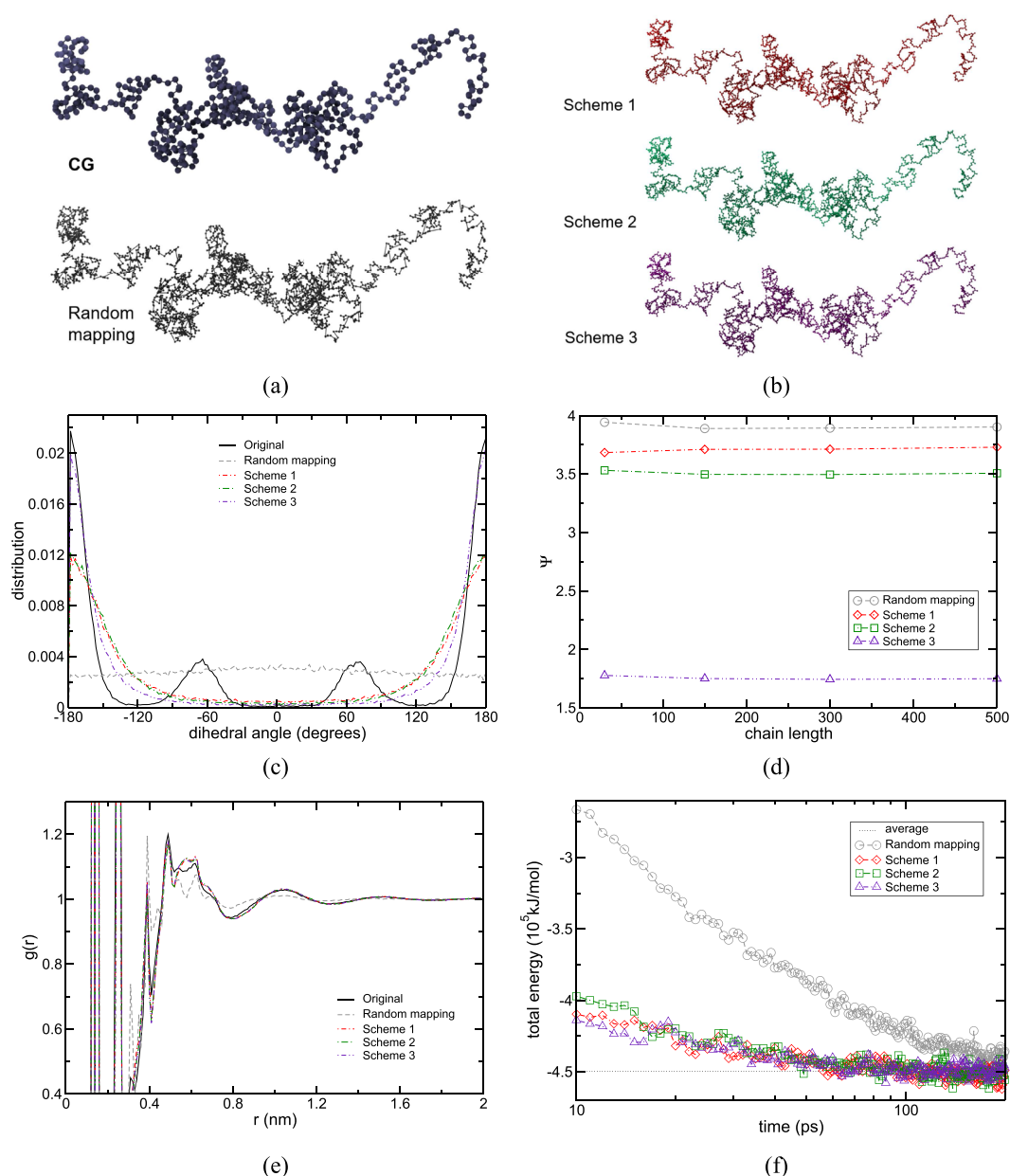
15 April 2024 05:25:08

FIG. 4. (a) A typical 500-mer CG conformation and backmapped UA configuration from random mapping. (b) Counterpart predictions from machine learning through different data representation schemes. (c) Distributions of dihedral angles defined by UA particles (3, 4, 6, 7) along the chain contour [cf. Fig. 1(a) for indices] from original and predicted 500-mer UA configurations. (d) Metric Ψ of constructed UA configurations for systems containing 100 30-mer, 125 150-mer, 135 300-mer, and 200 500-mer chains, respectively, using trained models based on a 30-mer dataset. (e) Comparison of the radial distribution function of UA particles between the 500-mer original and the obtained configurations after an energy minimization procedure on the predictions to remove steric overlapping (via a steepest descent algorithm with the maximum force tolerance of 10 000 kJ/(mol·nm) and an initial step size of 0.01 nm, which converged in a couple hundred of steps). (f) Temporal evolution of total energy in a subsequent MD simulation (in the NVT ensemble) for the prepared 500-mer UA systems. The dotted line indicates the converged value from a long MD simulation of the UA system.

dihedral angles. These variations enlarged the configurational space of predictions for backmapping.

We defined a metric, $\Psi = \sum_i \int |p_i^{\text{prediction}}(x) - p_i^{\text{original}}(x)| dx$, as the sum over L1 norm of the difference between p_i of prediction and

original for all the distributions (see Fig. S14 for the comparison of different dihedral distributions). Lower values of Ψ correspond to higher prediction quality. In Fig. 4(d), we plot Ψ of backmapped configurations from random mapping and machine learning for

systems of different chain lengths and numbers. Specifically, the backmapping from different schemes was performed using the same trained models based on 30-mer UA configurations. From the figure, one finds that all the schemes exhibit similar performance for different melt systems, demonstrating the transferability and efficiency of the machine learning approach using the chain/fragment-based adaptive reference strategy. Furthermore, machine learning shows much improved backmapping credibility in comparison to the random mapping. Scheme 3 provides the best predictions compared to Schemes 2 and 1, implying that data representation can influence the results of training. Comparatively, reasonable constraints in data representation simplify the statistical relationships that lead to better results.

To further validate the backmapping, we initiated atomistic MD simulations with predicted configurations and compared the trajectories with long UA simulations that present converged metrics of properties. As the adopted chain-based strategy does not fully take into account the local environment of chains, neighboring steric overlapping from different chains cannot be resolved. The resulting stiffness of interaction potentials can be relaxed via energy minimization, as is standard in backmapping methods.[15,24–30,34] In Fig. 4(e), we present the radial distribution function of UA particles in 500-mer configurations obtained after an energy minimization procedure on the predictions (see Fig. S15 for the inter-chain result). Particularly, the input CG configuration for backmapping was selected from a CG MD trajectory (rather than being mapped forwardly from a UA configuration). Figure 4(f) plots the temporal evolution of the system's total energy in subsequent MD simulations (in the NVT ensemble with initial velocities generated based on a Maxwell distribution). The energy evolution proceeds to an equilibrium value rapidly (in ~60 ps, which is close to the segmental relaxation time of PI at $T = 413$ K[15,49]). Upon reaching the equilibrium value, conformational properties recovered the original ones (see Fig. S16, for example). The quick convergence indicates that the predicted configurations from the trained models are close to the target one in the free energy landscape, as also implied by Fig. 4(e). In contrast, random mapping requires longer time to reach equilibrium, which can be attributed to kinetic trappings induced by steric constraints. From the comparison, we conclude that machine learning presents a significant improvement in capturing intra-chain correlations that provides faithful backmapping of long macromolecules in dense systems.

In summary, we have proposed an image-based approach for structural backmapping from coarse-grained to atomistic configurations. We note that the image mapping offers a straightforward and efficient representation to structure data of particle coordinates. Combined with machine learning tools, such as conditional generative adversarial networks, it provides an effective and reliable strategy for high throughput and high accuracy conversion. We demonstrated the success of the approach with a case study on *cis*-1,4 polyisoprene melt simulations using adaptive reference, which fully exhibits the generality and transferability of the strategy. Our approach separates configurational information from molecular topology and force field for backmapping, and it makes predictions only based on relative positions and hidden correlations. This underlying mechanism can be easily implemented and applied to other complex systems at different coarse-graining levels, such as heteropolymers and biomacromolecules, where

bookkeeping and tuning atomistic details via human intervention become particularly tedious and labor-intensive. Specifically, the mapping scheme and data representation can be modified to depict different molecular architectures and chemical species through the variation in color range, color block size, and spatial arrangement, while the optimized control parameters of the machine learning mechanism should be generally applicable. We expect that this backmapping approach further complements multiscale modeling techniques addressing material properties over a broad range of length and time scales.

See the supplementary material for the development of coarse-grained force field, simulation details, benchmark of parameters for machine learning, and additional results.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] E. Weinan, *Principles of Multiscale Modeling* (Cambridge University Press, 2011).

[2] M. F. Horstemeyer, *Practical Aspects of Computational Chemistry* (Springer, 2009), pp. 87–135.

[3] G. S. Ayton, W. G. Noid, and G. A. Voth, "Multiscale modeling of biomolecular systems: In serial and in parallel," Curr. Opin. Struct. Biol. **17**, 192–198 (2007).

[4] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, "Multiscale modeling of emergent materials: Biological and soft matter," Phys. Chem. Chem. Phys. **11**, 1869–1892 (2009).

[5] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt, "Systematic coarse-graining methods for soft matter simulations—A review," Soft Matter **9**, 2108–2119 (2013).

[6] S. Izvekov and G. A. Voth, "A multiscale coarse-graining method for biomolecular systems," J. Phys. Chem. B **109**, 2469–2473 (2005).

[7] 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**, 244114 (2008).

[8] 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**, 244115 (2008).

[9] Y. N. Pandey, G. J. Papakonstantopoulos, and M. Doxastakis, "Polymer/nanoparticle interactions: Bridging the gap," Macromolecules **46**, 5097–5106 (2013).

[10] J. Baschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müuller-Plathe, M. Murat, W. Paul *et al.*, *Viscoelasticity, Atomistic Models, Statistical Chemistry* (Springer, 2000), pp. 41–156.

[11] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Bürger, "Simulation of polymer melts. II. From coarse-grained models back to atomistic description," Acta Polym. **49**, 75–79 (1998).

[12] D. Fritz, V. A. Harmandaris, K. Kremer, and N. F. A. van der Vegt, "Coarse-grained polymer melts based on isolated atomistic chains: Simulation of polystyrene of different tacticities," Macromolecules **42**, 7579–7588 (2009).

[13]V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer, "Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations," Macromolecules **39**, 6708–6719 (2006).

[14]J. Zhao, S. Nagao, and Z. Zhang, "Thermomechanical properties dependence on chain length in bulk polyethylene: Coarse-grained molecular dynamics simulations," J. Mater. Res. **25**, 537–544 (2010).

[15]Y. N. Pandey, A. Brayton, C. Burkhart, G. J. Papakonstantopoulos, and M. Doxastakis, "Multiscale modeling of polyisoprene on graphite," J. Chem. Phys. **140**, 054908 (2014).

[16]T. Ohkuma and K. Kremer, "Comparison of two coarse-grained models of cis-polyisoprene with and without pressure correction," Polymer **130**, 88–101 (2017).

[17]N. Shahidi, A. Chazirakis, V. Harmandaris, and M. Doxastakis, "Coarse-graining of polyisoprene melts using inverse Monte Carlo and local density potentials," J. Chem. Phys. **152**, 124902 (2020).

[18]L. Yelash, M. Müller, W. Paul, and K. Binder, "How well can coarse-grained models of real polymers describe their structure? The case of polybutadiene," J. Chem. Theory Comput. **2**, 588–597 (2006).

[19]T. Strauch, L. Yelash, and W. Paul, "A coarse-graining procedure for polymer melts applied to 1,4-polybutadiene," Phys. Chem. Chem. Phys. **11**, 1942–1948 (2009).

[20]I. Y. Lyubimov and M. G. Guenza, "Theoretical reconstruction of realistic dynamics of highly coarse-grained cis-1,4-polybutadiene melts," J. Chem. Phys. **138**, 12A546 (2013).

[21]C. A. Lemarchand, M. Couty, and B. Rousseau, "Coarse-grained simulations of cis-and trans-polybutadiene: A bottom-up approach," J. Chem. Phys. **146**, 074904 (2017).

[22]K. Johnston and V. Harmandaris, "Hierarchical simulations of hybrid polymer–solid materials," Soft Matter **9**, 6696–6710 (2013).

[23]A. J. Brooks, W. Dai, M. L. O'Mara, A. Abankwa, Y. Chhabra, R. A. Pelekanos, O. Gardon, K. A. Tunny, K. M. Blucher, C. J. Morton et al., "Mechanism of activation of protein kinase JAK2 by the growth hormone receptor," Science **344**, 1249783 (2014).

[24]L.-J. Chen, H.-J. Qian, Z.-Y. Lu, Z.-S. Li, and C.-C. Sun, "An automatic coarse-graining and fine-graining simulation method: Application on polyethylene," J. Phys. Chem. B **110**, 24093–24100 (2006).

[25]G. Santangelo, A. Di Matteo, F. Müller-Plathe, and G. Milano, "From mesoscale back to atomistic models: A fast reverse-mapping procedure for vinyl polymer chains," J. Phys. Chem. B **111**, 2765–2773 (2007).

[26]T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Müller-Plathe, and G. Milano, "Coarse-grained and reverse-mapped united-atom simulations of long-chain atactic polystyrene melts: Structure, thermodynamic properties, chain conformation, and entanglements," Macromolecules **40**, 3876–3885 (2007).

[27]A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. De Vries, and S. J. Marrink, "Reconstruction of atomistic details from coarse-grained structures," J. Comput. Chem. **31**, 1333–1343 (2010).

[28]A. Ghanbari, M. C. Böhm, and F. Müller-Plathe, "A simple reverse mapping procedure for coarse-grained polymer models with rigid side groups," Macromolecules **44**, 5520–5526 (2011).

[29]T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, "Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models," J. Chem. Theory Comput. **10**, 676–690 (2014).

[30]L. E. Lombardi, M. A. Martí, and L. Capece, "CG2AA: Backmapping protein coarse-grained structures," Bioinformatics **32**, 1235–1237 (2016).

[31]J. Krajniak, S. Pandiyan, E. Nies, and G. Samaey, "Generic adaptive resolution method for reverse mapping of polymers from coarse-grained to atomistic descriptions," J. Chem. Theory Comput. **12**, 5549–5562 (2016).

[32]J. Krajniak, Z. Zhang, S. Pandiyan, E. Nies, and G. Samaey, "Reverse mapping method for complex polymer systems," J. Comput. Chem. **39**, 648–664 (2018).

[33]J. Peng, C. Yuan, R. Ma, and Z. Zhang, "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).

[34]G. Zhang, A. Chazirakis, V. A. Harmandaris, T. Stuehn, K. C. Daoulas, and K. Kremer, "Hierarchical modelling of polystyrene melts: From soft blobs to atomistic resolution," Soft Matter **15**, 289–302 (2019).

[35]L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, "DeePCG: Constructing coarse-grained models via deep neural networks," J. Chem. Phys. **149**, 034101 (2018).

[36]S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," Nat. Commun. **9**, 3887 (2018).

[37]J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. De Fabritiis, F. Noé, and C. Clementi, "Machine learning of coarse-grained molecular dynamics force fields," ACS Cent. Sci. **5**, 755–767 (2019).

[38]S. Hunkler, T. Lemke, C. Peter, and O. Kukharenko, "Back-mapping based sampling: Coarse grained free energy landscapes as a guideline for atomistic exploration," J. Chem. Phys. **151**, 154102 (2019).

[39]W. Wang and R. Gómez-Bombarelli, "Coarse-graining auto-encoders for molecular dynamics," npj Comput. Mater. **5**, 125 (2019).

[40]M. Stieffenhofer, M. Wand, and T. Bereau, "Adversarial Reverse Mapping of Equilibrated Condensed-Phase Molecular Structures," arXiv:2003.07753v1.

[41]S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," IEEE Signal Process. Mag. **20**, 21–36 (2003).

[42]W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," IEEE Trans. Multimedia **21**, 3106–3121 (2019).

[43]I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems (Curran Associates, 2014), pp. 2672–2680.

[44]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2017), pp. 1125–1134.

[45]O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer, 2015), pp. 234–241.

[46]M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, "Atomistic Monte Carlo simulation of cis-1,4 polyisoprene melts. I. Single temperature end-bridging Monte Carlo simulations," J. Chem. Phys. **115**, 11339–11351 (2001).

[47]M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, "Atomistic Monte Carlo simulation of cis-1,4 polyisoprene melts. II. Parallel tempering end-bridging Monte Carlo simulations," J. Chem. Phys. **115**, 11352–11361 (2001).

[48]V. A. Harmandaris, M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, "Detailed molecular dynamics simulation of the self-diffusion of n-alkane and cis-1,4 polyisoprene oligomer melts," J. Chem. Phys. **116**, 436–446 (2002).

[49]M. Doxastakis, D. N. Theodorou, G. Fytas, F. Kremer, R. Faller, and F. Müller-Plathe, "Chain and local dynamics of polyisoprene as probed by experiments and computer simulations," J. Chem. Phys. **119**, 6883–6894 (2003).

[50]A. K. Soper, "Empirical potential Monte Carlo simulation of fluid structure," Chem. Phys. **202**, 295–306 (1996).

[51]F. Müller-Plathe, "Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back," ChemPhysChem **3**, 754–769 (2002).

[52]D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," J. Comput. Chem. **24**, 1624–1636 (2003).

[53]B. Hess, C. Kutzner, D. van Der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," J. Chem. Theory Comput. **4**, 435–447 (2008).

[54]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015).