Machine Learning-Based Coarse Grained Interaction Potentials for Molecular Systems

D.-P. Gerakinis
National Centre for Scientific Research "Demokritos", Athens
dp.gerakinis@inn.demokritos.gr

E. Ricci
National Centre for Scientific Research "Demokritos", Athens, Greece
e.ricci@inn.demokritos.gr

G. Giannakopoulos
National Centre for Scientific Research "Demokritos", Athens, Greece & SciFY P.N.P.C., Greece
ggianna@iit.demokritos.gr

V. Karkaletsis
National Centre for Scientific Research "Demokritos", Athens, Greece
vangelis@iit.demokritos.gr

D. N. Theodorou
School of Chemical Engineering, National Technical University of Athens, Athens, Greece
doros@chemeng.ntua.gr

N. Vergadou
National Centre for Scientific Research "Demokritos", Athens
n.vergadou@inn.demokritos.gr

ABSTRACT

The development and implementation of hierarchical multiscale methods is necessary for the molecular simulation of complex chemical systems such as organic fluids and soft matter systems, in order to reach longer length and time scales. Coarse-Graining (CG) is at the core of multiscale methods. Machine Learning (ML) techniques have been investigated in the recent years for the development of improved atomistic force fields based on quantum mechanical calculations. However, the integration of ML methods into the development of CG force fields required for hierarchical multiscale modelling schemes for bulk organic systems, on the basis of atomistic simulations, is still very scarce. In this work, Graph Convolutional Neural Network (GCNN) architectures were adopted to develop CG Machine Learned potentials for bulk amorphous systems, implementing a strategy that includes a force-matching scheme using benzene liquid as a test system. Two CG representations were implemented for the benzene molecule: a single bead and a three-bead representation. The ML-based CG force fields were utilized to perform molecular dynamics (MD) simulations of the systems at the CG level and the properties of the CG systems simulated with a neural network potential were compared against the underlying atomistic ones, to quantify the effectiveness of the developed models. The effects of hyperparameters, loss function construction and GCNN architecture size were thoroughly investigated and discussed providing a wealth of information that can serve as a general basis for the reliable use of ML-based CG approaches in the wide field of bulk soft matter systems.

1. Introduction

The development of innovative technologies that can address the crucial contemporary environmental and social demands require the design of new materials with controlled properties. This challenge necessitates a fundamental understanding of the interactions and the microscopic mechanisms that underlie the macroscopic behavior of the materials. Molecular simulations can play a crucial role in gaining such an insight and enable the establishment of structure-property relationships. The study of complex chemical systems with molecular simulation methods requires the development of multiscale hierarchical methods [1]-[3] in order to access longer length and timescale phenomena.

Coarse-graining methodologies are a core component of multiscale methodologies. Coarse-graining involves the substitution of groups of atoms by effective interaction sites [4, 5], in a manner that preserves the degrees of freedom that are important for the study of the mechanisms and properties of interest. The interactions in this lower dimensional space are modeled through CG force fields which are approximations of the potentials of mean force (PMFs). Predefined analytical functional forms are usually employed and their parametrization is conducted as a top-down [4]–[8] fitting process, where it is ensured that they reproduce the macroscopic behavior of the material, or as a bottom-up [9]–[11] fitting process, for the reproduction of properties based on atomistic simulations, or from a combination of these two approaches.

Machine Learning techniques have been investigated in the recent years for the development of improved atomistic force fields based on quantum mechanical calculations [13]. However, the integration of ML methods into the development of CG force fields required for hierarchical multiscale modelling schemes for bulk amorphous systems on the basis of atomistic simulations is still rare [14]. In this work, Neural Network-based models capable of learning complex potential energy hypersurfaces are utilized for the approximation of PMFs, replacing pairwise predefined analytical functionals for the description of the CG interactions. A Graph Convolutional Neural Network (GCNN) architecture, SchNet [15], was adopted to develop CG ML potentials, implementing and optimizing a scheme [16] that includes a force-matching procedure [17]. Benzene liquid was used as a benchmark system using two CG representations for the benzene molecule: a single bead and a three-bead representation. The influence of hyperparameter tuning, loss function construction and GCNN architecture size were widely explored. The obtained models were used to perform CG Molecular Dynamics simulations and the structural and thermodynamic properties of

the CG systems were systematically compared with the underlying atomistic reference to quantify the effectiveness of the developed models.

2. Methodology

2.1 SchNet Graph Convolutional Neural Network Architecture

An essential step in the development of ML-assisted methods is the determination of input features for the ML model. Generally, such a transformation must end up in the construction of descriptors respecting invariance to translation, rotation, and permutation of the particles' order. Furthermore, it is important to have a descriptor independent of the number of particles in the system. Thereby, the resulting ML model can be efficiently deployed to larger or smaller systems than the ones upon which it has been trained. Rather than constructing a priori fixed feature representations, more general and transferable strategies have been proposed, which consist of learning the molecular representation instead [18-21] along with the desired property.

The Graph Convolutional NN architecture implemented in this work is based on the SchNet [22] architecture. A detailed description of the SchNet architecture can be found in Refs. [15], [22] and [23]. The SchNet architecture was initially developed for the analysis of small molecule structures and bulk crystals to predict atomistic energies and forces based on the quantum mechanical information, but its application can be extended for the study of molecular systems at the CG level as well, based on models trained on the atomistic reference [24], [25]. SchNet is a GCNN with learned local environment representation, requiring several hyperparameters. The graph nodes represent particles and edges are considered between all particles, CG moieties in the present case, within a cutoff distance. Edge features are defined as the distances between the beads, preserving translational and rotational invariance.

Each particle is represented through a feature vector, which is initialized to distinguish between the particle chemical identities (embedding layer). The feature is then updated in each NN layer depending on the chemical environment, by performing continuous filter convolutions across the particle neighborhoods, optimizing the convolutional filter weights during the training (convolutional layers). Afterwards, a fully connected section is included (readout layers). A sequence of continuous convolution layers and readout layers constitutes a SchNet "block". Series of blocks can be utilized in series to define the full network architecture. The output found at the end of the blocks can be interpreted as a learned feature representation, which encodes the many-body information from the particle neighborhood required to

predict the target property. Finally, a fully connected (dense) section transforms the fingerprints into the final scalar output, which is interpreted as a per-particle energy contribution. This local decomposition ensures invariance of the ML potential architecture to the total number of atoms. In the presence of intramolecular and intermolecular interactions, the possibility to construct two input features for each particle, one that would consider only intermolecular neighboring particles, and one including only the intramolecular ones was investigated, as implemented in Ref. [24]. In this case, the values for the learned inter- and intramolecular features were summed before the final dense layer. All energy contributions are summed per configuration to obtain the total energies, $\sum_j \widehat{U}_j$, which are then differentiated with respect to the positions of the particles to predict the force acting on each particle.

2.2 Loss Function Construction and CG Force Field Determination

The determination of CG force field is implemented with the inclusion of a the force-matching scheme [17]. This scheme considers the mean squared Euclidean (L2-Squared) distance between the forces predicted by the CG model, $\boldsymbol{F}^{CG} \in \mathbb{R}^{3N}$, for the N number of CG sites having coordinates $\boldsymbol{x}_i$ and the coarse-grained forces, as calculated based on the mapping of the atomistic forces, $\boldsymbol{F}^A \in \mathbb{R}^{3\nu}$, where ν is the number of atoms, onto the respective CG sites:

$$\chi^2 = \langle \left[\boldsymbol{F}^{CG}(\boldsymbol{x}_i) - \mathcal{M}\left(F^A(\boldsymbol{r}_j) \right) \right] \rangle \qquad (1)$$

In the expression above $\mathcal{M}$ is the mapping operator between the atomistic and the CG space, $\mathbf{r}_j$ are the atomistic coordinates and the brackets $\langle \dots \rangle$ denote average over CG degrees of freedom and the number of sampled atomistic configurations.

During the GCNN training, the loss function was initially constructed to be based on the minimization of the mean squared difference between the predicted forces and the forces associated with the atomistic molecular dynamics (MD) trajectory considered as input, using the following loss function L:

$$L = \frac{1}{3N} \sum_i^N \left\{ \left(-\boldsymbol{\nabla}_{x_i} U^{CG} \right) - \mathcal{M}\left(F^A(\boldsymbol{r}_j) \right) \right\}^2 \qquad (2)$$

Where U^{CG} is the estimated energy of the system in the CG representation and the first term in eq. (2) is the force acting on the CG particle at position $\mathbf{x}_i$ as predicted CG force field. The estimated CG energy of

the system in the ML-based procedure is usually obtained by the sum of the per-particle energy contributions, $\widehat{U}_i$, which is the output of the GCNN, plus potential prior terms:

$$U^{CG} = U_{prior} + \sum_i \widehat{U}_i \qquad (3)$$

The prior terms are oftentimes used to "inform" the model on unphysical states when these states are not included in the training data and prevent it from generating them. This is necessary for example in the case of training taking place using only physically valid configurations as the ones included in an MD trajectory. In order to ensure that the CG beads will not be allowed to approach too close to each other or even overlap, an excluded volume energy prior term of the following form is taken into account:

$$U_{ex} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\sigma}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right)^n \qquad (4)$$

Where σ is the excluded volume diameter and n the excluded volume exponent and they are additional hyperparameter, whose appropriate values need to be explored in each case. For multi-bead per molecule representations in which connectivity is to be taken into account, additional prior terms on specific geometric features, such as bonds and angles may need to be added. For example, in the three-bead representation of the benzene molecule, a harmonic bond prior energy term was included to hinder bonds overstretching:

$$U_{bond} = k_b(l - l_0)^2 \qquad (5)$$

Where k_b is the bond stiffness constant, l_0 is the equilibrium CG bond length, l is the actual bond length. The bond stiffness constant k_b is an additional model hyperparameter.

The appropriate loss construction is a very important component towards an efficient ML-based training procedure. In the case of the present test system, it was found that loss function solely based on force component is not sufficient to provide well-performing CG models, as discussed in Section 3 and also reported in previous work [16]. A modification in the loss function was thus investigated by including also the squared deviation between the energy value predicted by the model and an energy component from the atomistic simulation (U_A). The loss function definition has in this case the following form:

$$L = \frac{1}{3N} \sum_i^N \left(-\boldsymbol{\nabla}_{x_i} U^{CG} - \mathcal{M}\left(\boldsymbol{F}^A(\boldsymbol{r}_j) \right) \right)^2 + \lambda (U^{CG} - U_A)^2 \qquad (6)$$

and λ is a model hyperparameter that serves as weighting factor of the energy component with respect to the force component in the loss.

2.3   Simulation Details and Training Procedure

The liquid benzene system was studied considering two individual CG mapping (Figure 2.1) schemes of: (a) one bead per molecule and of (b) three CG beads per molecule. In the latter case, each bead represented two carbon atoms and their bonded hydrogens, and its centre of mass was located at the centre of mass of its four constituent atoms.
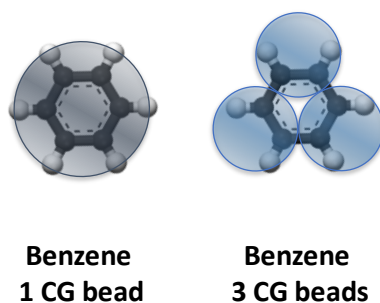
Figure 2.1. Graphical depiction of the two CG mappings implemented for the benzene molecule.

The training data were obtained from atomistic MD trajectory frames of liquid benzene simulated at 300 K and atmospheric pressure, under periodic boundary conditions using the Amber Cornell force field [26]. The system was equilibrated through a 1 ns NPT MD simulation, and then a 20 ns NVT run was conducted using LAMMPS [27]. The reference atomistic system for the 1 CG bead case consisted of 500 molecules that correspond to 500 CG interaction sites and for the 3 CG bead per molecule representation the reference atomistic system contained 300 molecules, corresponding to 900 CG interaction sites. The first half of the run was considered equilibration and discarded, and afterwards 10000 simulation frames were retained. 90% of the data was used for the training set, 10% was kept for the test set with the configurations from the MD trajectory being randomly assigned to either set.

The SchNet architecture has several hyperparameters that need to be tuned: these values are pre-defined and remain unchanged during a training run. The hyperparameters include: the size of the architecture (such as number of convolutional filters and SchNet blocks, filter size), the activation functions in use,

parameters related to the training procedure (such as batch size, learning rate, decay ratio), the input data sample size, the systems representation (via the feature size vector and the cutoff distance of neighbors) as well as hyperparameters related to the construction of the loss (weighting parameters for the various loss components or to tune the prior terms). All these factors may affect the quality of the training and the effectiveness of the obtained model. After the ML model training using a specific hyperparameter set, NVT CG simulations with the GCNN potentials were conducted using the Atomic Simulation Environment (ASE) python package [28], in order to evaluate the performance of the ML-based trained models.

3. Results

The training of a successful model was a very challenging and laborious procedure taking into account the wide range of hyperparameters that needed to be optimized and tested. The monitoring of the training procedure itself could not serve as a reliable verification means of an efficient training as the convergence in the loss function is not an indicator of the development a successful model. The performance of the obtained ML-based models had to be evaluated conducting CG simulations in order to ensure that a model could result in thermodynamically stable simulations and meaningful dynamic behavior and spatial organization, the latter being compared to the atomistic ground truth. Several models in which the loss function during training was well behaved and converged were found to result either in unstable or physically inconsistent CG simulations, as also reported in Ref. [16]. This problematic behavior involved very high temperature fluctuations that correspond to fast in place oscillations of the molecules, simulations in which the system ends up in a "frozen" state, or observation of unphysical structural characteristics, for example at very short distances. For the benzene liquid systems under study, the development of reasonable models became systematically feasible when the loss function was extended to include, in addition to the mean square error on the prediction of the force, also a second component that included an energy term (eq. (6)).

3.2 Single-bead representation

For the single-bead representation of the benzene liquid system, we investigated several training conditions, primarily focusing on: (a) the size of the architecture through the hyperparameter controlling the number of blocks constituting the network, (b) the value of the *cutoff* hyperparameter through which each particle's neighborhood is defined, (c) the hyperparameters σ and n, concerning the excluded volume

prior energy term, (d) we investigated the transferability of the trained models by employing them to perform CG simulations at different temperatures or for different system sizes than the ones upon which they were trained in the first place, (e) the size of datasets for the training and (f) the employment of the softplus function as activation function, as an alternative to the tanh function.

More than 150 sets of tests were conducted, in which we investigated different hyperparameter combinations exploring how specific hyperparameters affect the training, the stability and the performance of the obtained models. The information on the investigated hyperparameter range is shown in Table 3.1.

Table 3.1. SchNet hyperparameter ranges investigated for the tests with liquid benzene, for the cases of the 1-CG-bead-per-molecule representation.

Hyperparameter	Values	Hyperparameter	Values
Excluded volume radius (σ) [Å]	2, 3, 4, 4.5, 5, 6, 6.5, 7, 8, 9	Filter size	40, 80
Excluded volume exponent (n)	5, 6, 7, 8, 9	Number of filters	128, 256
Local neighborhood cutoff radius [Å]	5, 7, 8, 9, 12	Activation function	Tanh, shifted softplus
Number of SchNet blocks	1, 2, 3, 4, 6	Energy scaling in the loss function (λ)	0, 10^{-7}, 10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10
Feature size	120, 240	Batch size	1, 5, 10, 15, 30, 50

The influence of hyperparameter combinations on the performance of the obtained models was thoroughly investigated. Specific hyperparameter combinations were found to produce models that resulted in stable ML-assisted CG simulations and reasonable results in terms of dynamical behavior and of structural characteristics compared to the atomistic ground truth provided that the energy term was included in the loss functions as shown in eq. (6) In this mapping scheme of 1 bead per molecule only intermolecular interactions are present in the system. The energy prior terms used in this case, involve the inclusion of the excluded volume prior term.

In Figure 3.1, the effect of changing the energy scaling factor λ, keeping all other hyperparameters the same (*cutoff* = 8 Å, n = 7, σ = 5 Å) is shown. From these comparative graphs we observe that for $\lambda = 0$, an unstable simulation is obtained while positive results are obtained for $\lambda = 10^{-1}$. In Figures 3.2(a) and 3.2(b), results are presented for hyperparameter combinations with n = 6, σ = 5 Å, λ = 1 and examining the effect of the local neighborhood cutoff radius hyperparameter. From these comparative graphs we

observe that for *cutoff* = 6 Å we get a thermodynamically unstable simulation. For larger values of the *cutoff*, the simulation is stable while the position of the first peak of the intermolecular g(r) remains rather unchanged.
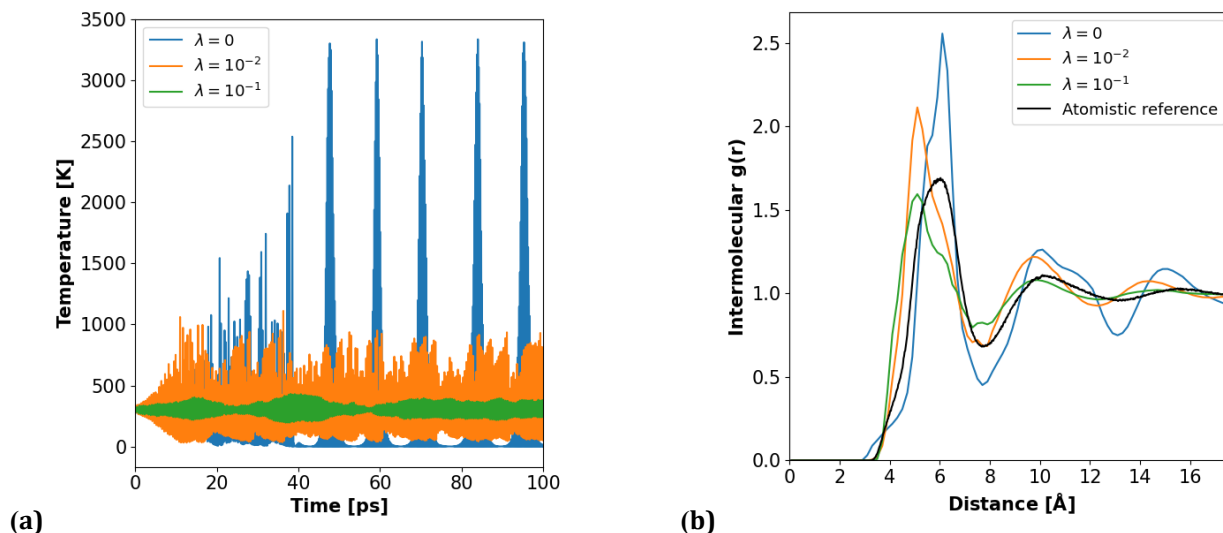

(a)

(b)

Figure 3.1. Comparative graphs of (a) the evolution of the temperature during the ML-based CG simulation [left] and (b) Comparison between the intermolecular radial distribution functions of the system after the simulation [colored] and the atomistic ground truth [black], for *cutoff* = 8 Å, $n = 7$, $\sigma = 5$ Å, various values of λ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 500 molecules.

(a)

(b)
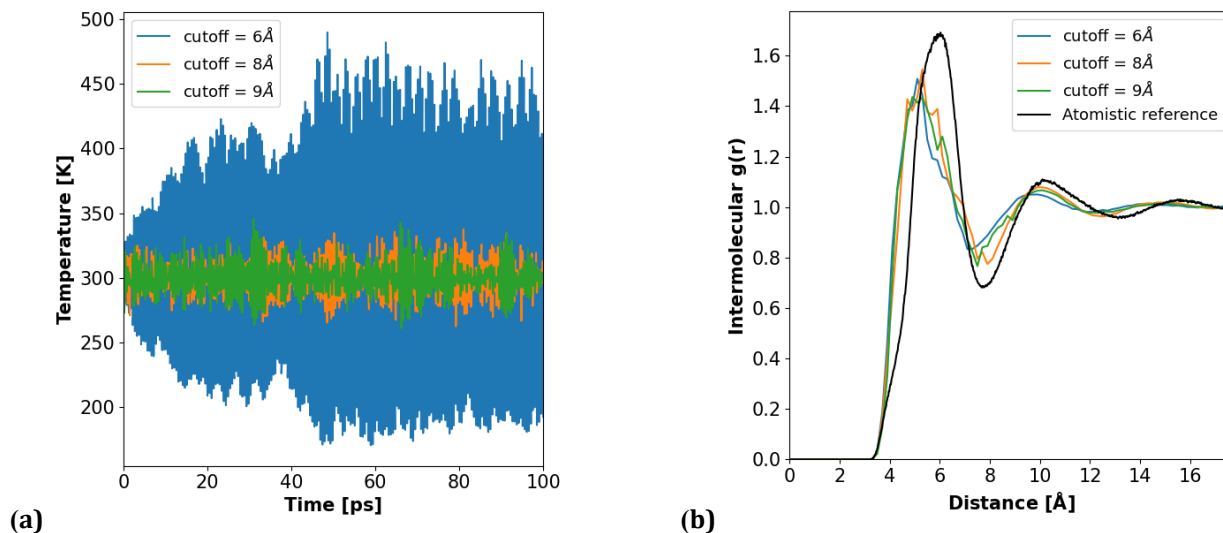
Figure 3.2. Comparison of the (a) temperature trend and (b) radial distribution function during ML-based CG NVT MD simulations of liquid benzene mapped using 1 CG bead per molecule, performed at 300K and atmospheric pressure with GCNN potentials trained with different values of the *cutoff* radius of 6Å, 8Å and 9Å for $n = 6$, $\sigma = 5$ Å, $\lambda = 1$, various values of *cutoff*, $T = 300$K.

Figure 3.3 depicts an example of results of a satisfactory CG model, that were obtained by a model for the training of which the loss function defined in eq.(6) was used, and *cutoff* = 8 Å, n = 7, σ = 5 Å, λ = 10^{-1}, is presented. The corresponding CG simulation was thermodynamically stable and reasonable results were observed in terms of dynamical behavior and of structural characteristics, compared to the atomistic ground truth.
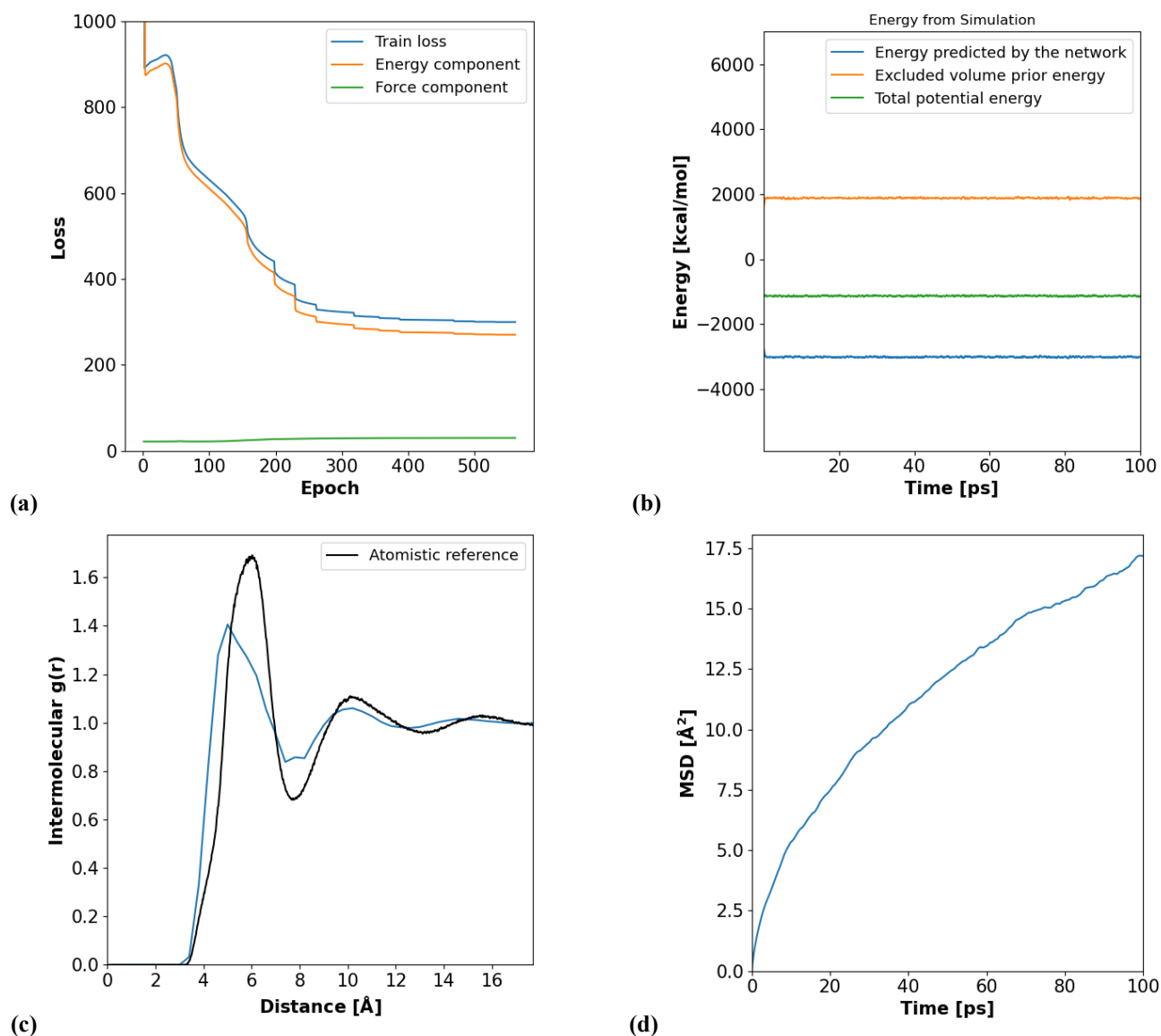

(a)

(b)

(c)

(d)

Figure 3.3. Example of a well-behaved model trained considering both force and energy differences in the loss function for liquid benzene mapped using 1 CG bead per molecule. (a) Trends of the loss components and total values over the training set. (b) Energy values calculated during the CG NVT MD simulation performed at T = 300 K and p = 1atm with the GCNN potential. (c) Mean Squared Displacement of the centers of mass of the particles. (d) Radial distribution for the centers of mass of the particles.

In Figure 3.4., the system size effect for the single-bead representation is shown on the values of the energy per particle. A model that was trained on data from a 500 benzene molecules system is used for a simulation of a system of 200 benzene molecules and another one of 1000 benzene molecules. It is apparent that the results of the initial system are reproduced also for the other systems.
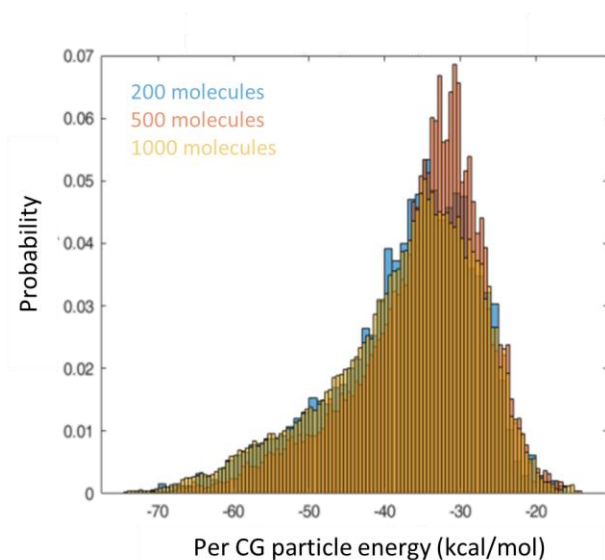

Figure 3.4. System size effect for the single-bead representation on "Energy per particle" values.

3.3 Three-bead representation

In total, more than 150 tests were conducted with different hyperparameter combinations in order to investigate the effect that certain hyperparameters have onto the training and the performance of the obtained models. The hyperparameter range is investigated is summarized in Table 3.2 and specific cases are presented in the analysis that follows.

.

Table 3.2. SchNet hyperparameter ranges investigated for the tests with liquid benzene, for the cases of the 3-CG-bead-per-molecule representation

Hyperparameter	Values	Hyperparameter	Values
Excluded volume radius (σ) [Å]	0.9735, 2, 2.1667, 2.334, 2.5, 3.5, 4	Number of intermolecular SchNet blocks	2, 3, 4
Excluded volume exponent (n)	5, 6, 7, 8	Number of intramolecular SchNet blocks	0, 1
Local neighborhood cutoff radius [Å]	6, 7	Bond stiffness k_b [$\frac{kcal}{mol\,Å^2}$]	10, 15, 20, 50, 100, 400, 1000
Energy scaling in the loss function (λ)	0, 10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10		

In the case of the three-bead representation, intramolecular interactions are also present and a bond prior term (see eq.(5)) was also included, with the addition of the hyperparameter k of effective CG bond stiffness. In Figures 3.5. and 3.6., results are presented from ML CG simulations with a model trained with the hyperparameter combinations $\sigma = 2$ Å, *cutoff* $= 6$ Å, $\lambda = 10^{-1}$ and $n = 6$. The conducted simulations were thermodynamically stable and the radial distributions functions between benzene molecules centers-of-mass exhibit a reasonable behavior in comparison to the atomistic reference (Fig. 3.6a). Changing the k does not seem to significantly affect the obtained results for this set of hyperparameters. Figure 3.6b depicts the intermolecular radial distribution function in which, though, an unphysical first peak at around ~2.3 Å is observed. To further examine this effect, hyperparameters related to the excluded volume prior term were investigated.
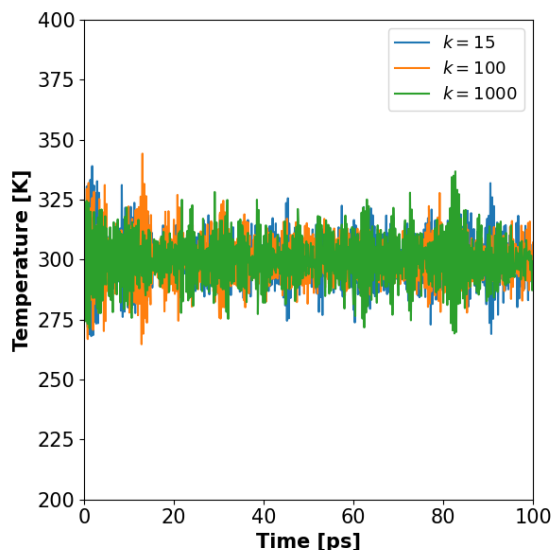
Figure 3.5. Comparative graphs of the evolution of the temperature during the ML-based CG simulation for *cutoff* $= 6$ Å, $\sigma = 2$ Å, $\lambda = 10^{-1}$, $n = 6$, various values of k, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.
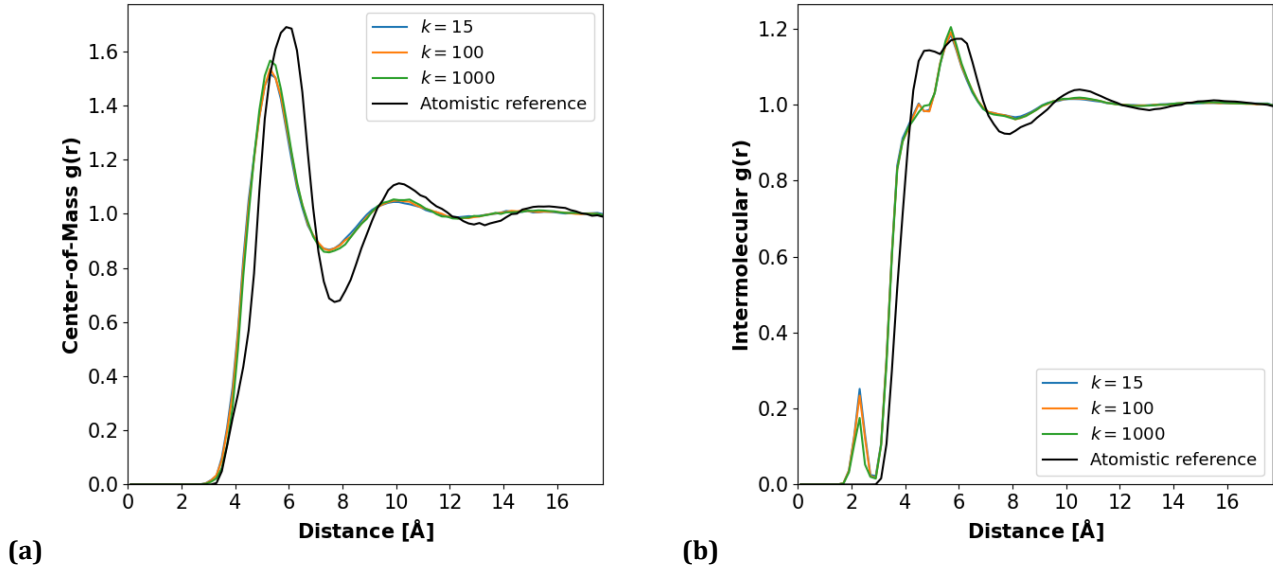
Figure 3.6. (a) Center-of-mass radial distribution functions and (b) intermolecular radial distribution functions of various systems (colored) with the respective atomistic ground truths (black) for $cutoff = 6$ Å, $\sigma = 2$ Å, $\lambda = 10^{-1}$, $n = 6$, various values of k, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.

In Figures 3.7. and 3.8., results are presented for the hyperparameter combinations $\sigma = 2$ Å, $cutoff = 6$ Å, $k = 1000$, $\lambda = 10^{-1}$ and various values of the excluded volume exponent n. The radial distribution between the benzene molecules' centers of masses is shown (Fig. 3.8(b)) in comparison to the atomistic reference. In Figure 3.8(a) the intermolecular radial distribution functions are shown in which the unphysical first peak at short distances remains unaffected in the range values of the excluded volume exponent that were tested.
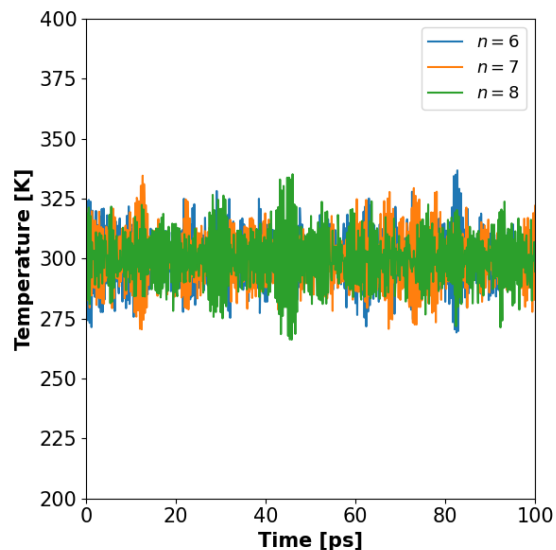
Figure 3.7. Comparative graphs of the evolution of the temperature during the ML-based CG simulation for *cutoff* = 6Å, $\sigma = 2$, $\lambda = 10^{-1}$, various values of n, $T = 300$K, $p = 1$atm, for a benzene liquid system of 300 molecules.
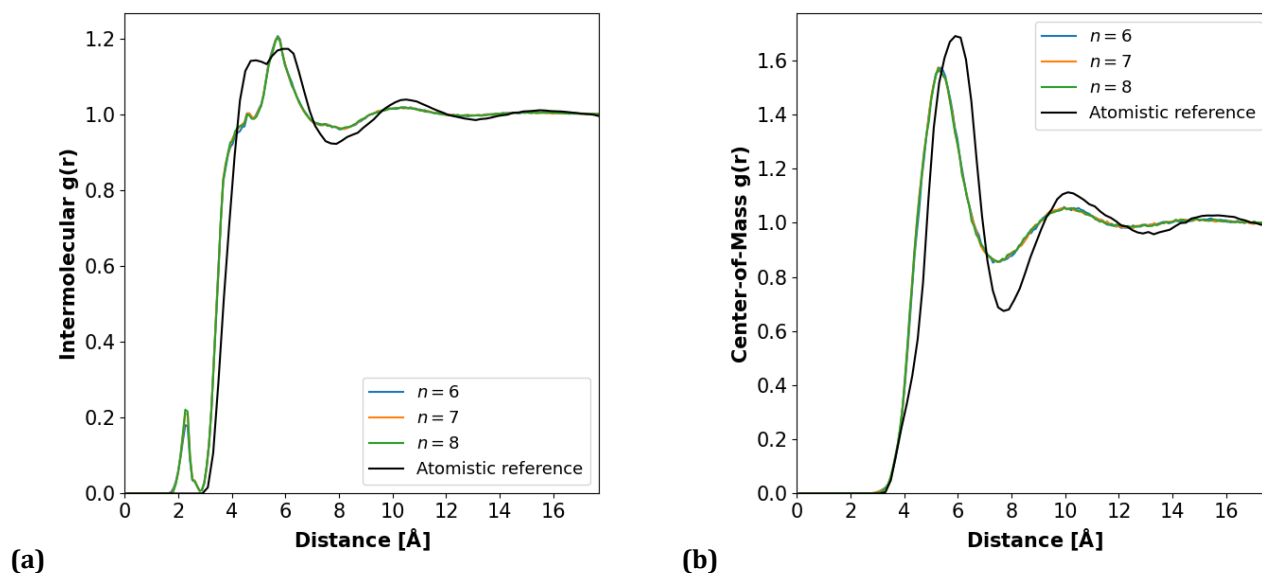

(a)

(b)

Figure 3.8. (a) Comparison between the intermolecular radial distribution functions and the center-of-mass radial distribution functions from ML CG simulations of the various models (colored) with the respective atomistic ground truths (black) *cutoff* = 6Å, $\sigma = 2$, $\lambda = 10^{-1}$, various values of n, $T = 300$K, $p = 1$atm, for a benzene liquid system of 300 molecules.

In Figures 3.9. and 3.10., ML CG simulations results are shown for models trained with *cutoff* = 6 Å, $n = 6$, $\lambda = 10^{-1}$, $k = 1000$ and various values of σ. From these comparative graphs we observe a monotonic behavior as far as the height of the first unphysical peak in the intermolecular g(r) graph is concerned, with respect to the value of σ. More specifically, as the value of σ increases, the height of the

peak becomes smaller and disappears for $\sigma \geq 3$ Å (Fig. 3.10(a)). Simultaneously, in the center-of-mass radial distribution functions (Fig. 3.10(b)) a more intense peak is also observed for the higher σ values along with a slight movement of the peak to right, improving the agreement with the atomistic reference.
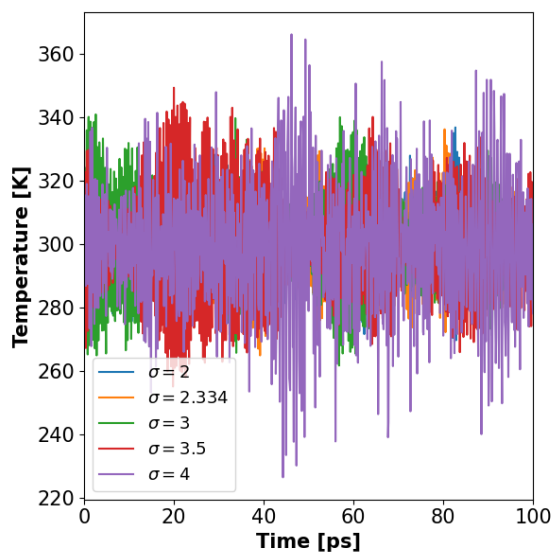

Figure 3.9. Comparative graphs of the evolution of the temperature during the ML-based CG simulation for *cutoff* = 6 Å, n = 6, $\lambda = 10^{-1}$, various values of σ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.
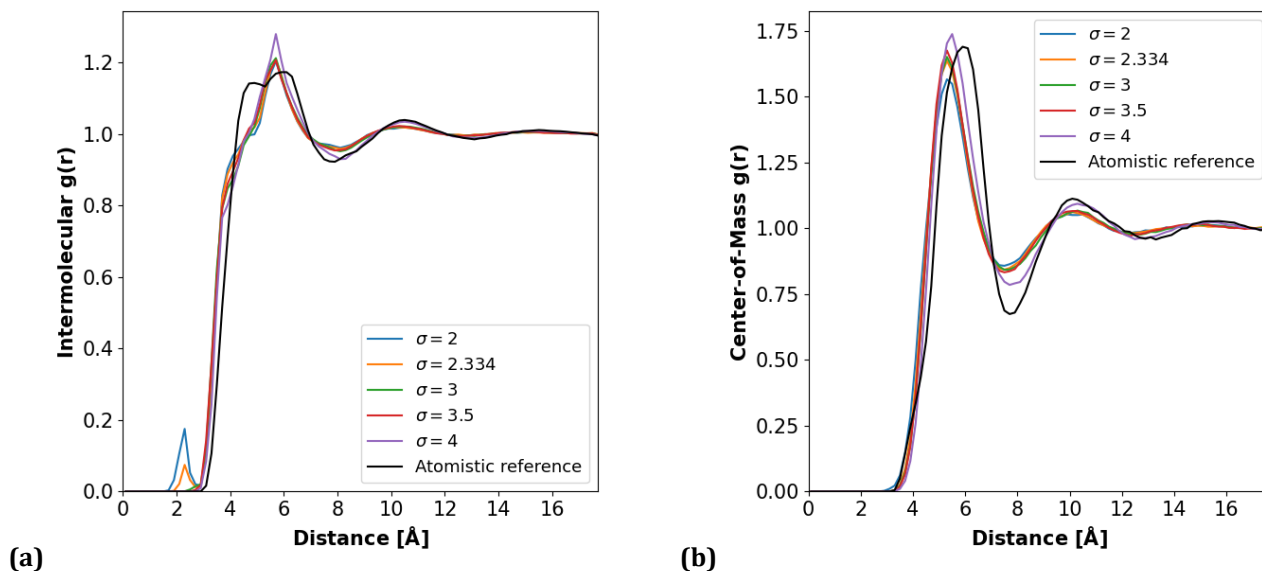

(a)

(b)

Figure 3.10. (a) Comparison between the intermolecular radial distribution functions and the center-of-mass radial distribution functions of ML CG simulations of various models (colored) with the respective atomistic ground truth (black) for *cutoff* = 6 Å, n = 6, $\lambda = 10^{-1}$, various values of σ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.

Figures 3.11. and 3.12., depict results for the hyperparameter combinations $\sigma = 3$, *cutoff* = 6 Å, $n = 6$ and $k = 1000$ and varying values of the loss energy component weighting factor λ. The temperature evolution is shown in Fig. 3.11 which depicts unstable simulations for the case of $\lambda \leq 10^{-2}$. For $\lambda = 0$, no energy component is considered in loss function which includes solely the mean squared difference between the predicted forces and the forces associated with the atomistic MD configurations. Very small effects are observed on the structural characteristics of the various systems.
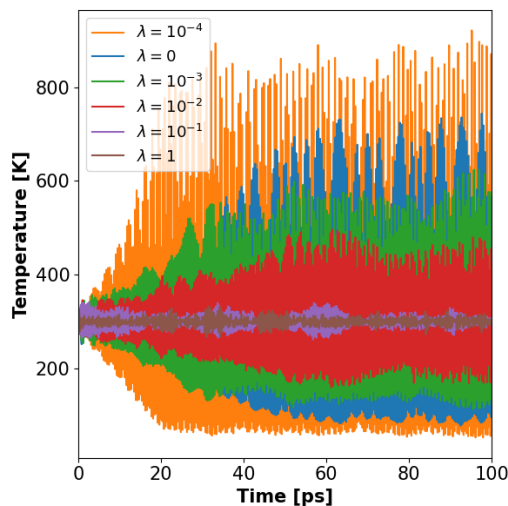

Figure 3.11. Comparative graphs of the evolution of the temperature during the ML-based CG simulation for *cutoff* = 6 Å, $\sigma = 3$, $n = 6$, $k = 1000$, various values of λ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.
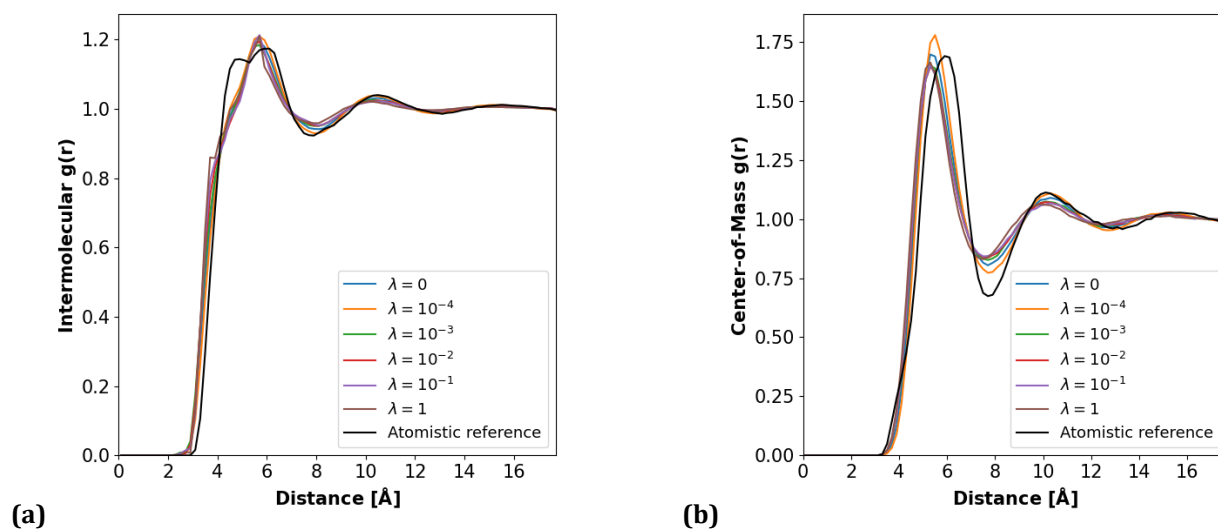

(a)

(b)

Figure 3.12. (a) Comparison between the inter-molecular radial distribution functions and (b) the center-of-mass radial distribution functions from the ML CG MD simulation using the various models (colored) with the respective atomistic ground truth (black) *cutoff* = 6Å, $\sigma = 3$, $n = 6$, $k = 1000$, various values of λ, $T = 300$K, $p = 1$atm, for a benzene liquid system of 300 molecules.

In Figures 3.13. and 3.14., ML CG MD simulation results are presented varying again the values of the loss energy component weighting factor λ and for the hyperparameter combinations $\sigma = 4$ Å, $cutoff = 6$ Å, $n = 6$ and $k = 1000$. In this case, also it is depicted that incorporation of the energy component in the loss function is crucial for obtaining a stable simulation. The hyperparameter combinations presented here (for $\lambda \neq 0$) apart from stable simulations exhibit good agreement in terms structure for the center-of-mass radial distribution function while the case of $\lambda = 1$ seems to be able to capture better the double peak that is observed in the inter-molecular radial distribution function.
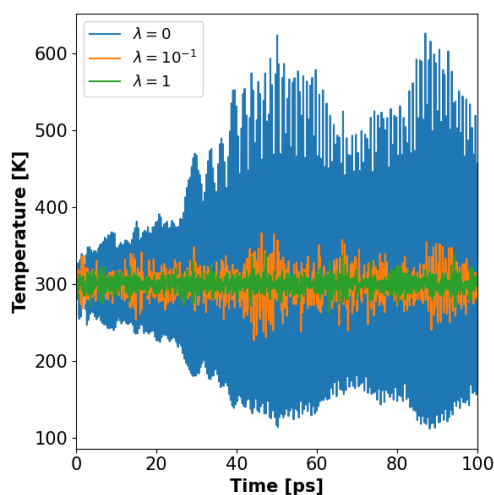

Figure 3.13. Comparative graphs of the evolution of the temperature during the ML-based CG simulation for $cutoff = 6$ Å, $\sigma = 4$ Å, $n = 6$, $k = 1000$, various values of λ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.
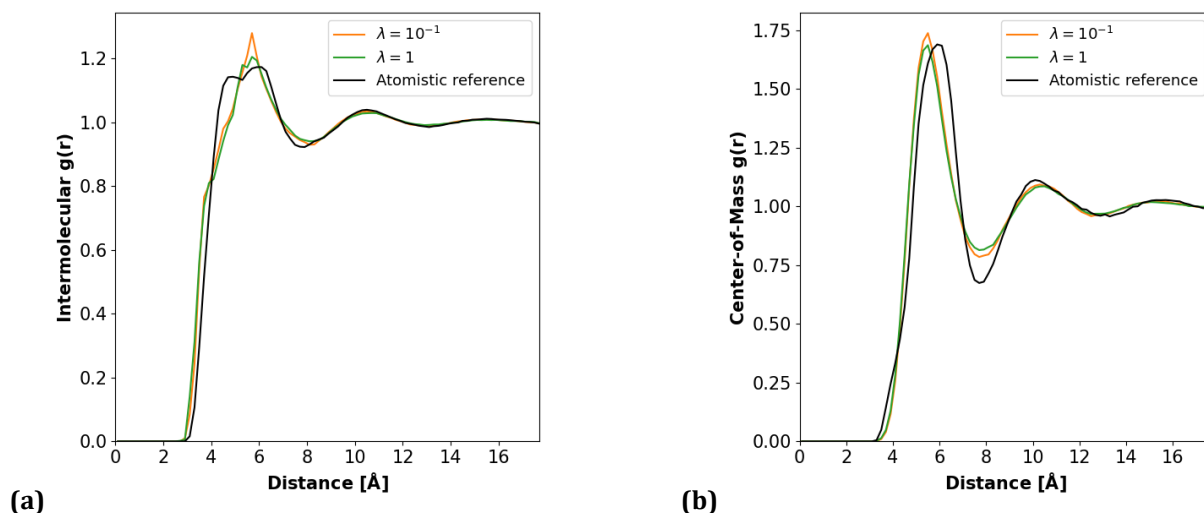

Figure 3.14. (a) Comparison among the intermolecular radial distribution functions [left] and the center-of-mass radial distribution functions [right] of several systems (colored) with the respective atomistic ground truths (black) $cutoff = 6$ Å, $\sigma = 3$ Å, $n = 6$, $k = 1000$, various values of λ, $T = 300$ K, $p = 1$atm, for a benzene liquid system of 300 molecules.

4. Conclusions

In this study, the implementation of an ML-based scheme that utilizes a graph convolutional neural network architecture was investigated for development of coarse-grained interaction potentials for the simulation of bulk amorphous systems. Liquid benzene was chosen as the benchmark test system and two different CG mappings were applied for the CG representation of the benzene molecule: a single-bead representation, in which only intermolecular interactions between the CG moieties are present, and a three-bead one. The procedure following involved training on data obtained from atomistic MD simulations of the system under study. A thorough investigation was conducted on the effect of various hyperparameters and loss construction on the training and the performance of the obtained models. The single-bead representation models were found to be generally more sensitive to changes in the values of the hyperparameters than the three-bead representation models. The efficiency of the developed ML-based models was evaluated by performing CG MD simulation to inspect the stability of the simulation and quantifying the dynamic behavior as well as the system structure in comparison to the atomistic reference, investigating also system size transferability prospects.

In this work several open challenges were detected and discussed for the CG simulations at bulk conditions using ML-based force fields, contributing to a deeper understanding in this still rather nascent domain. Unlike to several conventional applications of ML-methods, the convergence in the (currently used) loss functions during training, does not serve as a reliable evaluation metric of the performance of the developed models during CG simulations. In parallel, the potential need for multicomponent loss functions has been identified, a fact that also poses that need for appropriate balancing or mixing strategies. Future steps will also involve extension for the inclusion of directional characteristics in the description of local environments. The above aspects are crucial towards the scope of developing accurate, transferable and interpretable, higher-order CG potentials, that could be also streamlined with further developed and validated forward- and reverse-mapping ML-based schemes [29]–[31], for hierarchical multiscale simulations of bulk soft matter systems.

Acknowledgments

References

[1]   D. N. Theodorou, "Hierarchical modelling of polymeric materials," *Chem. Eng. Sci.*, vol. 62, no. 21, pp. 5697–5714, 2007, doi: 10.1016/j.ces.2007.04.048.

[2]   A. Gooneie, S. Schuschnigg, and C. Holzer, "A Review of Multiscale Computational Methods in Polymeric Materials," *Polymers (Basel).*, vol. 9, no. 12, p. 16, Jan. 2017, doi: 10.3390/polym9010016.

[3]   N. Vergadou and D. N. Theodorou, "Molecular Modeling Investigations of Sorption and Diffusion of Small Molecules in Glassy Polymers," *Membranes (Basel).*, vol. 9, no. 8, p. 98, Aug. 2019, doi: 10.3390/membranes9080098.

[4]   W. Noid *et al.*, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *J. Chem. Phys.*, vol. 128, p. 244114, 2008, doi: 10.1063/1.2938860.

[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*, vol. 9, no. 7, pp. 2108–2119, Jul. 2013, doi: 10.1039/C2SM27201F.

[6]   S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *J. Phys. Chem. B*, vol. 111, no. 27, pp. 7812–7824, Jul. 2007, doi: 10.1021/jp071097f.

[7]   S. J. Marrink, L. Monticelli, M. N. Melo, R. Alessandri, D. P. Tieleman, and P. C. T. Souza, "Two decades of Martini: Better beads, broader scope," *WIREs Comput. Mol. Sci.*, vol. 13, no. 1, Jan. 2023, doi: 10.1002/wcms.1620.

[8]   K. A. Maerzke and J. I. Siepmann, "Transferable Potentials for Phase Equilibria−Coarse-Grain Description for Linear Alkanes," *J. Phys. Chem. B*, vol. 115, no. 13, pp. 3452–3465, Apr. 2011, doi: 10.1021/jp1063935.

[9]   B. M. Mognetti *et al.*, "Coarse-grained models for fluids and their mixtures: Comparison of Monte Carlo studies of their phase behavior with perturbation theory and experiment," *J. Chem. Phys.*, vol. 130, no. 4, p. 044101, Jan. 2009, doi: 10.1063/1.3050353.

[10]   M. G. Martin and J. I. Siepmann, "Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes," *J. Phys. Chem. B*, vol. 103, no. 21, pp. 4508–4517, May 1999, doi: 10.1021/jp984742e.

[11]   A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Phys. Rev. E*, vol. 52, no. 4, pp. 3730–3737, Oct. 1995, doi: 10.1103/PhysRevE.52.3730.

[12]   D. Reith, M. Puetz, and F. Mueller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," *J. Comput. Chem.*, vol. 24, pp. 1624–1636, Nov. 2003.

[13]   O. T. Unke *et al.*, "Machine Learning Force Fields," *Chem. Rev.*, vol. 121, no. 16, pp. 10142–10186, Aug. 2021, doi: 10.1021/acs.chemrev.0c01111.

[14]   E. Ricci and N. Vergadou, "Integrating Machine Learning in the Coarse-Grained Molecular Simulation of Polymers," *J. Phys. Chem. B*, vol. 127, no. 11, pp. 2302–2322, Mar. 2023, doi: 10.1021/acs.jpcb.2c06354.

[15]   K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 992–1002. Accessed: Dec. 20, 2023. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/3294771.3294866

[16]   E. Ricci, G. Giannakopoulos, V. Karkaletsis, D. N. Theodorou, and N. Vergadou, "Developing Machine-Learned Potentials for Coarse-Grained Molecular Simulations: Challenges and Pitfalls," in *Proceedings of the 12th Hellenic*

Conference on Artificial Intelligence, New York, NY, USA: ACM, Sep. 2022, pp. 1–6. doi: 10.1145/3549737.3549793.

[17] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," *Europhys. Lett.*, vol. 26, no. 8, pp. 583–588, Jun. 1994, doi: 10.1209/0295-5075/26/8/005.

[18] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – A deep learning architecture for molecules and materials," *J. Chem. Phys.*, vol. 148, no. 24, p. 241722, Jun. 2018, doi: 10.1063/1.5019779.

[19] O. T. Unke and M. Meuwly, "PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges," *J. Chem. Theory Comput.*, vol. 15, no. 6, pp. 3678–3693, Jun. 2019, doi: 10.1021/acs.jctc.9b00181.

[20] S. Batzner *et al.*, "E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials," Jan. 2021, doi: 2101.03164.

[21] Y. Shao, M. Hellström, P. D. Mitev, L. Knijff, and C. Zhang, "PiNN: A Python Library for Building Atomic Neural Networks of Molecules and Materials," *J. Chem. Inf. Model.*, vol. 60, no. 3, pp. 1184–1193, Mar. 2020, doi: 10.1021/acs.jcim.9b00994.

[22] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," *Sci. Adv.*, vol. 3, no. 5, May 2017, doi: 10.1126/sciadv.1603015.

[23] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, "SchNetPack: A Deep Learning Toolbox For Atomistic Systems," *J. Chem. Theory Comput.*, vol. 15, no. 1, pp. 448–455, Jan. 2019, doi: 10.1021/acs.jctc.8b00908.

[24] J. Ruza, W. Wang, D. Schwalbe-Koda, S. Axelrod, W. H. Harris, and R. Gómez-Bombarelli, "Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks," *J. Chem. Phys.*, vol. 153, no. 16, p. 164501, Oct. 2020, doi: 10.1063/5.0022431.

[25] B. E. Husic *et al.*, "Coarse graining molecular dynamics with graph neural networks," *J. Chem. Phys.*, vol. 153, no. 19, 2020, doi: 10.1063/5.0026133.

[26] W. D. Cornell *et al.*, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995, doi: 10.1021/ja00124a002.

[27] S. Plimpton, "Fast Parallel Algorithms for Short–Range Molecular Dynamics," *J. Comput. Phys.*, vol. 117, no. 1, pp. 1–19, 1995.

[28] A. Hjorth Larsen *et al.*, "The atomic simulation environment—a Python library for working with atoms," *J. Phys. Condens. Matter*, vol. 29, no. 27, p. 273002, 2017, doi: 10.1088/1361-648X/aa680e.

[29] W. Wang and R. Gómez-Bombarelli, "Coarse-graining auto-encoders for molecular dynamics," *npj Comput. Mater.*, vol. 5, no. 1, p. 125, Dec. 2019, doi: 10.1038/s41524-019-0261-5.

[30] D. Nasikas, E. Ricci, G. Giannakopoulos, V. Karkaletsis, D. N. Theodorou, and N. Vergadou, "Investigation of Machine Learning-based Coarse-Grained Mapping Schemes for Organic Molecules," in *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, New York, NY, USA: ACM, Sep. 2022, pp. 1–8. doi: 10.1145/3549737.3549792.

[31] E. Christofi *et al.*, "Deep convolutional neural networks for generating atomistic configurations of multi-component macromolecules from coarse-grained models," *J. Chem. Phys.*, vol. 157, no. 18, p. 184903, Nov. 2022, doi: 10.1063/5.0110322.