Structural Coarse-Graining via Multi-Objective Optimization with Differentiable Simulation

Zhenghao Wu*,† and Tianhang Zhou‡

†*Department of Chemistry, Xi'an Jiaotong-Liverpool University, Suzhou 215123, Jiangsu, P. R. China*

‡*College of Carbon Neutrality Future Technology, State Key Laboratory of Heavy Oil Processing, China University of Petroleum (Beijing), Beijing 102249, China*

E-mail: zhenghao.wu@xjtlu.edu.cn

Abstract

In the realm of multiscale molecular simulations, structure-based coarse graining is a prominent approach for creating efficient coarse-grained (CG) representations of soft matter systems such as polymers. This involves optimizing CG interactions by matching static correlation functions of corresponding degrees of freedom in all-atom (AA) models. Here, we present a versatile method, namely, differentiable coarse-graining (DiffCG), which combines multi-objective optimization and differentiable simulation. The DiffCG approach is capable of constructing robust CG models by iteratively optimizing effective potentials to simultaneously match multiple target properties. We demonstrate our approach by concurrently optimizing bonded and non-bonded potentials of a CG model of polystyrene (PS) melts. The resulting CG-PS model accurately reproduces both structural and thermodynamic properties of the AA counterpart. More importantly, leveraging the multi-objective optimization capability, we develop a precise and efficient CG model for PS melts that is transferable across a wide range of

temperatures, *i.e.*, from 400 to 600 K. It is achieved via optimizing a pairwise potential with non-linear temperature dependence in the CG model to simultaneously match target data from AA-MD simulations at multiple thermodynamic states. Our work showcases a promising route for developing accurate and transferable CG models of complex soft-matter systems through multi-objective optimization with differentiable simulation.

Introduction

Complementary to experimental methods, molecular simulations have emerged as powerful tools for studying the behaviors of complex soft-matter systems at the microscopic level. In spite of the fact that all-atom models can give direct, precise predictions of structure-property relations of intricate systems with abundant atomic details, a considerable amount of phenomena of interest in systems such as macromolecules (e.g., polymers and proteins) that usually take place at the mesoscopic level are still inaccessible through all-atom simulations. To overcome these limitations, coarse-grained (CG) models have been developed, where multiple atoms or molecules are represented by a single interacting entity, reducing computational cost while still capturing essential features of the system.[1,2] Using the CG models, molecular systems with extensive spatio-temporal scales are allowed to explore, facilitating investigations of phenomena, e.g., long-time dynamics,[3–6] interfacial/phase behaviors,[7,8] self-assembly,[9–11] etc.

Typically, coarse-graining methods can be divided into two categories: top-down and bottom-up approaches.[12,13] In the top-down approach such as Martini,[14] relatively simple functional forms with a fixed number of ajustable parameters are selected for parameterization of CG models to match the experimentally measured macroscropic properties. However, it is uncertain whether the top-down CG models can accurately reflect the underlying microscopic physics due to its lack of a rigorous CG mapping from atomic degrees of freedom. In the bottom-up parameterization scheme, usually more complicated potentials are pa-

rameterized with atomistically detailed information. The bottom-up CG method attempts to capture the molecular details and reproduce the microscopic behaviors of the mapped atomistic model. There are several promising algorithms to develop systematic bottom-up coarse-grained models, including multiscale coarse-graining (matching effective force residuals),[15] relative entropy (matching information loss),[16] and Iterative Boltzmann inversion[17] and Inverse Monte-Carlo (matching structural properties).[18]

Although bottom-up approaches hold promise in the development of faithful CG models, there are still several significant challenges that remain unresolved.[12] The first pertains to the *transferability* of the CG models across varying environments such as (thermodynamic) state points and compositions. Another challenge revolves around ensuring the *consistency* and accurate *representation* including structural, dynamical, and thermodynamical properties. Numerous efforts have been dedicated to improving bottom-up CG modeling. The problem of transferability within such models often arises due to inadequate training data to parameterize the selected CG model. To address this limitation, for instance, one viable strategy involves the utilization of the "extended ensemble" variational principle.[19] This method attempts to determine a set of transferable potentials that optimally describe CG interactions for multiple systems across multiple state points under the multiscale coarse-graining framework. A similar idea that leverages information from multiple state points was also developed in the IBI method.[20] A single set of potential functions, i.e., a force field, that provides an optimal compromise for modeling a range of systems and thermodynamic state points. More recent advances in improving the transferability of coarse-grained models include the dual-potential method by Noid et al.,[21,22] ultra-coarse-graining with semi-global density potentials by Jin, Yu, and Voth,[23,24] microcononical relative entropy by Pretti and Shell,[25] and a recent extension of IBI by Qian and co-workers.[26]

Aside from the transferability issue, the representability of CG models in comparison to their corresponding AA models remains a persisting concern. One of the most common examples is the trade-off between accurately reproducing pairwise correlations and a dramatic

overestimation of pressure. This issue is evident in methodologies like the IBI approach.[17] To mitigate the pressure overestimation predicament, the IBI method employs a technique involving linear rescaling within pairwise interactions.[17] This rescaling is iteratively refined to align with the desired pressure target. However, a drawback of this approach is that, while it aims to improve pressure prediction, it cannot explicitly guarantee concurrent matching of key structural attributes, such as radial distribution functions (RDFs). Consequently, this approach could potentially lead to a reduction in the accuracy of structural predictions.

Recently, machine learning has emerged as a unique and powerful tool for the development of bottom-up coarse-graining methods.[27–29] It has shown remarkable potential in improving essential components of coarse-graining methods, including determining optimal coarse-grained mapping schemes,[30,31] facilitating back-mapping processes,[32–34] refining coarse-grained force fields, and optimizing associated parameters.[27,35,36] Moreover, an increasingly noteworthy technique garnering attention is the differentiable simulations (DiffSim). DiffSim built on automatic differentiation has been utilized to accelerate quantum calculations, generate protein structures, and computational fluid dynamics. In molecular simulations, Greener et al. used DiffSim to learn parameters of simple force-field functionals for coarse-grained simulations of small proteins.[37] Gómez-Bombarelli and co-workers improved the memory efficiency of DiffSim using the adjoint ordinary differentiation equation.[38] This augmented DiffSim is successfully applied to learn pairwise neural network potentials in simple condensed systems[39] and free energy barriers in systems with rare events.[40] Although it is significant for learning dynamics and non-equilibrium behaviors, differentiating through a (long) simulation trajectory may generate unreasonable gradients due to the chaotic nature of the dynamical system of molecular dynamics.[41] To alleviate this issue, Differentiable Trajectory Reweighting (DiffTre) provides a promising alternative that bypasses differentiation through MD simulation for time-independent observables via combinations of automatic differentiation with statistical reweighting.[42]

The utilization of DiffSim has proven to be remarkably successful in developing CG mod-

els of simple liquids in condensed phase.[42] Building upon this foundation, we introduce a novel structural coarse-graining approach, namely, differentiable coarse-graining (DiffCG). This novel method extends the capabilities of DiffSim to coarse-graining complex systems in condensed phases, such as polymer melts. To alleviate the instability of gradients from differentiating through a long trajectory in DiffSim, DiffTre is employed to obtain stable gradients of CG potentials.[42] Notably, the extension to more complex condensed-phase systems is achieved through the integration of multi-objective optimization techniques. Specifically, three multi-objective optimization algorithms are selected, which are popular in machine learning community. These algorithms are meticulously benchmarked using a representative system, i.e., mono-disperse polystyrene (PS) melts, which has been widely used for the evaluation of coarse-grained models.[43,44] The conventional CG force fields (bond, angle, dihedral and pairwise interactions) are employed to describe CG-PS system and they are optimized to simultaneously match multiple target properties, encompassing the corresponding structural correlations and pressure characteristics. More importantly, leveraging the capability of multi-objective optimization in DiffCG, we have developed a temperature-transferable CG model of PS melts using (non-linear) temperature-dependent potentials. The effectiveness of DiffCG in developing CG models of PS melts is comprehensively examined through a systematic evaluation in terms of the reproduction of target properties, transferability across a wide range of temperatures and molecular weights, and the optimization efficiency.

Methodology

Structure-Based Coarse-Graining

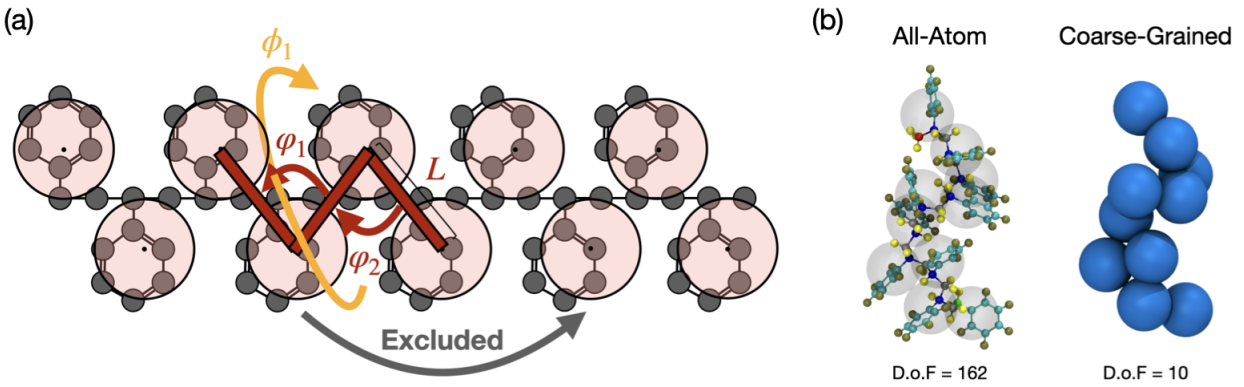

Figure 1: (a) Schematics of a coarse-grained polystyrene chain for the illustration of separations in bonded interactions: bond length L, bond angle φ, and dihedral angle ϕ; (b) Coarse-grained mapping between all-atom model and coarse-grained model of a ten-monomer polystyrene chain.

The essential concept underlying the structure-based coarse-graining methods, e.g., IBI and IMC, is the Boltzmann inversion, where independent degrees of freedom $\mathbf{q}$ in a canonical ensemble obey the Boltzmann distribution,

$$P(\mathbf{q}) = Z^{-1} \ln \int \exp[\beta U(\mathbf{q})] \tag{1}$$

where $Z = \int \exp[-\beta U(q)]\mathrm{d}q$ is a partition function of the system with respect to $\mathbf{q}$; $\beta = 1/k_B T$, k_B is Boltzmann constant and T denotes temperature. The primary objective of structure-based coarse-graining lies in approximating the (many-body) potential of mean force (PMF) with effective potentials that reproduce a set of reference distribution functions from AA models. Hence, deriving the effective CG potentials can be considered as solving an inverse problem: The forward function from the potential $U(\mathbf{q})$ to the probability distribution functions $P(\mathbf{q})$ is well established (through MD simulation), while the opposite direction, from $P(\mathbf{q})$ to $U(\mathbf{q})$, is to be determined. Furthermore, this inverse problem can be treated as an (iterative) optimization problem: (iteratively) optimize the effective potentials

until it reproduces all target observables (e.g., distribution functions) through the CG-MD simulation. It should be noted that the Boltzmann-inverted potential serves as an exact solution only for non-interacting systems, such as ideal gases, but it often fails to reproduce the reference $g^{\mathrm{ref}}(r)$ for condensed-phase systems due to strong many-body correlations.[17] Therefore, optimization is necessary to improve the initial Boltzmann-inverted potential.

It is common to assume that the bonded and nonbonded interactions are uncorrelated for molecular systems, giving:

$$U^{\mathrm{CG}} = U^{\mathrm{CG}}_{\mathrm{bonded}} + U^{\mathrm{CG}}_{\mathrm{nonbonded}} \tag{2}$$

As shown in Figure 1 (a), the bonded potentials $U^{\mathrm{CG}}_{\mathrm{bonded}}$ can usually be further decomposed into three terms: $U^{\mathrm{CG}}_{\mathrm{bonded}} = U^{\mathrm{CG}}_{\mathrm{bond}} + U^{\mathrm{CG}}_{\mathrm{angle}} + U^{\mathrm{CG}}_{\mathrm{dihedral}}$. The bond ($U^{\mathrm{CG}}_{\mathrm{bond}}$), angle ($U^{\mathrm{CG}}_{\mathrm{angle}}$), and dihedral ($U^{\mathrm{CG}}_{\mathrm{dihedral}}$) potentials describe the two-body interactions between connected beads, three-body interactions between the bond angles, and four-body interactions between two intersecting planes, respectively. The initial CG potentials can be derived from the corresponding distribution functions according to Boltzmann inversion (Equation 1):

$$U^{\mathrm{CG}}_{\mathrm{bond}}(l) = -k_{\mathrm{B}}T \ln\left[P^{\mathrm{CG}}(l)/4\pi l^2\right] \tag{3}$$

$$U^{\mathrm{CG}}_{\mathrm{angle}}(\theta) = -k_{\mathrm{B}}T \ln\left[P^{\mathrm{CG}}(\theta)/\sin(\theta)\right] \tag{4}$$

$$U^{\mathrm{CG}}_{\mathrm{dihedral}}(\varphi) = -k_{\mathrm{B}}T \ln\left[P^{\mathrm{CG}}(\varphi)\right] \tag{5}$$

where $P^{\mathrm{CG}}(l)$, $P^{\mathrm{CG}}(\theta)$, and $P^{\mathrm{CG}}(\varphi)$ are distribution functions of bond length l, bond angles θ, and dihedral angle φ, respectively; $4\pi l^2$ and $1/\sin(\theta)$ are rescaling factors in order to obtain the volume normalized distribution functions.

The approximation of the additive pairwise potentials is employed to describe the nonbonded interactions. It is noted that while higher body-order terms in the nonbonded potential might be significant for accurate CG models,[45–48] the investigation in this direction falls

outside the scope of the current work. The initial trial pairwise potential for the nonbonded interaction can be obtained through the Boltzmann inversion:

$$U^{\mathrm{CG}}_{\mathrm{nonbonded}}(r) = -k_{\mathrm{B}}T \ln g^{\mathrm{ref}}(r) \tag{6}$$

where $g^{\mathrm{ref}}(r)$ denotes the radial distribution function from the reference fine-grained simulation. Effective potentials derived from structure-based coarse-graining often show (thermodynamic) state dependence, e.g., temperature,[25,49] composition,[50] etc, even after the optimization. Additional constraints can be introduced to the coarse-graining process to improve the transferability and representability of the structure-based CG model.[51–53] In structural coarse-graining, we believe that achieving an accurate CG model can be transformed into a problem of efficient multi-objective optimization with constraints.

Multi-Objective Optimization with Differentiable Simulation

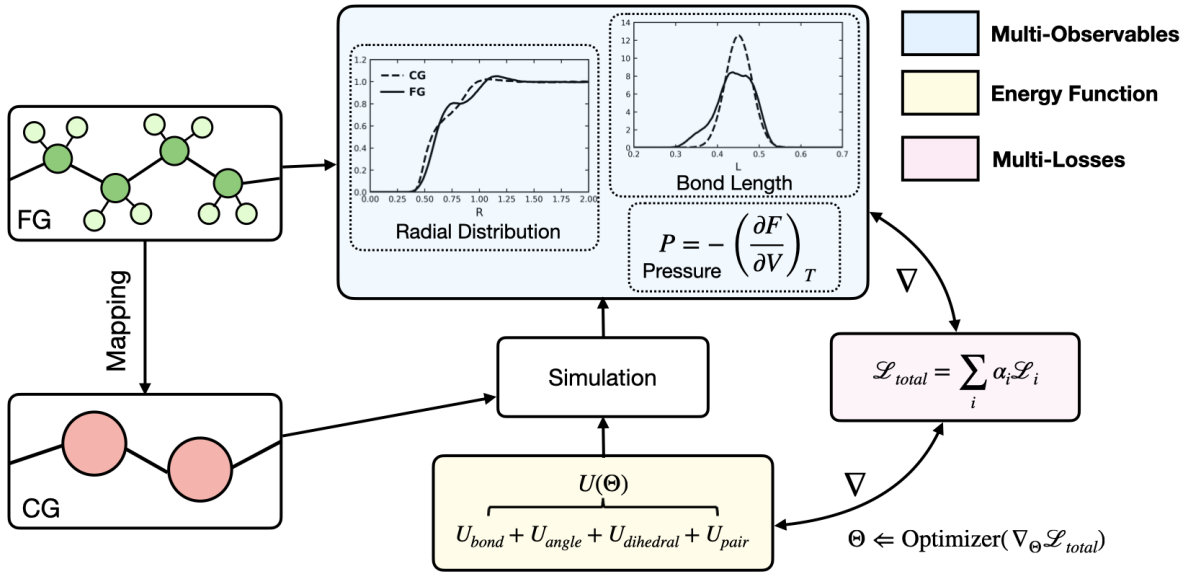

Figure 2: A general workflow for the differentiable coarse-graining: mapping between fine-grained (FG) and coarse-grained (CG) systems; simulation sampling; potential energy functions $U(\Theta)$; multiple observables (e.g., radial distribution functions, bond length distributions, pressure, etc.); multi-objective loss $\mathcal{L}_{total} = \sum_i \alpha_i \mathcal{L}_i$; and ∇: gradients from automatic differentiation.

In this section, we begin with a brief introduction to DiffSim, explaining its fundamental principles and methodologies for optimizing CG potentials. Following that, we delve into the concept of multi-objective optimization algorithms, discussing how they can be integrated into DiffSim for faithful structure-based bottom-up coarse-graining.

DiffSim uses automatic differentiation techniques that automatically calculate gradients by differentiating through the simulator, allowing gradient-based optimization using "back-propagation" to optimize simulation inputs such as force field parameters.[38] Gradient-based optimization requires the computation of reliable gradients. To overcome the issue of instability of gradient computation through the MD solver, we employ the importance reweighting technique as DiffTre[42] to improve the stability of gradients. DiffTre basically creates a proxy function using the thermodynamic perturbation theorem to bypass the differentiation through the entire trajectory for time-independent observables. Given that the structure-based coarse-graining method only addresses static properties as targets, this method is inherently well-suited for such applications.

As discussed in the previous section, the goal in structure-base coarse-graining is to optimize potential energy functions of the CG model to minimize a loss function as follows:

$$\mathcal{L}_\theta = \sum_{i \in \{P(l), P(\theta) \ldots\}} \mathcal{L}_{\theta,i}(\langle O_{CG} \rangle, \langle O_{AA} \rangle) \tag{7}$$

where $\langle \cdot \rangle$ is the ensemble averaging operator; O_{CG} and O_{AA} are microscopic properties/observables from CG simulations and AA reference simulations, respectively; $\sum(\cdot)$ represents the summation of various observables such as probability distributions of bond length, bond angles, and pressure. DiffTre constructs a well-behaved proxy objective for differentiation of static properties, utilizing the statistical reweighting based on the Bolzmann probability distribution:[42]

$$\langle O(R, U_\theta) \rangle = \sum_{i=1}^{N} w_i O(R, U_\theta) \tag{8}$$

$$w_i = \frac{e^{-\beta\left(U_{\lambda,\theta}(r_i) - U_{\lambda,\hat{\theta}}(r_i)\right)}}{\sum_{j=1}^{K} e^{-\beta\left(U_{\lambda,\theta}(r_j) - U_{\lambda,\hat{\theta}}(r_j)\right)}} \tag{9}$$

where w_i denotes thermodynamic weights where $U_{\lambda,\theta}$ represents the potential energy calculated with the parameters to be updated; $U_{\lambda,\hat{\theta}}$ represents the potential energy calculated with the reference parameters that generate the trajectory. More details of DiffTre can be found in the reference.[42]

As depicted in Equation 7, the loss function comprises a summation of various loss functions associated with the target observables. Consequently, the utilization of a multi-objective optimization algorithm becomes imperative. In numerical optimization, a natural concept to combine losses from multiple objectives is to perform a weighted linear summation of the losses, as shown below:

$$\mathcal{L}_{total} = \sum_i w_i L_i \tag{10}$$

where w_i and L_i are the weight coefficient and loss function of objective i, respectively. Weight w_i can be an additional optimizable parameter or can be determined self-consistently during the optimization process. Searching for these optimal weightings is usually expensive and increasingly difficult for large models with numerous objectives. Although balancing the loss contribution for each objective is still an open research area in multi-objective optimization/learning, several attempts have been made towards this direction. Here we test three different multi-objective optimization strategies that are popular in the machine learning community, namely, *Uncertainty Weighting*[54], *Auxiliary Multi-Task*[55] (a revised version of Uncertainty Weighting), and *Coefficient-of-Variations Weighting*[56], integrated with DiffSim for structure-based coarse-graining. The *uniform weighting* scheme in which the weights are kept at $w = 1$ throughout the optimization is used as a baseline for comparison. A list of different weighting schemes considered in this work with their definitions and characteristics is summarized in Table 1.

Table 1: A summary of different weighting schemes considered in this work with definition and main characteristics of the weight coefficient w

Algorithm	Definition of w	Main characteristics
Uniform Weighting	1	Constant
Uncertainty Weighting	$\frac{1}{2\sigma_i^2} + \frac{\log \sigma_i}{\mathcal{L}_i}$	Jointly optimized, with σ_i as variable; loss may become negative
Auxiliary Multi-Task	$\frac{1}{2\sigma_i^2} + \frac{\log(1+\sigma_i^2)}{\mathcal{L}_i}$	Jointly optimized, with σ_i as variable
CoV Weighting	$\frac{\sigma_{l_i}}{\mu_{l_i}}$	Observed, without additional variable

Uncertainty Weighting

Uncertainty weighting is a prominent method used in multitask machine learning.[54] Multi-task learning inherently entails a multi-objective nature due to potential conflicts between distinct tasks. Tuning these weights by hand proves to be difficult and resource-intensive, making multi-task learning prohibitive in practice. In particular, this concept resembles the objective loss function outlined in Equation 7. Therefore, we are evaluating the efficacy of this method in optimizing the CG model. The uncertainty weighting learns to numerically balance a linear combination of loss functions tied to individual tasks, considering the homoscedastic uncertainty of each task. The multi-task learning/optimization is treated as a probabilistic model,[54] and a set of learnable parameters is used to signify the task-specific uncertainty. The combination of multi-objective loss functions can be written as:

$$\mathcal{L}_{total} = \sum_i \frac{\mathcal{L}_i}{2\sigma_i^2} + \log \sigma_i \tag{11}$$

where $\mathcal{L}_i$ is the loss of individual objective function; σ_i is an additional learnable parameter that regulates the relative weights across multiple objective functions.

Auxiliary Multi-Task

This multi-objective optimization algorithm is a revised version of the uncertainty weighting method from previous work of auxiliary tasks in multi-task learning.[55] This weighting scheme replaces the regularization term $\log \sigma_i$ in the uncertainty weighting (Equation 11) with $\log(1+$

σ_i^2) to avoid negative loss values.

$$\mathcal{L}_{total} = \sum_i \frac{\mathcal{L}_i}{2\sigma_i^2} + \log(1 + \sigma_i^2) \tag{12}$$

This reformalism of the uncertainty weighting has demonstrated enhanced performance compared to the uncertainty weighting in the task of scene geometry and semantics. The performance of the algorithm might be task dependent; thus we also included this reformulated algorithm into our benchmarks for the purpose of comprehensive evaluation.

Coefficient-of-Variations Weighting

The Coefficient-of-Variations (CoV) weighting scheme is also designed for optimally learning machine learning tasks where the objective function is a weighted linear combination of multiple losses.[56] In brief, this weighting scheme uses the statistical properties inherent in loss functions to explicitly determine their relative importance. This is achieved using the CoV, which quantifies the relationship between the standard deviation (σ) and the mean (μ). Essentially, CoV signifies the degree of variability exhibited by the observed losses relative to their mean. The expression of CoV for the individual loss function L is given by:

$$c_L = \frac{\sigma_L}{\mu_L} \tag{13}$$

To facilitate a standardized comparison of various loss functions, the CoV is computed based on the observed ratio between the current observation and the mean of that particular loss, rather than the direct loss value itself. The loss-ratio l employed as a measurement, instead of the loss value itself, is defined as:

$$l_t = \frac{\mathcal{L}_t}{\mu_{\mathcal{L}_{t-1}}} \tag{14}$$

where $\mathcal{L}_t$ is the observed loss value at iteration step t; $\mu_{\mathcal{L}_{t-1}}$ is the mean of the observed losses up to iteration step $t-1$. Notably, it is essential that the observations are derived

from a ratio-scale to ensure accurate computation of the CoV. This approach enables an equitable comparison of uncertainty across a diverse series of measurements, even when their magnitudes differ.

The weight w_i of individual loss function L_i (Equation 7) based on the CoV of loss-ratios $c_{l_{it}}$ at time step t is given by:

$$w_{it} = \frac{1}{z_t} c_{l_{it}} = \frac{1}{z_t} \frac{\sigma_{l_{it}}}{\mu_{l_{it}}} \tag{15}$$

Here, z_t denotes a normalization constant that remains unaffected by the specific objective i: $z_t = \sum_i c_{l_{it}}$. This formulation guarantees $\sum_i c_{l_{it}} = 1$, a crucial aspect to disentangle the loss weighting from the learning rate. Obviously, the weight coefficient w_i in the CoV approach does not require additional parameters during optimization, making it different from the probabilistic model as uncertainty weighting.

Simulation Details

Fine-Grained Simulations

All-atom simulations of isotactic polystyrene (PS) melts are performed in LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator).[57] The OPLS (Optimized Potentials for Liquid Simulations) force field is employed to describe the potential energies of the system.[58] The initial configuration prepared by AutoPoly (https://github.com/Chenghao-Wu/AutoPoly) consists of 50 chains, each with 10 monomers. Periodic boundary conditions are applied in the X, Y, and Z directions to simulate the bulk properties of PS melts. The equation of motion is integrated based on a velocity-verlot algorithm with a time step of 1 fs. We first equilibrate the systems for 10 ns at a constant pressure $P = 1$ atm and a constant temperature $T = 600$ K. The Nóse-Hoover method is employed for both the barostat and the thermostat with a coupling constant $\tau_P = 1$ ps and $\tau_T = 0.1$ ps, respectively. Subsequently, we quench the system with a cooling rate 1 K/ps to low temperature. Further equilibration with 10 ns runs under the NPT ensemble is taken to ensure that the density of the system

converges to a plateau value. The production run of the system at each temperature is carried out under the NVT ensemble for calculations of various observables. Simulation observables, such as radial distribution functions and pressure, are calculated from the production runs. All other parameters are the same as in the equilibration step.

Coarse-Grained Simulations

The mapping scheme of the PS CG model is shown in Figure 1 (b), where a CG bead represents one monomer. The CG bead is centered on the corresponding centers of mass of the monomer in the all-atom model. All CG simulations are performed using the differentiable and hardware-accelerated molecular simulator based on JAX[59] (JAX-MD, version 0.1.28[60]), enabling fast running on modern computing accelerators (e.g., graphical computing unit/ tensor computing unit). All forces are computed via automatic differentiation on potential energy functions in JAX. It is noted that the derived tabulated potentials can also be used in LAMMPS. Same as in the AA simulations, Nosé–Hoover thermostat is used to regulate the temperature for all simulated systems. The initial configuration of the CG simulation is taken from the last snapshot of the production run in AA simulations to accelerate the equilibration of CG-MD simulations. The coarse-grained potential energy surface allows for a longer integration time step. We select time step $\Delta t = 4$ fs, which is consistent with previous CG-PS models.[44,61] Production runs with 1 ns sampling of CG-MD simulations are used to evaluate the performance of the final optimized CG-PS models.

Details of Differentiable Coarse-Graining

The workflow of the DiffCG is shown in Figure 2. All potential energy functions are represented by tabulated forms using cubic spline functions. As discussed in Section **Structure-Based Coarse-Graining**, the initial trial parameters are obtained via the direct Bolztamnn Inversion method to facilitate convergence of optimization. The optimization is carried out over 300 iterations using the stochastic gradient descent optimizer, i.e., Adam, implemented

in Optax,[59] with learning rate 0.003. The coarse-graining optimization is finished when it reaches the 300-iteration limit. This number is selected so that the total loss remains constant within a specified error range, and further optimization is unlikely to yield significant improvements. The initial weighting factors for the Uncertainty Weighting (Uncertainty) and Auxiliry Multitask Learning (Auxiliry) are 1 for each objective, while the Coefficient-of-Variantions Weighting (CoV) does not need an extra parameter. In the construction of the temperature-transferable CG model, we keep the bonded potentials (from $T = 600$ K) same for all CG-PS models and only optimize the pairwise potentials during the coasre-graining optimization process at target temperatures $T = 400, 500, 600$ K. RDFs and pressure from the AA-MD simulations at respective temperatures are used as target properties for DiffCG optimization. More details about the differentiable coarse-graining optimization can be found in Supporting Information.

Results

Comparison of Multi-Objective Optimization Algorithms for Structural Coarse-Graining

First, we examine the effectiveness of DiffCG with various multi-objective optimization algorithms in the construction of CG models of PS melts at a single thermodynamic state, i.e., temperature $T = 600$ K and pressure $P = 1$ atm. The total mean square error (MSE) defined as: $E_{total} = \epsilon_{bond} + \epsilon_{angle} + \epsilon_{dihedral} + \epsilon_{RDF} + \epsilon_{pressure}$ as a function of iterations during the optimization process using various multi-objective optimization algorithms is shown in Figure 3. Among all these algorithms, the CoV algorithm achieves the best performance in terms of minimizing E_{total}. Surprisingly, the Auxiliary method even behaves worse than the uniform one, although its error minimization is stable than others. Figure 3 (b) shows the decomposition of E_{total} of individual loss functions. Specifically, the CoV method achieves a balance of low MSEs in all these individual optimization objectives, in-

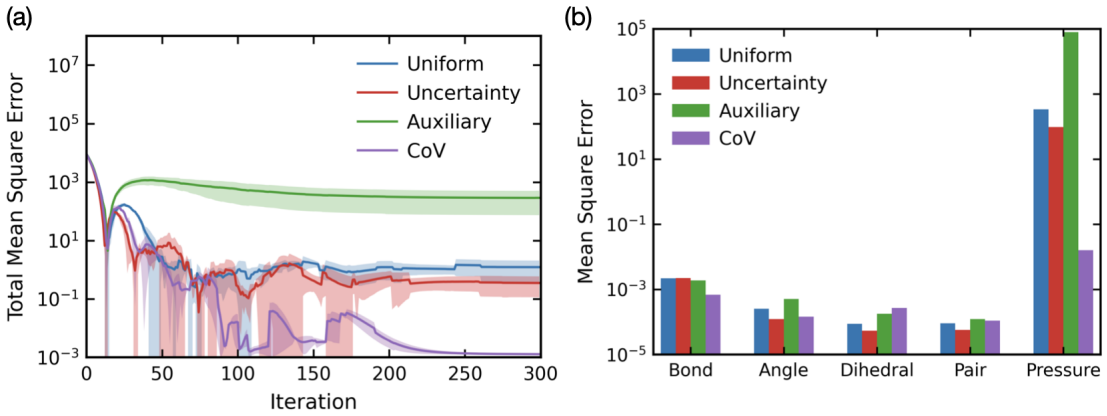

Figure 3: (a) A comparison of total mean square error as a function of the optimization iteration from DiffCG with various multi-objective optimization algorithms: uniform weighting (blue), uncertainty weighting (red), auxiliary multi-task (green), and covariance-of-variation weighting (purple). The shaded areas are corresponded to standard deviations of three independent optimizations with different random seeds; (b) The decomposition of total mean square error into individual sources or errors: probability distrubution functions of bond length, bond angle, dihedral angle, pair/radial distribution functions, and pressure.

dicating its advantages over other multi-objective algorithms in structural coarse-graining. The thermodynamic constraint (pressure) exhibits a high degree of sensitivity to the choice of multi-objective optimization algorithms. Among all the coarse-graining optimizations performed, the primary source of MSE is attributed to pressure, with errors stemming from the probability distributions of bond, angle, dihedral, and radial distribution function (RDF) being comparatively minor.

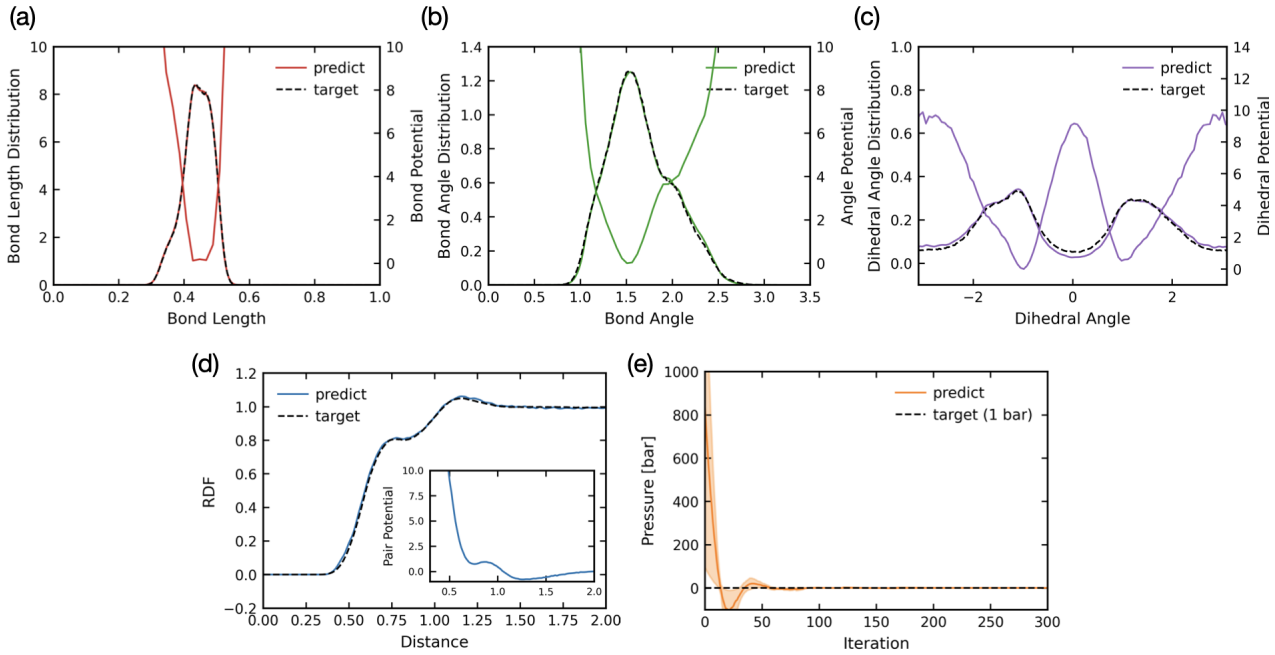

Figure 4: (a) Bond length distribution functions; (b) bond angle distribution functions; (c) dihedral angle distribution functions; (d) radial distribution function; and (d) pressure as a function of DiffCG optimization iterations.

Figure 4 (a)-(d) display the probability distribution functions of the bond length, bond angle, dihedral angles, and radial distribution functions obtained from a 1 ns CG MD simulation of PS melts with parameters from the final iteration of the CoV-based DiffCG. The apparent overlaps indicate that the CoV weighting algorithm with DiffSim is capable of optimizing the CG model to simultaneously match multiple target observables including (a) bond length, (b) bond angle, (c) dihedral, and (d) RDF. Furthermore, the thermodynamic quantity, i.e., pressure, as a function of optimization iterations is depicted in Figure 4 (e). The initial CG potential exhibits a notable tendency to overestimate pressure, similar to other structural coarse-graining methods. Through a few optimization iterations, this discrepancy decreases significantly. Typically, within approximately 100 optimization iterations, the potential stabilizes to the target values.

Trade-off between Sampling and Accuracy

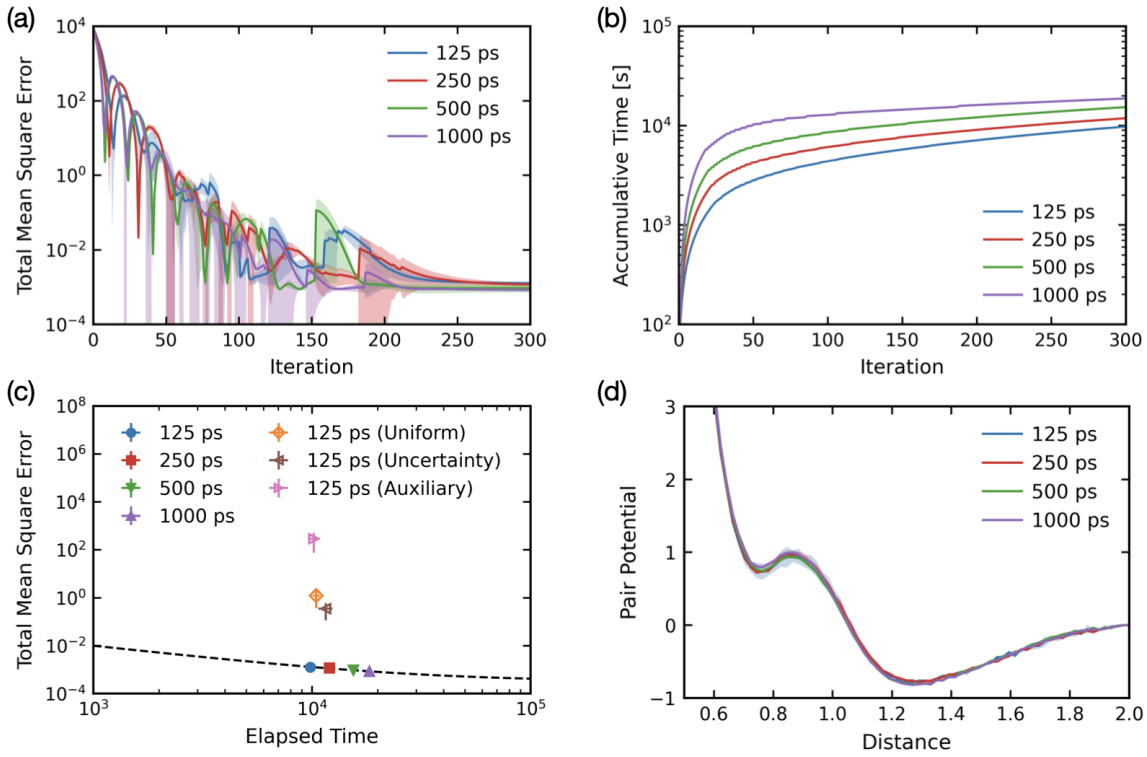

Figure 5: Comparisons of (a) total mean square error and (b) accumulative time of entire DiffCG optimization as a function of the optimization iteration; (c) Relationships between total mean square error at the 300th iteration and the elapsed time of the DiffCG optimization; (d) The pairwise potentials of CG-PS models obtained at the 300th iteration from CoV-based DiffCG optimizations with various MD sampling lengths: t_{sim} = 125 ps (blue), t_{sim} = 250 ps (red), t_{sim} = 500 ps (green), and t_{sim} = 1000 ps (purple).

The effect of MD sampling on the optimized potentials is important for the iterative coarse-graining approach. Therefore, we conduct a detailed analysis to carefully investigate the influence of the MD sampling length on the performance of the resulting CG-PS model in the DiffCG framework. Figure 5 (a) compares the total MSE E_{taotal} during the coarse-graining optimization with various sampling lengths from t_{sim} = 125 ps to t_{sim} = 1000 ps. The decay rate of MSE along the optimization iterations is similar for different sampling lengths, while the CG optimization with longer sampling length achieves lower MSE, i.e., higher accuracy. The cumulative time for the 300 iterative optimization with various sampling lengths is shown in Figure 5 (b). Obviously, optimization with a longer sampling requires a

longer accumulative time. However, the difference between each other is not proportional to the sampling length due to the reweighting scheme in DiffTre (see a detailed discussion in **Supporting Information**). Figure 5 (c) depicts the trade-off between accuracy and MD sampling length. The data from DiffCG using other multi-objective optimization algorithms are also included for comparison. The CoV-based DiffCG method surpasses all alternative approaches in achieving convergence towards the Pareto frontier. This analysis serves as a valuable reference to determine the appropriate sampling length in alignment with the desired level of accuracy. Figure 5 (d) shows that optimizations with various sampling lengths converge to the same potentials within errors after 300 iterations. The converged potentials fluctuate slightly between iterations, which may explain the small differences that can be observed. A noticeable dependence of the simulation lengths is the optimized dihedral potentials. As seen in Figure S1, the longer simulation length yields a better reproduction of the probability distribution of dihedral angles. It can probably be attributed to the slow relaxation times of the dihedral angles in comparison to other degrees of freedom, e.g., bond length and bond angle. This insight suggests the potential incorporation of enhanced sampling techniques to expedite the sampling process during the coarse-graining optimization. Additionally, it is worth noting that even the shortest sample length (t_{sim} = 125 ps) still yields satisfactory CG models of PS melts, showing the computational efficiency of DiffCG in the development of CG models.

Molecular Weight Transferability of CG PS model

In this section, the molecular weight transferability of the derived CG PS model is thoroughly examined. We conducted molecular simulations of CG PS models with various chain lengths from 10 to 100 monomers using the effective potentials derived at T = 600 K. Figure 6(a) shows the chain-length-dependent density compared to previous simulation results.[62] Despite the effect of temperature, both models show similar asymptotic behavior of chain-length dependence of the density, which converges to a plateau value for $N > 60$. In addition, we

analyzed the chain dimensions of PS melts with various chain lengths. As seen in Figure 6 (b), the radius of gyration R_g and chain end-to-end distances R_{ete} are plotted as a function of chain length N. A power law fit is performed for both radius of gyration and end-to-end distance, giving $R_g \sim N^{\nu}$ with $\nu \approx 0.565$ and $R_{ete} \sim N^{\nu}$ with $\nu \approx 0.555$, respectively. As a comparison, $\nu \approx 0.551$ for PS melts over a similar chain-length range was found in a recent work.[26] Moreover, the ratio $\langle R_{ete}^2 \rangle / \langle R_g^2 \rangle \approx 6.03$ for CG-PS model with $N = 100$ monomers is very close to the value of 6 predicted for Gaussian chains. These results underscore the robust transferability of the molecular weight within the derived CG PS model.

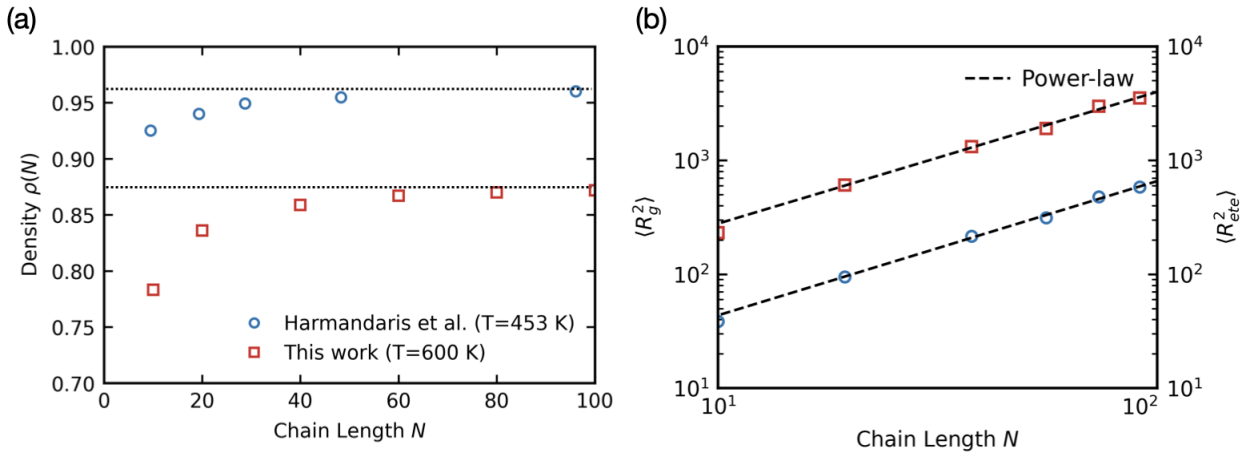

Figure 6: (a) Density predictions as a function of chain length N from CG MD simulations of PS model at $T = 600$ K (red), compared with previous simulation results of Harmandaris et al.[62] for PS at $T = 453$ K; (b) Chain sizes, ensemble-averaged square radius of gyration (blue) and ensemble-averaged square end-to-end distance of CG-PS model as a function of chain sizes from molecular simulations. The dashed lines are power law fits for radius of gyration and end-to-end distance: $R_g \sim N^{\nu}$ with $\nu \approx 0.565$ and $R_{ete} \sim N^{\nu}$ with $\nu \approx 0.555$, respectively.

Temperature Transferability of CG PS model

The CG models derived via the structure-based approaches often suffer from the issue of state dependence, where the effective CG potential derived at a certain thermodynamic state (e.g., temperature, concentration, pressure, etc.) is not guaranteed to be transferable to other states for accurate predictions of either thermodynamical or structural properties. We attempt to improve the transferability of structure-based coarse-graining through the

utilization of the DiffCG method, with a particular focus on developing a systematic CG model of PS melts that can be transferable over a wide range of temperature. Our strategy to constructing the thermodynamically-consistent temperature-transferable CG model under the DiffCG framework is from two sides. On the one hand, information at multiple states from fine-grained MD simulation is configured as the targets to optimize the CG potentials. This strategy was found to be useful for improving the derived CG potentials in the IBI framework. On the other hand, thermodynamic dependence (e.g., temperature, density, or volume) is embeded in the CG potential functional form, which could be necessary to improve the transferability of CG models as suggested in previous studies.[21,25] Specifically, we employ a dual-potential-like model[21] to incorporate temperature dependence in the pairwise potential:

$$U_{\text{pair}}(\mathbf{R}) = U(\mathbf{R}) + U(\mathbf{R}, k_B T) \tag{16}$$

where $U(\mathbf{R})$ and $U(\mathbf{R}, k_B T)$ are a temperature-independent (T-independent) and a temperature-dependent (T-dependent) potential energy functions, respectively. $U(\mathbf{R}, k_B T)$ is in principle a flexible functional term as a function of both configuration $\mathbf{R}$ and temperature T. The separation of two terms, as seen in Equation 16, is expected to correspond to the mean and variance of energy distributions, or alternatively, the energetic and entropic contributions to the free energy.[25] This choice is based on established principles and aligns with well-founded theoretical considerations.[21,25]

We define $U(\mathbf{R}, k_B T) = f(k_B T) U'(\mathbf{R})$, where $f(k_B T)$ is a simple functional of temperature and $U'(\mathbf{R})$ is a function invariant to temperature, neglecting the cross-correlation between configuration $\mathbf{R}$ and temperature T. To our knowledge, it is still unclear how the CG potentials would vary with temperature. Thus both linear (Equation 17) and non-linear (Equation 18) T-dependence are examined here:

$$U(\mathbf{R}, k_B T) = (\alpha k_B T + \beta) U'(\mathbf{R}) \tag{17}$$

$$U(\mathbf{R}, k_B T) = (k_B T^{\alpha} + \beta) U'(\mathbf{R}) \tag{18}$$

where α and β are optimizable parameters and $U'(R)$ is a temperature-independent pairwise potential represented by a spline function. Power-law function is selected here since it is a common non-linear model. The coarse-graining optimizations of these two types of potentials employ the same hyperparameters and simulation setups.

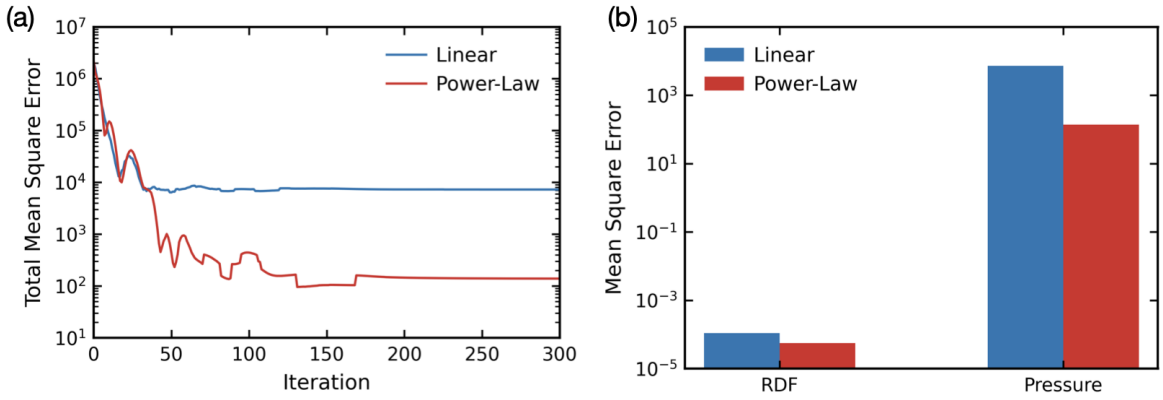

Figure 7: (a) A comparison of total mean square error as a function of the optimization iteration from applying DiffCG in constructing temperature-transferable CG-PS model with linear temperature-dependent pairwise potentials (blue) and power-law temperature-dependent pairwise potentials (red). The shaded areas are corresponded standard deviations from three independent optimizations with different random seeds; (b) The decomposition of total mean square error into individual sources or errors: radial distribution functions and pressure.

Figure 7 (a) displays evolution of the total MSE, E_{total}, as a function of optimization iterations for T-transferable CG-PS models, where $E_{total} = \sum^{T}(\epsilon_{bond} + \epsilon_{angle} + \epsilon_{dihedral} + \epsilon_{RDF} + \epsilon_{pressure})$, $T = 400, 500, 600$ K. The total MSE of the optimization with power-law T-dependent potential decreases much faster than the linear one in all iterations. Moreover, the CG model with power-law T-dependent potentials achieves a lower E_{total} compared with the linear counterpart, which is $\sim 10^3$ times. In addition, the MSE of the individual source in the final iteration is shown separately in Figure 7 (b). Both the RDF and pressure predictions have smaller error from the power-law model than the linear one, suggesting the underlying non-linear dependence of the CG potentials on the temperature.

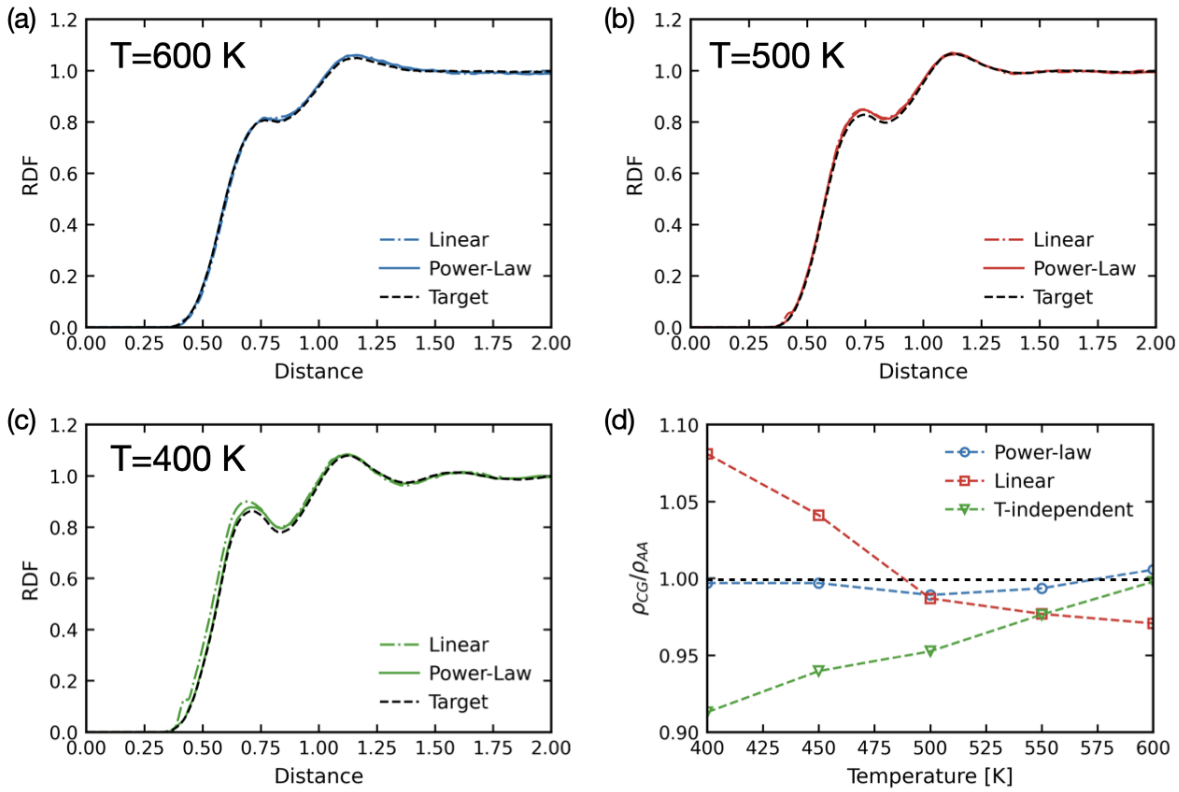

Figure 8: Radial distribution functions from temperature-transferable CG-PS model with linear temperature-dependent pairwise potentials (dotted line) and power-law temperature-dependent pairwise potentials (solid line) at (a) $T = 600$ K, (b) $T = 500$ K, (c)$T = 400$ K compared with respective target AA-MD data (dashed line); (d) Ratio of mass densities ρ_{CG}/ρ_{AA} between AA and CG models with various temperature-dependent pairwise potentials: Power-law (blue), Linear (red), and temperature-independent (green) obtained from 1-ns NPT simulations.

The predictions of the RDFs at temperatures $T = 600$ K, $T = 500$ K and $T = 400$ K are shown in Figure 8 (a)-(c), respectively. Good agreement with the AA-MD reference simulations indicates that the CG-PS model with power-law T-dependent potentials can capture the temperature transferability of structures. Although less accurate than the power-law model (particularly at low temperature $T = 400$ K), the linear T-dependent model generally shows satisfactory predictions in structural properties from NPT simulations of the CG-PS model at high temperatures. However, as displayed in Figure 8 (d), a substantial deviation in mass density prediction is evident when the linear T-dependent model is used. The situation is much worse when the T-independent potential is used (obtained from DiffCG

optimization at $T = 600$ K), consistent with previous findings of CG models developed using the IBI method.[63] On the contrary, the mass densities from the power-law T-dependent model match very well with those of the AA model over temperatures ($T = 450, 550$ K) that are not even used for optimizations. The advantages of power-law temperature dependence might be attributed to the fact that it allows larger variations as a function of temperature in pairwise potentials. These observations underscore the efficacy of the non-linear formulation of temperature dependence in CG pairwise potentials, which is the key in achieving accurate structure and pressure predictions in a wide range of temperatures. The underlying mechanisms for the nonlinearity observed in the CG model calls for detailed studies in the future.

Discussions and Conclusion

We have successfully introduced an efficient and versatile structural coarse-graining method for complex soft-matter systems, namely, differentiable coarse-graining (DiffCG). This approach combines the principles of multi-objective optimization and differentiable simulation techniques with trajectory reweighting. In order to identify the most suitable candidate for integration into the DiffCG methodology, we conducted a comprehensive benchmark of three widely-used multi-objective optimization algorithms in the machine learning community. Our meticulous assessments and systematic comparisons unequivocally demonstrate that the CoV weighting algorithm consistently outperforms other prominent multi-objective optimization algorithms examined in this study for the structural coarse-graining task. With the integration of CoV, we demonstrate that DiffCG is capable of constructing CG models of polymers with both structural and thermodynamic properties optimized at the same time during coarse-graining optimization. We have also evaluated the relationship between the accuracy of the CG model and MD sampling, which is the most time-intensive step in DiffCG. Our result reveals a discernible trade-off between these two factors. Specifically, the accu-

racy of the CG model increases with increasing MD sampling, while the improvement is not sufficiently significant even when extending the sampling to $t_{sim} = 1000$ ps from $t_{sim} = 125$ ps. Our findings from the example CG-PS model indicate that the required MD sampling is inherently influenced by slow modes within the corresponding AA model, e.g., dihedral angles for polymers. To expedite the DiffCG process for coarse-graining more intricate systems, enhanced sampling techniques such as tempering-based or collective-variable-based methods may prove to be invaluable.[64,65]

Moreover, we have shown how a temperature-transferable CG model of PS melts can be developed using DiffCG. To achieve this, we harness the multi-objective optimization capabilities of DiffCG, utilizing target data collected from multiple thermodynamic states. In our approach, a dual-potential-like form is adopted to represent the pairwise potential within the transferable CG model. We conduct a systematic comparison between linear and non-linear T-dependent pairwise potentials to assess their effectiveness in ensuring temperature transferability. Our results indicate that the reproduction of both structural and thermodynamic properties in CG-PS models rely on the non-linear T-dependence of the pairwise potential. This finding underscores the significance of considering non-linearity when striving for precise temperature transferability in CG models, which has not yet been thoroughly considered in previous studies.[22,25,49] It is noteworthy that a similar non-linear scheme has been successfully applied in previous studies involving temperature-transferable coarse-graining.[61,66] In future research, it might be imperative to develop a rigorous theoretical framework that can provide insights into the underlying mechanisms of the observed nonlinearity in the temperature dependence of the CG model.

As DiffCG has been successfully demonstrated to optimize CG models with transferability over a wide range of temperatures, it is reasonable to expect that our approach is able to construct CG models transferable across other conditions, such as interfaces, multicomponent systems, diverse compositions, and so on. Extending this methodology to handle such intricate environments may necessitate the development of CG potentials with improved

expressiveness. Recent breakthroughs in machine-learning force-fields (MLFF) hold great promise as a viable approach for developing accurate and transferable CG models.[27,67,67] In particular, MLFFs have shown their effectiveness in CG simulations of various systems, including organic molecules,[27] liquids,[68,69] and fast-folding proteins.[28] However, the extent to which MLFFs can be applied to condensed-phase macromolecular systems remains an intriguing question for further exploration.[1,29,70] Furthermore, an ideal scenario for the development of CG models envisions integration with an automated optimization loop, possibly through active learning schemes, as done in MLFFs of interatomic potential.[71,72] This integration would unlock the full potential of coarse-graining techniques, enabling the modeling of complex systems across extensive spatio-temporal scales. DiffCG is anticipated to be a transformative tool in structural coarse-graining, with the potential to address a wide range of challenges and pave the way for innovative developments in CG modeling.

Supporting Information Available

Details about differentiable coarse-graining optimization for polystyrene melts; Probability distribution functions of bond length, bond angle, dihedral angles, and average pressure from CG-PS models developed by DiffCG with various MD sampling; Tabulated potentials for the transferable CG-PS model.

Acknowledgement

Wu Z. thanks Wujie Wang, Xiang Fu, Tian Xie for their helpful comments and suggestions. Wu Z. thanks Prof. Sinan Keten for his kind guidance and helpful advice at Northwestern University. Wu Z. acknowledges the computing resource granted by supercomputing center at XJTLU. Wu Z. ackowledges the support by State Key Laboratory of Heavy Oil Processing (RRSP10120230193). Zhou T. is supported by the National Natural Science Foundation of China (22308376 and 20220242), the Outstanding Young Scholars Foundation of China

University of Petroleum (Beijing) (ZX20230080), and the Foundation of United Institute for Carbon Neutrality (CNIF20230209).

References

(1) Dhamankar, S.; Webb, M. A. Chemically specific coarse-graining of polymers: Methods and prospects. *J. Polym. Sci.* **2021**, *59*, 2613–2643.

(2) Shi, R.; Qian, H.; Lu, Z. Coarse-grained molecular dynamics simulation of polymers: Structures and dynamics. *WIREs Comput Mol Sci* **2023**, e1683.

(3) Padding, J. T.; Briels, W. J. Systematic coarse-graining of the dynamics of entangled polymer melts: the road from chemistry to rheology. *J. Phys.: Condens. Matter* **2011**, *23*, 233101.

(4) Everaers, R.; Karimi-Varzaneh, H. A.; Fleck, F.; Hojdis, N.; Svaneborg, C. Kremer–Grest Models for Commodity Polymer Melts: Linking Theory, Experiment, and Simulation at the Kuhn Scale. *Macromolecules* **2020**, *53*, 1901–1916.

(5) Wu, Z.; Milano, G.; Müller-Plathe, F. Combination of Hybrid Particle-Field Molecular Dynamics and Slip-Springs for the Efficient Simulation of Coarse-Grained Polymer Models: Static and Dynamic Properties of Polystyrene Melts. *J. Chem. Theory Comput.* **2021**, *17*, 474–487.

(6) Wu, Z.; Müller-Plathe, F. Slip-Spring Hybrid Particle-Field Molecular Dynamics for Coarse-Graining Branched Polymer Melts: Polystyrene Melts as an Example. *J. Chem. Theory Comput.* **2022**, *18*, 3814–3828.

(7) Ardham, V. R.; Leroy, F. Atomistic and Coarse-Grained Modeling of the Adsorption of Graphene Nanoflakes at the Oil–Water Interface. *J. Phys. Chem. B* **2018**, *122*, 2396–2407.

(8) Sherck, N.; Shen, K.; Nguyen, M.; Yoo, B.; Köhler, S.; Speros, J. C.; Delaney, K. T.; Shell, M. S.; Fredrickson, G. H. Molecularly Informed Field Theories from Bottom-up Coarse-Graining. *ACS Macro Lett.* **2021**, *10*, 576–583.

(9) Carmichael, S. P.; Shell, M. S. A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly. *J. Phys. Chem. B* **2012**, *116*, 8383–8393.

(10) Wang, S.; Larson, R. G. Coarse-Grained Molecular Dynamics Simulation of Self-Assembly and Surface Adsorption of Ionic Surfactants Using an Implicit Water Model. *Langmuir* **2015**, *31*, 1262–1271.

(11) Shmilovich, K.; Mansbach, R. A.; Sidky, H.; Dunne, O. E.; Panda, S. S.; Tovar, J. D.; Ferguson, A. L. Discovery of Self-Assembling -Conjugated Peptides by Active Learning-Directed Coarse-Grained Molecular Simulation. *J. Phys. Chem. B* **2020**, *124*, 3873–3891.

(12) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759–5791.

(13) Noid, W. G. Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models. *J. Phys. Chem. B* **2023**, *127*, 4174–4207.

(14) Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods* **2021**, *18*, 382–388.

(15) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(16) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.

(17) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations: Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.

(18) Soper, A. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **1996**, *202*, 295–306.

(19) Mullinax, J. W.; Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **2009**, *131*, 104110.

(20) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, 224104.

(21) Lebold, K. M.; Noid, W. G. Dual approach for effective potentials that accurately model structure and energetics. *J. Chem. Phys.* **2019**, *150*, 234107.

(22) Lebold, K. M.; Noid, W. G. Dual-potential approach for coarse-grained implicit solvent models with accurate, internally consistent energetics and predictive transferability. *J. Chem. Phys.* **2019**, *151*, 164113.

(23) Jin, J.; Voth, G. A. Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *J. Chem. Theory Comput.* **2018**, *14*, 2180–2197.

(24) Jin, J.; Yu, A.; Voth, G. A. Temperature and Phase Transferable Bottom-up Coarse-Grained Models. *J. Chem. Theory Comput.* **2020**, *16*, 6823–6842.

(25) Pretti, E.; Shell, M. S. A microcanonical approach to temperature-transferable coarse-grained models using the relative entropy. *J. Chem. Phys.* **2021**, *155*, 094102.

(26) Zhang, X.-Z.; Lu, Z.-Y.; Qian, H.-J. Temperature Transferable and Thermodynamically Consistent Coarse-Grained Model for Binary Polymer Systems. *Macromolecules* **2023**, *56*, 3739–3753.

(27) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; De Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755–767.

(28) Durumeric, A. E.; Charron, N. E.; Templeton, C.; Musil, F.; Bonneau, K.; Pasos-Trejo, A. S.; Chen, Y.; Kelkar, A.; Noé, F.; Clementi, C. Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.* **2023**, *79*, 102533.

(29) Ricci, E.; Vergadou, N. Integrating Machine Learning in the Coarse-Grained Molecular Simulation of Polymers. *J. Phys. Chem. B* **2023**, *127*, 2302–2322.

(30) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput Mater* **2019**, *5*, 125.

(31) Yang, W.; Templeton, C.; Rosenberger, D.; Bittracher, A.; Nüske, F.; Noé, F.; Clementi, C. Slicing and Dicing: Optimal Coarse-Grained Representation to Preserve Molecular Kinetics. *ACS Cent. Sci.* **2023**, *9*, 186–196.

(32) An, Y.; Deshmukh, S. A. Machine learning approach for accurate backmapping of coarse-grained models to all-atom models. *Chem. Commun.* **2020**, *56*, 9312–9315.

(33) Stieffenhofer, M.; Wand, M.; Bereau, T. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 045014.

(34) Li, W.; Burkhart, C.; Polińska, P.; Harmandaris, V.; Doxastakis, M. Backmapping coarse-grained macromolecules: An efficient and versatile machine learning approach. *J. Chem. Phys.* **2020**, *153*, 041101.

(35) Ding, X.; Zhang, B. Contrastive Learning of Coarse-Grained Force Fields. *J. Chem. Theory Comput.* **2022**, *18*, 6334–6344.

(36) Krämer, A.; Durumeric, A. E. P.; Charron, N. E.; Chen, Y.; Clementi, C.; Noé, F.

Statistically Optimal Force Aggregation for Coarse-Graining Molecular Dynamics. *J. Phys. Chem. Lett.* **2023**, *14*, 3970–3979.

(37) Greener, J. G.; Jones, D. T. Differentiable molecular simulation can learn all the parameters in a coarse-grained force field for proteins. *PLoS ONE* **2021**, *16*, e0256990.

(38) Wang, W.; Axelrod, S.; Gómez-Bombarelli, R. Differentiable Molecular Simulations for Control and Learning. 2020; `http://arxiv.org/abs/2003.00868`, arXiv:2003.00868 [physics, stat].

(39) Wang, W.; Wu, Z.; Dietschreit, J. C. B.; Gómez-Bombarelli, R. Learning pair potentials using differentiable simulations. *J. Chem. Phys.* **2023**, *158*, 044113.

(40) Šípka, M.; Dietschreit, J. C. B.; Grajciar, L.; Gómez-Bombarelli, R. Differentiable Simulations for Enhanced Sampling of Rare Events.

(41) Metz, L.; Freeman, C. D.; Schoenholz, S. S.; Kachman, T. Gradients are Not All You Need. 2022; `http://arxiv.org/abs/2111.05803`, arXiv:2111.05803 [cs, stat].

(42) Thaler, S.; Zavadlav, J. Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting. *Nat Commun* **2021**, *12*, 6884.

(43) Carbone, P.; Varzaneh, H. A. K.; Chen, X.; Müller-Plathe, F. Transferability of coarse-grained force fields: The polymer case. *J. Chem. Phys.* **2008**, *128*, 064904.

(44) Karimi-Varzaneh, H. A.; Van Der Vegt, N. F. A.; Müller-Plathe, F.; Carbone, P. How Good Are Coarse-Grained Polymer Models? A Comparison for Atactic Polystyrene. *ChemPhysChem* **2012**, *13*, 3428–3439.

(45) Rudzinski, J. F.; Noid, W. G. The Role of Many-Body Correlations in Determining Potentials for Coarse-Grained Models of Equilibrium Structure. *J. Phys. Chem. B* **2012**, *116*, 8621–8635.

(46) John, S. T.; Csányi, G. Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials. *J. Phys. Chem. B* **2017**, *121*, 10934–10949.

(47) Scherer, C.; Andrienko, D. Understanding three-body contributions to coarse-grained force fields. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22387–22394.

(48) Wang, J.; Charron, N.; Husic, B.; Olsson, S.; Noé, F.; Clementi, C. Multi-body effects in a coarse-grained protein force field. *J. Chem. Phys.* **2021**, *154*, 164113.

(49) Rosenberger, D.; Van Der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6617–6628.

(50) Rosenberger, D.; Van Der Vegt, N. F. A. Relative entropy indicates an ideal concentration for structure-based coarse graining of binary mixtures. *Phys. Rev. E* **2019**, *99*, 053308.

(51) Moradzadeh, A.; Motevaselian, M. H.; Mashayak, S. Y.; Aluru, N. R. Coarse-Grained Force Field for Imidazolium-Based Ionic Liquids. *J. Chem. Theory Comput.* **2018**, *14*, 3252–3261.

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

(53) Bernhardt, M. P.; Hanke, M.; Van Der Vegt, N. F. Stability, Speed, and Constraints for Structural Coarse-Graining in VOTCA. *J. Chem. Theory Comput.* **2023**, *19*, 580–595.

(54) Kendall, A.; Gal, Y.; Cipolla, R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. 2018; `http://arxiv.org/abs/1705.07115`, arXiv:1705.07115 [cs].

(55) Liebel, L.; Körner, M. Auxiliary Tasks in Multi-task Learning. 2018; `http://arxiv.org/abs/1805.06334`, arXiv:1805.06334 [cs].

(56) Groenendijk, R.; Karaoglu, S.; Gevers, T.; Mensink, T. Multi-Loss Weighting with Coefficient of Variations. 2020; http://arxiv.org/abs/2009.01717, arXiv:2009.01717 [cs].

(57) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In 'T Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(58) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(59) Babuschkin, I. et al. The DeepMind JAX Ecosystem. 2020; http://github.com/deepmind.

(60) Schoenholz, S. S.; Cubuk, E. D. JAX, M.D.: A Framework for Differentiable Physics. 2020; http://arxiv.org/abs/1912.04232, arXiv:1912.04232 [cond-mat, physics:physics, stat].

(61) Xia, W.; Song, J.; Jeong, C.; Hsu, D. D.; Phelan, F. R.; Douglas, J. F.; Keten, S. Energy-Renormalization for Achieving Temperature Transferable Coarse-Graining of Polymer Dynamics. *Macromolecules* **2017**, *50*, 8787–8796.

(62) Harmandaris, V. A.; Kremer, K. Dynamics of Polystyrene Melts through Hierarchical Multiscale Simulations. *Macromolecules* **2009**, *42*, 791–802.

(63) Farah, K.; Fogarty, A. C.; Böhm, M. C.; Müller-Plathe, F. Temperature dependence of coarse-grained potentials for liquid hexane. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2894–2902.

(64) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced sampling in molecular dynamics. *J. Chem. Phys.* **2019**, *151*, 070902.

(65) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *LiveCoMS* **2022**, *4*.

(66) Qian, H.-J.; Carbone, P.; Chen, X.; Karimi-Varzaneh, H. A.; Liew, C. C.; Müller-Plathe, F. Temperature-Transferable Coarse-Grained Potentials for Ethylbenzene, Polystyrene, and Their Mixtures. *Macromolecules* **2008**, *41*, 9919–9929.

(67) Ruza, J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Harris, W. H.; Gómez-Bombarelli, R. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.* **2020**, *153*, 164501.

(68) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. DeePCG: Constructing coarse-grained models via deep neural networks. *J. Chem. Phys.* **2018**, *149*, 034101.

(69) Thaler, S.; Stupp, M.; Zavadlav, J. Deep coarse-grained potentials via relative entropy minimization. *J. Chem. Phys.* **2022**, *157*, 244103.

(70) Ye, H.; Xian, W.; Li, Y. Machine Learning of Coarse-Grained Models for Organic Molecules and Polymers: Progress, Opportunities, and Challenges. *ACS Omega* **2021**, *6*, 1758–1772.

(71) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Materials* **2019**, *3*, 023804.

(72) Vandermause, J.; Xie, Y.; Lim, J. S.; Owen, C. J.; Kozinsky, B. Active learning of reactive Bayesian force fields applied to heterogeneous catalysis dynamics of H/Pt. *Nat Commun* **2022**, *13*, 5183.

TOC Graphic