The Journal of Chemical Physics

Bayesian calibration of force-fields from experimental data: TIP4P water

Cite as: J. Chem. Phys. **149**, 154110 (2018); https://doi.org/10.1063/1.5030950
Submitted: 27 March 2018 . Accepted: 27 September 2018 . Published Online: 18 October 2018

Ritabrata Dutta, Zacharias Faidon Brotzakis, and Antonietta Mira

COLLECTIONS

Paper published as part of the special topic on Enhanced Sampling for Molecular Systems
Note: This article was intended as part of the Special Topic "Enhanced Sampling for Molecular Systems" in Issue 7 of Volume 149 of J. Chem. Phys.

View Online Export Citation CrossMark

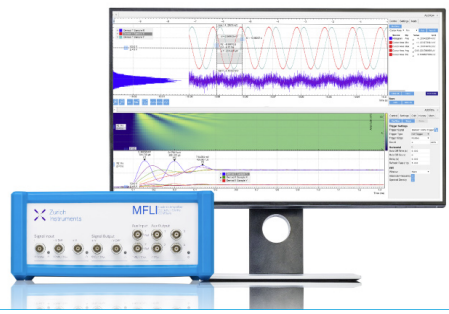
ARTICLES YOU MAY BE INTERESTED IN

Adaptive coupling of a deep neural network potential to a classical force field
The Journal of Chemical Physics **149**, 154107 (2018); https://doi.org/10.1063/1.5042714

Perspective: Computational chemistry software and its advancement as illustrated through three grand challenge cases for molecular science
The Journal of Chemical Physics **149**, 180901 (2018); https://doi.org/10.1063/1.5052551

Perspective: Identification of collective variables and metastable states of protein dynamics
The Journal of Chemical Physics **149**, 150901 (2018); https://doi.org/10.1063/1.5049637

Challenge us.
What are your needs for periodic signal detection?
Zurich Instruments
Watch



AIP Publishing

© 2018 Author(s).

Bayesian calibration of force-fields from experimental data: TIP4P water

Ritabrata Dutta,[1,a)] Zacharias Faidon Brotzakis,[1,2] and Antonietta Mira[1,3]
[1]*Institute of Computational Science, Università della Svizzera Italiana, Lugano, Switzerland*
[2]*Department of Chemistry and Applied Bioscience, ETH Zürich, Zürich, Switzerland*
[3]*Department of Science and High Technology, Università degli Studi dell'Insubria, Varese, Italy*

(Received 27 March 2018; accepted 27 September 2018; published online 18 October 2018)

Molecular dynamics (MD) simulations give access to equilibrium structures and dynamic properties given an ergodic sampling and an accurate force-field. The force-field parameters are calibrated to reproduce properties measured by experiments or simulations. The main contribution of this paper is an approximate Bayesian framework for the calibration and uncertainty quantification of the force-field parameters, without assuming parameter uncertainty to be Gaussian. To this aim, since the likelihood function of the MD simulation models is intractable in the absence of Gaussianity assumption, we use a likelihood-free inference scheme known as approximate Bayesian computation (ABC) and propose an adaptive population Monte Carlo ABC algorithm, which is illustrated to converge faster and scales better than the previously used ABCsubsim algorithm for the calibration of the force-field of a helium system. The second contribution is the adaptation of ABC algorithms for High Performance Computing to MD simulations within the *Python* ecosystem *ABCpy*. This adaptation includes a novel use of a dynamic allocation scheme for Message Passing Interface (MPI). We illustrate the performance of the developed methodology to learn posterior distribution and Bayesian estimates of Lennard-Jones force-field parameters of helium and the TIP4P system of water implemented for both *simulated* and *experimental datasets* collected using neutron and X-ray diffraction. For simulated data, the Bayesian estimate is in close agreement with the true parameter value used to generate the dataset. For experimental as well as for simulated data, the Bayesian posterior distribution shows a strong correlation pattern between the force-field parameters. Providing an estimate of the entire posterior distribution, our methodology also allows us to perform the uncertainty quantification of model prediction. This research opens up the possibility to rigorously calibrate force-fields from available experimental datasets of any structural and dynamic property. *Published by AIP Publishing.*
https://doi.org/10.1063/1.5030950

I. INTRODUCTION

In the last decades, molecular simulations have become a cornerstone for computing the equilibrium and/or dynamic properties of classical many body systems as well for bridging microscopic with macroscopic observables that could be of both experimental and basic importance.[1–3] Given a force-field formalism, in this paper we assume that the phase space can be ergodically explored by evolving Newton's equation of motion under a molecular mechanics force-field of interactions. For this purpose, we consider Molecular Dynamics (MD) simulations, which sample the phase space by integrating the deterministic Newton's equations of motion, hence giving access to both dynamic and thermodynamic properties. The accuracy of the underlying molecular mechanics force-field used to solve the equations of motion defines the approximation in the phase space exploration.

Each force-field reproduces specific properties and is indexed by a set of parameters, whose values are unknown (e.g., bonded and non-bonded force-field parameters). If the parameters of the force-field could be rigorously learned with an automated data-driven methodology, we could reparametrize the force-field for different experimentally obtained target properties (e.g., radial distribution functions, self-diffusion coefficient, density, etc.). There exist different force-field formalisms, constraining ourself only to water we have TIP3P, TIP4P,[4] TIP4P/2005,[5] and TIP5P,[6] among others. This raises the need for calibrations and force-field comparisons in a data-driven manner. Given an experimentally observed dataset, the Bayesian inferential framework can address both questions, namely, calibrating the parameters of a given force-field formalism and, given many formalisms, choosing one. This is achieved, correspondingly, through Bayesian parameter inference and model selection.[7] In this paper, we only focus on the former.

The underlying uncertainties in MD simulations can be classified into four main categories:[8] (1) *Modeling uncertainty* due to the specific choice of the model (e.g., a specific choice of the functional form for the force-field). (2) *Parametric uncertainty* due to the unknown values of the set of parameters defining the selected model (e.g., the parameters of the selected force-field function). (3) *Computational uncertainty* due to the particular computational setup (e.g., the systematic error due to the box size, fixed number of molecules, finite sampling time, and solution of Newton's equation by time-integrators) or due to the stochastic components in the computational model

Note: This article was intended as part of the Special Topic "Enhanced Sampling for Molecular Systems" in Issue 7 of Volume 149 of J. Chem. Phys.
[a)]Author to whom correspondence should be addressed: duttar@usi.ch

(e.g., stochastic thermostats). (4) *Measurement uncertainty* due to the experimental or observational error, only occurring while calibration of a force-field is performed based on an experimentally observed dataset.

The use of Bayesian inference for MD simulations has a long history;[8–17] for a recent review and a comparative study of different inferential approaches, we direct the readers to Ref. 18. In most of these studies, the different types of uncertainties are assumed to be Gaussian in nature. This provides us with a simple functional form for likelihood which in turn allows the use of standard Bayesian tools for inference and calibration of force-fields. In the ergodic limit (where the entire state space is explored), and if there is no bias due to under-sampling from the Molecular dynamics simulation [e.g., using GROningen MAchine for Chemical Simulations (GROMACS)[19] or Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[20]], approximating the uncertainty as Gaussian is theoretically meaningful. But for real life systems, the state space is explored very slowly due to the existence of large free energy barriers; hence a Gaussian model of uncertainty may be a poor approximation. Furthermore, the Gaussianity assumption does not hold when one wants to calibrate force-fields based on structural and dynamical properties with non-Gaussian data distribution (e.g., the radial distribution functions and the self-diffusion coefficient) or the partition function of the NPT-ensemble, which is known to be non-Gaussian due to its thermodynamics definition.[21]

If we do not assume Gaussianity, one of the main difficulties in applying Bayesian inference is the intractability of the likelihood functions of the force-field parameters for an observed dataset. In this paper, we solve this problem by approximate Bayesian computation (ABC),[22] a likelihood-free inference scheme recently developed in the field of statistical science. ABC has been used for the calibration of MD simulations in Ref. 15 under Gaussianity assumption and in Ref. 21 to calibrate parameters of Lennard-Jones (LJ) potential of helium without the assumption of Gaussian uncertainty. The authors of Ref. 18 concluded that most of the existing algorithms poorly quantify model prediction uncertainty, with ABC (under Gaussian assumption of uncertainty) being the most promising one. In Ref. 21, the ABC algorithm, more specifically ABCsubsim,[23] is proven to be equivalent to an algorithm that exploits the Gaussian assumption for one of their examples but is not able to further improve inferential results. This may be due to the use of ABC under Gaussian assumption of uncertainty[18] or due to the deficiencies in specific ABC algorithms.[21]

Following these studies, we propose the adaptive population Monte Carlo ABC (APMCABC) algorithm for the calibration of force-fields, without assuming any functional form for the likelihood, and illustrate its speed-up, faster convergence, and better estimate of the parameter values relative to the ABCsubsim algorithm used in Ref. 21 for the calibration of Lennard-Jones potential of helium (Secs. III E and IV A). As noted in Ref. 21, High Performance Computing (HPC) and parallelization are essential in the context of MD due to the long simulation time and chaotic and unstable behavior of more realistic and challenging force-field formalisms. This need is addressed thanks to our recently developed *Python* eco-system called ABCpy,[24] which implements and parallelizes most of the existing ABC algorithms and adapts them optimally to the HPC infrastructure. Furthermore, to mitigate the imbalance in simulation times of different MD simulation models for different values of force-field parameters, we use a new dynamic allocation scheme for Message Passing Interface (MPI), developed in Ref. 25.

Having found that the APMCABC is more efficient than the ABCsubsim algorithm to calibrate the force-field of helium, where no Gaussianity assumption is made, we pursue using it to calibrate the force-field of a more complicated system such as water. In this paper, we only address the issue of calibrating the parameters given a force-field formalism, under the assumption of no measurement uncertainty. For this purpose, we consider a force-field formalism of water, specifically the rigid non-polarizable TIP4P force-field developed by Ref. 4, with all bonds and angles constrained using the LINCS algorithm mentioned in Sec. 2. Water is universal in science and life. From protein folding[26] to ion mobility[27] and from enzymatic activity[28] to anti-freeze activity,[29] local and/or global water structure and dynamics regulate biological and physicochemical processes. Hence, force-field parametrization by targeting the structural or dynamical properties of water is of great impact in biology and science. We illustrate the performance of the inferential scheme for the simulated dataset, generated using GROMACS, and for experimental datasets, obtained under ambient conditions using neutron diffraction[30] and X-ray diffraction.[31]

In Sec. II, we explain the Lennard-Jones potential of helium and TIP4P force-field formalism of water and describe the models used to forward-simulate the dataset using LAMMPS and GROMACS, correspondingly. A short introduction to Bayesian inference and approximate Bayesian computation used for inferring the force-field parameters is given in Sec. III. We infer the posterior distribution of the parameters and their Bayes estimates, given a simulated or experimentally obtained dataset, in Sec. IV. As the proposed Bayesian inferential scheme provides us with a posterior distribution of the parameters, we can quantify the model prediction uncertainty by simulating the dataset from the MD simulation model using values of the parameters randomly drawn from the inferred posterior distribution. Finally, in Sec. V, we illustrate the ability to quantify prediction uncertainty by our inferential scheme for the experimental dataset of water, collected using neutron and X-ray diffraction technology.

II. FORCE-FIELDS

A. Lennard-Jones potential of helium

We first consider the calibration of the parameters for 6-12 Lennard-Jones potential of helium. The potential is given by

$$V_{LJ}(\sigma_{LJ}, \epsilon_{LJ}) = \sum_i \sum_j 4\epsilon_{LJ}\left(\left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{6}\right), \quad (1)$$

where ϵ_{LJ} (zJ), σ_{LJ} (nm),[32] and r_{ij} (nm) are, correspondingly, the depth of the potential well, the finite distance at which the inter-particle potential is zero, and the distance

between the i and j particles. Using data, we want to calibrate parameters ϵ_{LJ} and σ_{LJ}. Thus, for the purpose of the present study, the LJ model of potential, which we denote by $\mathcal{M}_{LJ}$, is only parametrized in terms of the two non-bonded force-field parameters σ_{LJ} and ϵ_{LJ}, which we jointly denote with

$$\boldsymbol{\phi} = (\sigma_{LJ}, \epsilon_{LJ}).$$

If the initial values for coordinates of helium atoms in an enclosed space are known, we can forward simulate the coordinates of the atoms over time using model $\mathcal{M}_{LJ}$ for given values of these parameters $\boldsymbol{\phi} = \boldsymbol{\phi}^*$,

$$\mathcal{M}_{LJ}[\boldsymbol{\phi} = \boldsymbol{\phi}^*] \rightarrow \{(\mathbf{R}(t)) : \ t = 0, \ldots, t_{end}\}, \tag{2}$$

where $\mathbf{R}(t)$ are the position of the molecules at time t in an enclosed space. Following Ref. 21 and 33, the forward simulation in the NPT-ensemble of 300 K temperature and atmospheric pressure is performed using LAMMPS[20] in a simulation box of $27.3 \times 27.3 \times 27.3$ nm with 1000 helium atoms, for $t_0 = 0$, $t_{end} = 5$ ns, and time steps of 2 fs, after the system has been equilibrated for 2 ns. The damping frequency for temperature and pressure was set at 2 ps and 20 ps, respectively. Furthermore, the Lennard-Jones interaction cutoff and the mass of helium atom were fixed at 0.639 nm and 6.64×10^{-6} attogram, respectively.

B. TIP4P force-field of water

The TIP4P force-field[4] of water is a four interaction site water force-field with an extra charge placed on a dummy atom, besides the charges existing on the oxygen and hydrogens. It has been parametrized to reproduce the enthalpy of vaporization.[34] The formalism for water is shown in the following equation:

$$U_{noncov} = \sum_i \sum_j 4\epsilon_{TP}\left[\left(\frac{\sigma_{TP}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{TP}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_j \sum_\alpha \sum_\beta \frac{q_{i\alpha} q_{j\beta}}{r_{ij}}, \tag{3}$$

where the potential energy function U_{noncov} (kJ mol^{-1}) is the sum of 6-12 Lennard-Jones and electrostatic interactions. The r_{ij} (nm) term corresponds to the intermolecular distance between ith and jth water molecules. The Pauli repulsion and the van der Waals attraction are parametrized with terms $\propto r^{-12}$ and $\propto r^{-6}$, correspondingly. The Lennard-Jones interaction term includes σ_{TP} (nm) and ϵ_{TP} (kJ mol^{-1}) parameters, respectively, the distance at which the inter-particle potential is zero and the value of the minimum energy. Furthermore, we keep the bond length and angles fixed in the TIP4P rigid water model. Finally, the electrostatic term in Eq. (3) involves the charges of atoms, where i and j correspond to different water molecules and α and β are the indices of the partial charges q(e) of each molecule.

For the purpose of the present study, the TIP4P model of the force-field, which we denote by $\mathcal{M}_{TP}$, is parametrized in terms of the two non-bonded force-field parameters σ_{TP} and ϵ_{TP}, which we jointly denote with

$$\boldsymbol{\phi} = (\sigma_{TP}, \epsilon_{TP}),$$

representing the repulsion and attraction of the van der Waals forces. Although the proposed model contains additional parameters [e.g., the charges of Eq. (3)], we chose not to infer them and assume they are constant. However, we stress that the ABC method is not bound by the number of parameters to infer, and our choice is based on the illustrative purpose of this paper. If the initial values for coordinates of water molecules in an enclosed space are known, we can forward simulate the coordinates of the molecules over time using model $\mathcal{M}_{TP}$ for given values of these parameters $\boldsymbol{\phi} = \boldsymbol{\phi}^*$,

$$\mathcal{M}_{TP}[\boldsymbol{\phi} = \boldsymbol{\phi}^*] \rightarrow \{(\mathbf{R}(t)) : \ t = 0, \ldots, t_{end}\}, \tag{4}$$

where $\mathbf{R}(t)$ are the position of the molecules at time t in an enclosed space. The forward simulation is performed using MD (TIP4P implementation in GROMACS) for $t_0 = 0$, $t_{end} = 100$ ps, and time steps of 2 fs. We first construct a simulation box of $2.5 \times 2.5 \times 2.5$ nm and of 515 water molecules. Then, after compiling the TIP4P force-field with $\boldsymbol{\phi}^*$, we perform an energy minimization, followed by an NPT simulation. In the energy minimization step, we use the steepest descend algorithm for energy minimization.[35] We further use the LINCS algorithm[36] for the bond and angle constraints and the Particle Mesh Ewald (PME)[37] with a cutoff of 1 nm to treat the electrostatics. The cutoff for the van der Waals interactions is 1 nm. In the NPT simulation, we set the temperature and the pressure at 298 K and 1 atm, correspondingly, using the stochastic velocity rescaling thermostat[38] and the Parrinello-Rahman barostat,[39] respectively. The neighbor list is updated every 10 time steps. The electrostatic and van der Walls parameters and the LINCS algorithm for the bond and angle constraints are treated in the same way as in the energy minimization.

III. INFERENCE FRAMEWORK

We use the term observed data, denoted by $\boldsymbol{x}^{\mathbf{0}}$, to refer to a dataset generated by some real-world process (e.g., experimental studies using X-ray and neutron diffraction that allow us to measure different properties of water molecules), and our goal is to learn values of the force-field parameters characterizing this process. Assuming that the first observation of the process occurs at time t_0 and the last at time t_{end}, the observed dataset is $\boldsymbol{x}^{\mathbf{0}} \equiv \{\mathbf{R}(t) : t_0, \ldots, t_{end}\}$. From experimental studies, it is not possible to track the time dependent position of water molecules, but we can learn their properties, e.g., different radial distribution functions, using different diffraction techniques.

Here we address the question of calibrating force-fields for any available structural and dynamical properties, such as the radial distribution function and self-diffusion coefficient. To this aim, we develop an approximate Bayesian inference scheme that allows us to quantify the uncertainty in the inferred model parameters, uncertainty which is inherent to the inferential process given the chaotic nature of the models described in Eqs. (2) and (4). We only calibrate force-fields targeting structural and dynamical properties, but we stress that this methodology can be used for any property which is available experimentally.

A. Bayesian inference

We can quantify the uncertainty of the parameter $\boldsymbol{\phi}$ by its posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x})$ given the observed dataset $\boldsymbol{x} = \boldsymbol{x^0}$. The posterior distribution is obtained by Bayes' theorem as

$$p(\boldsymbol{\phi}|\boldsymbol{x^0}) = \frac{\pi(\boldsymbol{\phi})p(\boldsymbol{x^0}|\boldsymbol{\phi})}{m(\boldsymbol{x^0})}, \tag{5}$$

where $\pi(\boldsymbol{\phi})$, $p(\boldsymbol{x^0}|\boldsymbol{\phi})$, and $m(\boldsymbol{x^0}) = \int \pi(\boldsymbol{\phi})p(\boldsymbol{x^0}|\boldsymbol{\phi})d\boldsymbol{\phi}$ are, correspondingly, the prior distribution on the parameter $\boldsymbol{\phi}$, the likelihood function, and the marginal likelihood. The prior distribution $\pi(\boldsymbol{\phi})$ enables us to incorporate, in the inferential process, prior knowledge on the parameter values. If the likelihood function could be evaluated, at least up to a normalizing constant, then the posterior distribution could be approximated by drawing a representative sample of parameter values from it using (Markov chain) Monte Carlo sampling schemes.[40] Unfortunately, the likelihood function induced by the TIP4P water and helium model is analytically intractable because we do not assume Gaussianity in the model. In this setting, approximate Bayesian computation (ABC)[22] offers a way to sample from the approximate posterior distribution and opens up the possibility of sound statistical inference on the parameter $\boldsymbol{\phi}$. In this paper, we only focus on parameter estimation/calibration and uncertainty quantification, but we stress that ABC also allows us to easily perform parameter hypothesis testing and model selection.

B. Approximate Bayesian computation (ABC)

Models that are easy to forward simulate, given values of the parameters, are called simulator-based models in the ABC literature and are used in a wide range of scientific disciplines to describe and understand different aspects of nature ranging from dynamics of sub-atomic particles[41] to evolution of human societies[42] and formation of universes.[43] In the fundamental rejection ABC sampling scheme, we simulate a synthetic dataset $\boldsymbol{x}^{\text{sim}}$ from the simulator-based model $\mathcal{M}(\boldsymbol{\phi})$ for a fixed parameter value, $\boldsymbol{\phi}$, and measure the closeness between $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x^0}$ using a pre-defined discrepancy measure $d(\boldsymbol{x}^{\text{sim}}, \boldsymbol{x^0})$. Based on this discrepancy, ABC accepts the parameter value $\boldsymbol{\phi}$ when $d(\boldsymbol{x}^{\text{sim}}, \boldsymbol{x^0})$ is less than a pre-specified threshold value δ.

The intractable likelihood $p(\boldsymbol{x^0}|\boldsymbol{\phi})$ is approximated by $p_{d,\delta}(\boldsymbol{x^0}|\boldsymbol{\phi})$ for some choice of distance, d, and threshold, $\delta > 0$, where

$$p_{d,\delta}(\boldsymbol{x^0}|\boldsymbol{\phi}) \propto P(d(\boldsymbol{x}^{\text{sim}}, \boldsymbol{x^0}) < \delta) \tag{6}$$

and, as a consequence, the sampled parameters follow the posterior distribution of $\boldsymbol{\phi}$ conditional on $d(\boldsymbol{x}^{\text{sim}}, \boldsymbol{x^0}) < \delta$,

$$p_{d,\delta}(\boldsymbol{\phi}|\boldsymbol{x^0}) \propto P(d(\boldsymbol{x}^{\text{sim}}, \boldsymbol{x^0}) < \delta)\pi(\boldsymbol{\phi}).$$

For a better approximation of the likelihood function, computationally efficient sequential ABC algorithms[7] decrease the value of the threshold δ adaptively while exploring the parameter space.

In this manuscript, we consider two sequential ABC algorithms: ABCsubsim[23] and APMCABC.[44] At the first step of these two algorithms, N_{sample}-many parameter values are randomly drawn from the prior distribution and the value of δ is decreased adaptively depending on the pseudodata simulated from the model using those randomly sampled parameter values. In the next step, they produce N_{sample}-many parameter values approximately distributed from the distribution $p_{d,\delta}(\boldsymbol{\phi}|\boldsymbol{x^0})$, for the adapted δ value from the last step and again decrease the δ depending on the new samples. This procedure is continued N_{step} many times or until some stopping criterion is reached. We note that the adapted δ values at each step are strictly decreasing and converge to zero, in turn improving the approximation to the posterior distribution. Here we denote the value of δ at the final step as δ_{final}.

C. Discrepancy, prior distribution, and perturbation kernel

The discrepancy measure between $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x^0}$ is often defined through a distance between summary statistics computed from $\boldsymbol{x}^{\text{sim}}$ and $\boldsymbol{x^0}$. The choice of these summary statistics is a crucial aspect for a good ABC approximation. The summary statistics are usually chosen to minimize the loss of information on $\boldsymbol{\phi}$ contained in the data and picking low-dimensional summaries to avoid the curse of dimensionality.[45] Here we use intuitive discrepancy measures between interpretable and domain-driven summary statistics. The subjectivity of these decisions can be removed through automatic summary selection for ABC, described in Refs. 45–48, where an informative linear or non-linear combination of the summaries is chosen. Next, we describe the summary statistics, the discrepancy measure, the prior distribution of the parameters, and the perturbation kernel, used for the calibration of the LJ potential and TIP4P force-field of helium and water, correspondingly. The discrepancy measure used for the LJ force-field of helium was previously used in Ref. 21.

1. LJ potential of helium

a. Summary statistics. Given a dataset $\boldsymbol{x} \equiv \{\mathbf{R}(t) : t = t_0, \ldots, t_{end}\}$ for LJ potential of helium simulated using LAMMPS, we compute the following summary statistics:

$$\mathcal{F}_{LJ} : \boldsymbol{x} \to (\mathbf{f}_B \equiv f_B(t) : t = t_0, \ldots, t_{end}),$$

where $f_B(t) = \langle \exp\{-H(t)/(k_B T)\} \rangle$, k_B is the Boltzmann constant, T is the temperature of the system, $H(t)$ is the enthalpy contribution of a helium atom in the system at time t, and $\langle\ \rangle$ denotes the ensemble average over all the atoms in the system at a time instance t.

b. Discrepancy measure. The discrepancy measure between two datasets $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ is constructed by considering the Kullback-Leibler divergence between the probability distribution functions $\chi^{(1)}$ and $\chi^{(2)}$ of $\mathbf{f}_B^{(1)}$ and $\mathbf{f}_B^{(2)}$ over time instances extracted from $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$,

$$\begin{aligned} d_{LJ}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) &:= d_{LJ}\left(\mathcal{F}_{LJ}(\boldsymbol{x}^{(1)}), \mathcal{F}_{LJ}(\boldsymbol{x}^{(2)})\right) \\ &:= d_{LJ}\left(\mathbf{f}_B^{(1)}, \mathbf{f}_B^{(2)}\right) \\ &= \int \chi^{(1)}(z) \log \frac{\chi^{(1)}(z)}{\chi^{(2)}(z)} dz. \end{aligned}$$

c. Prior distributions. For σ_{LJ} and ϵ_{LJ}, we use independent continuous uniform prior distributions on the range [0.1(nm), 0.8(nm)] and [0.01(zJ), 1.0(zJ)], respectively.

d. Perturbation kernel. To explore the parameter space of $\boldsymbol{\phi} = (\sigma_{LJ}, \epsilon_{LJ}) \in [0.1(\text{nm}), 0.8(\text{nm})] \times [0.01(\text{zJ}), 1.0(\text{zJ})]$, we consider a truncated two-dimensional multivariate Gaussian distribution on the above space as the perturbation kernel. APMCABC inference scheme centers the perturbation kernel at the parameter value it is perturbing and updates the variance-covariance matrix of the perturbation kernel based on the parameter values sampled from the previous step.

2. TIP4P force-filed of water

a. Summary statistics. Given a dataset $\boldsymbol{x} \equiv \{\mathbf{R}(t) : t = t_0, \ldots, t_{end}\}$ for the TIP4P model of water simulated using GROMACS, we compute an array of summary statistics

$$\mathcal{F}_{TP} : \boldsymbol{x} \to (S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9)$$

defined as follows:

- S_1: Estimate of the number of hydrogen bonds per water molecule—The area under the curve r_{OH} vs g_{OH} until the first minimum.
- S_2: Estimate of the donor acceptor hydrogen bond distance—The value of r_{OH} (nm) at the first minimum of the radial distribution function g_{OH}.
- S_3: Mean of g_{OH}.
- S_4: Estimate of the number of water molecules in the first hydration shell—The area under the curve r_{OO} vs g_{OO} until the first minimum.
- S_5: Estimate of the maximum distance of the first hydration shell—The value of r_{OO} (nm) at the first minimum of the radial distribution function g_{OO}.
- S_6: Mean of g_{OO}.
- S_7: The height of g_{OO} at the first maximum of g_{OO}.
- S_8: The value of r_{OO} (nm) at the first maximum of the radial distribution function g_{OO}.
- S_9: The slope of the line, fitted to (M)ean (S)quare (D)isplacement (MSD), which is an estimate of the 6× self-diffusion coefficient. The MSD is in units of $\text{nm}^2\ \text{N}^{-1}$.

To compute the above summary statistics, we first compute the radial distribution functions for the O–H and O–O atoms and the mean square displacement (MSD) from the simulated coordinates of the dynamical system. We then compute the above-mentioned summary statistics from the radial distribution functions and the MSD as explained in Fig. 1. The intuition behind choosing the above mentioned quantities is that they are broadly used characteristic quantities of the structure and dynamics of liquids.[1,2]

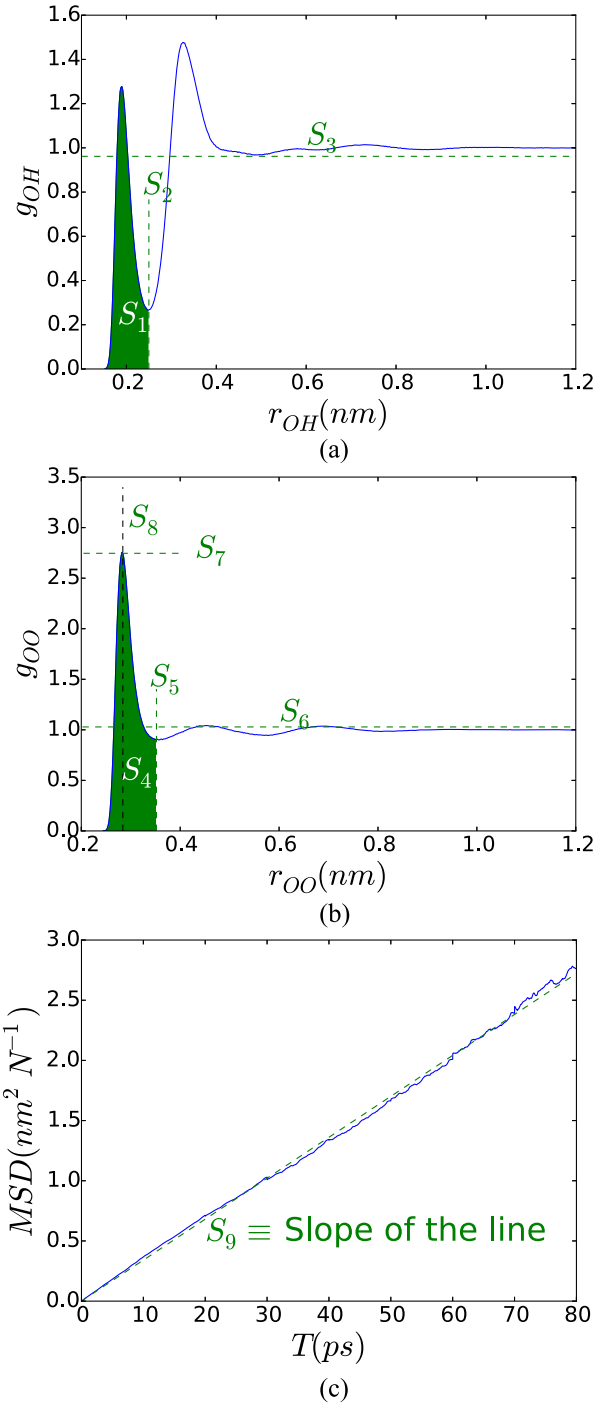


FIG. 1. **Summary statistics (water)**: $(S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9)$ are computed from the radial distribution functions for O–H (a), O–O (b) atoms, and the mean square displacement (c), generated from the simulated coordinates of the dynamical system.

b. Discrepancy measure. The discrepancy measure between two datasets $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ is constructed by considering the distance functions between the summary statistics extracted from them,

$$d_{TP}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := d_{TP}\left(\mathcal{F}_{TP}(\boldsymbol{x}^{(1)}), \mathcal{F}_{TP}(\boldsymbol{x}^{(2)})\right) = \frac{1}{9}\sum_{i=1}^{9} |S_i^{(1)} - S_i^{(2)}|.$$

c. Prior distributions. We use independent continuous uniform prior distributions on the range [0.281(nm), 0.53(nm)], and [0.2(kJ mol^{-1}), 0.9(kJ mol^{-1})], correspondingly, for σ_{TP} and ϵ_{TP}. Outside of this parameter range, the

TIP4P model of water in GROMACS becomes extremely chaotic and the simulated dataset cannot be obtained in a reasonable time span.

d. Perturbation kernel. The perturbation kernel used to explore the parameter space of $\boldsymbol{\phi} = (\sigma_{TP}, \epsilon_{TP}) \in [0.281(\text{nm}), 0.53(\text{nm})] \times [0.2(\text{kJ mol}^{-1}), 0.9(\text{kJ mol}^{-1})]$, is a truncated two-dimensional multivariate Gaussian distribution. APMCABC inference scheme centers the perturbation kernel at the parameter value it is perturbing and updates the variance-covariance matrix of the perturbation kernel based on the parameter values sampled in the previous step.

D. Time imbalance and dynamic allocation for MPI

For inference, we use the ABCpy *Python* package,[24] which implements some of the most advanced ABC algorithms with an optimal exploitation of an HPC environment. The parallelization schemes in ABCpy were primarily meant for inferring parameters from models which take almost equal time to simulate the dataset for any values of $\boldsymbol{\phi}$. Though, due to the chaotic nature of the MD systems, we observe in Fig. 2 that, for different values of $\boldsymbol{\phi}$, the simulation-time, for a fixed t_{end} value, is quite variable (with 36-core of Piz Daint Cray architecture: Intel Broadwell with NVidia TESLA P100). To solve this imbalance, we use a new dynamic allocation scheme for MPI, developed in Ref. 25.

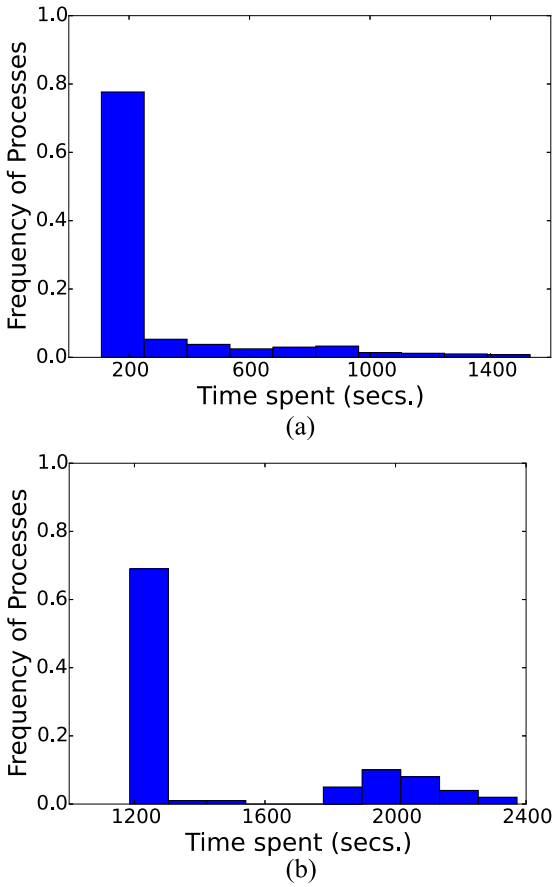


FIG. 2. Imbalance in time spent to simulate a pseudodata for different values of $\boldsymbol{\theta}$. (a) Lennard-Jones potential of helium and (b) TIP4P force-field of water.

Here, we briefly explain the dynamic work allocation strategy for map-reduce in comparison to a straightforward allocation approach. In the straightforward approach, the allocation scheme initially distributes m tasks to n executors, sends the map function to each executor, which in turn applies the map function, one after the other, to its m/n map tasks. This approach is visualized in Fig. 3(a), where a chunk represents the set of m/n map tasks. For example, if we want to draw 10 000 samples from the posterior distribution and we have $n = 100$ cores available, at each step of APMCABC we create groups of 100 parameters and each group is assigned to one individual core.

On the other hand, the dynamic allocation scheme initially distributes $k < m$ tasks to the k executors, sends the map function to each executor, which in turn applies it to the single task available. In contrast to the straightforward allocation, the executor requests a new map task as soon as the previous one is terminated. This clearly results in a better work balance. The dynamic allocation strategy is an implementation of the famous greedy algorithm for job-shop scheduling, which can be shown to have an overall processing time (makespan) up to twofold better than the best makespan.[49] This approach is illustrated in Fig. 3(b). The unbalancedness is not a problem that can be overcome easily by adding resources, rather speed-up and efficiency can drop drastically compared to the dynamic allocation strategy with increasing number of executors. For a detailed description and comparison, we direct readers to Ref. 25.

E. Efficient scale-up using adaptive population Monte Carlo ABC

Given the time imbalance and the high simulation cost of obtaining a pseudodataset from $\mathcal{M}_{TP}$, we need to choose an inference algorithm which converges fast with few numbers of pseudodata simulation and can thus better exploit the available computational resources. The ABCsubsim algorithm was used for the calibration of the Lennard-Jones

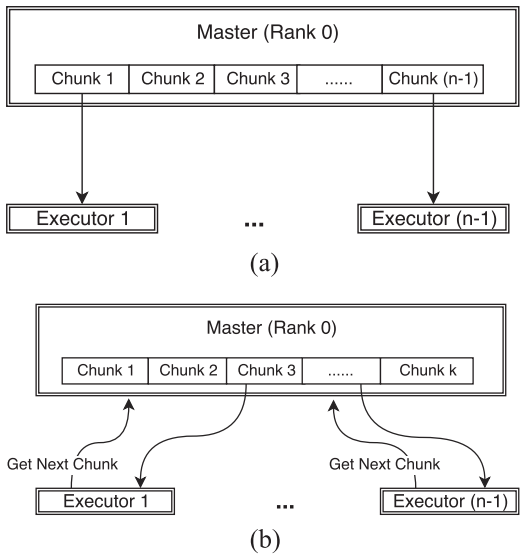


FIG. 3. Comparison of work-flow between the MPI (a) and dynamic-MPI (b) backend.

force-filed of helium in Ref. 21. Here, we propose instead to use the adaptive population Monte Carlo ABC (APMCABC)[44] algorithm.

In Fig. 4, we compare the performance of the ABCsubsim algorithm with the APMCABC algorithm, with reference to the speedups obtained while increasing the number of cores. A higher order scale-up in performance can be noted for the APMCABC algorithm. The discrepancy in speedup can be explained by the purely sequential nature of multiple Metropolis-Hastings chains used to update the samples at each step of ABCsubsim, compared to the purely parallel nature of APMCABC at each step, where the main job can be divided into N_{sample}-many parallel jobs.

Hence, in this paper, we choose APMCABC for the calibration of force-fields. Later in Sec. IV A, we also compare these two algorithms depending on their achieved threshold (δ) value and Bayes estimate after a fixed number of steps for the calibration of the helium system, illustrating a faster convergence gained by APMCABC. Thanks to APMCABC, we can approximate the posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x}^{\mathbf{0}})$ by drawing samples from it.

F. Parameter estimation

Our main goal is to estimate $\boldsymbol{\phi}$, given $\boldsymbol{x}^{\mathbf{0}}$. In decision theory, the Bayes estimator minimizes the posterior expected loss, $E_{p(\boldsymbol{\phi}|\boldsymbol{x}^{\mathbf{0}})}(\mathcal{L}(\boldsymbol{\phi},\cdot)|\boldsymbol{x}^{\mathbf{0}})$. Here we consider the following loss function:

$$\mathcal{L}(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2) := d_E(\phi_1,\phi_2),$$

where d_E is the Euclidean distance. If we have Z samples $(\boldsymbol{\phi}_i)_{i=1}^{Z}$ from the posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x}^{\mathbf{0}})$, the Bayes estimator can be approximated by

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{Z}\sum_{i=1}^{Z}\mathcal{L}(\boldsymbol{\phi}_i,\boldsymbol{\phi}), \tag{7}$$

which is also the estimated posterior mean $\hat{\boldsymbol{\phi}} = \frac{1}{Z}\sum_{i=1}^{Z}\boldsymbol{\phi}_i$.

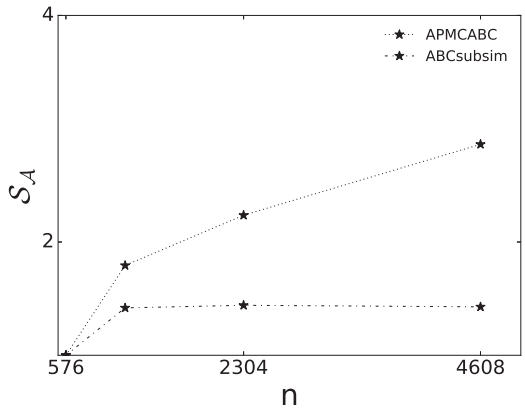


FIG. 4. **Performance comparison**: Computational speedup $\mathcal{S}_{\mathcal{A}}(n)$ of the APMCABC versus ABCsubsim algorithm for the calibration of the Lennard-Jones force-field of helium using the dynamic-MPI backend with different numbers of cores, n.

IV. RESULT

We now illustrate how the Bayesian inference scheme introduced in Sec. III can be used to infer the posterior distribution and Bayes estimates of parameter $\boldsymbol{\phi}$ given an observed dataset $\boldsymbol{x}^{\mathbf{0}}$ in both simulated and experimental settings.

A. LJ potential of helium

1. Simulated data

The simulated data were generated from $\mathcal{M}_{LJ}$ using the software LAMMPS for a system consisting of 1000 atoms, with t_{end} = 20 ns and a time step of 2 fs, under NPT conditions. To generate the simulated data that, in our proof of concept, will play the role of observed data, the parameter values were fixed at $\boldsymbol{\phi}^0 \equiv (\sigma^0_{LJ},\epsilon^0_{LJ}) = (0.2556(\text{nm}), 0.141(\text{zJ}))$ as in Ref. 21. We note that the value of the t_{end} simulation model is 5 ns compared to t_{end} = 20 ns for the simulated dataset. As explained in Ref. 21, the simulation model cannot replicate exactly the target data due to the smaller sampling time and this causes uncertainty due to modeling error, in addition to computational and parametric uncertainty. For inference, we use the APMCABC algorithm, with the tuning parameters fixed at the default values recommended in the ABCpy package with the exception of the N_{sample}, N_{step}, and the acceptance rate cutoff, which are set to 5000, 6, and 0.03, respectively. In Fig. 5, we illustrate the inferred posterior distribution and the Bayes estimate ($\widehat{\boldsymbol{\phi}}$) which is in close agreement with the true parameter values ($\boldsymbol{\phi}^0$) used to simulate the dataset.

The ABCsubsim algorithm was used to infer the parameters of the above problem using the exact same setup (e.g., simulated dataset, discrepancy measure, prior distribution, etc.) as in Ref. 21. In Table I, we compare the performance of APMCABC and ABCsubsim, by comparing the finally achieved threshold value (δ_{final}) and the Euclidean distance of the Bayes estimate from the true value after running APMCABC for N_{step} = 6, with the ones reported in Ref. 21 for ABCsubsim. We note that APMCABC achieves a much

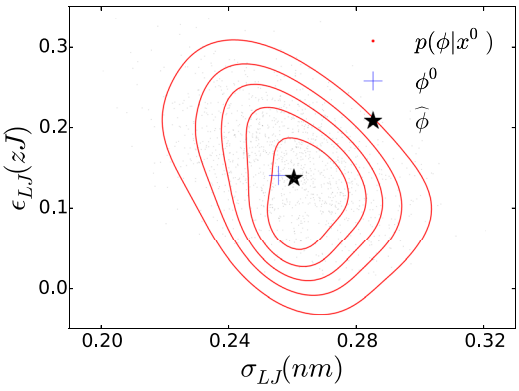


FIG. 5. **Inference on simulated data (helium)**: Contour plot of posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x}^{\mathbf{0}})$ (red), Bayes estimate $\widehat{\boldsymbol{\phi}} \equiv (\widehat{\sigma}_{LJ},\widehat{\epsilon}_{LJ}) = (0.26(\text{nm}), 0.137(\text{zJ}))$ (black star), and true value $\boldsymbol{\phi}^0 \equiv (\sigma^0_{LJ},\epsilon^0_{LJ}) = (0.255(\text{nm}), 0.141(\text{zJ}))$ (blue cross) used to generate the simulated dataset. The posterior correlation between the parameters is −0.38. The posterior distribution is obtained using a Gaussian kernel density estimator with bandwidth 0.7.

TABLE I. **Comparison of inference (helium)**: Euclidean distance of the Bayes estimate from the true parameter value used to simulate the dataset $d_E(\widehat{\boldsymbol{\phi}}, \boldsymbol{\phi}^0)$ and the final threshold value (δ_{final}) achieved by APMCABC and ABCsubsim algorithms for the calibration of the Lennard-Jones force-field of helium, after $N_{\text{step}} = 6$.

Algorithm	$d_E(\widehat{\boldsymbol{\phi}}, \boldsymbol{\phi}^0)$	N_{step}	δ_{final}
APMCABC	0.007 44	6	0.0138
ABCsubsim	0.033 65	6	0.67

smaller δ_{final} value of 0.0138 compared to the value 0.67, achieved by ABCsubsim after $N_{\text{step}} = 6$. Furthermore, the Bayes estimate learned using the inferred posterior samples of APMCABC is in better agreement with the true parameter values used to simulate the dataset, as illustrated by a smaller value of $d_E(\widehat{\boldsymbol{\phi}}, \boldsymbol{\phi}^0)$, the Euclidean distance of the Bayes estimate from the true parameter value used to simulate the dataset. This shows a faster convergence and superior inferential performance of the APMCABC algorithm compared to ABCsubsim, in addition to the better speedups achieved in Fig. 4.

B. TIP4P force-field of water

1. Simulated data

We first consider a simulated setting where the observed data have been generated, using GROMACS, from the TIP4P model $\mathcal{M}_{\text{TP}}$ of water reported in Eq. (4). To generate the simulated observed data, we fix the non-bonded force-field parameter values at $\boldsymbol{\phi}^0 \equiv (\sigma^0_{TP}, \epsilon^0_{TP}) = (0.315(\text{nm}), 0.648(\text{kJ mol}^{-1}))$, which has been found to accurately reproduce the heat of vaporization, critical density, temperature, and liquid density at 298 K.[4] As the true parameter value $\boldsymbol{\phi}^0$ is known, we can assess the performance of the posterior distribution and the Bayes estimate, correspondingly, by their concentration and closeness to $\boldsymbol{\phi}^0$.

For inference, all the tuning parameters of the APMCABC algorithm are fixed at the default values in the ABCpy package with the exception of N_{sample}, N_{step}, and the acceptance rate cutoff, which are set to 100, 10, and 0.03, correspondingly. In Fig. 6, the inferred posterior distribution, Bayes estimate ($\widehat{\boldsymbol{\phi}}$), and the true parameter value ($\boldsymbol{\phi}^0$) clearly show a very good performance of our inference scheme in estimating the underlying parametrization and quantifying the uncertainty in the inference.

Here we would like to note that the summary statistics (S_1, S_2, …, S_9) may have not converged for each of the parameter values at $t_{end} = 100$ ps, contributing to the modeling uncertainty, as $t_{end} = 100$ ps is part of our model specification. The concentrated nature of the posterior distribution in Fig. 6 shows, however, a negligible presence of modeling uncertainty and we can implicitly conclude that the values of these statistics did converge for most of the considered parameter values in the range of prior distribution. Additionally, to point the strength of having a posterior distribution for the parameters, we compute the posterior correlation between σ_{TP} and ϵ_{TP}, highlighting a strong negative correlation of −0.994.

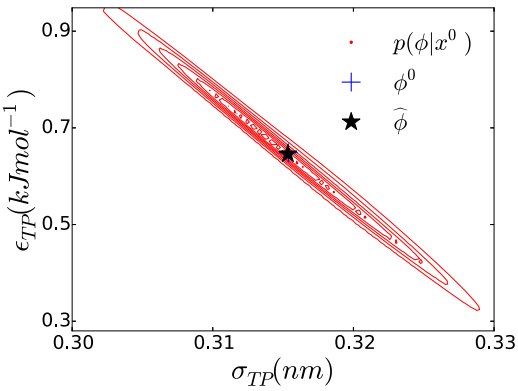


FIG. 6. **Inference on simulated data (water)**: Contour plot of posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x}^{\boldsymbol{0}})$ (red), Bayes estimate $\widehat{\boldsymbol{\phi}} \equiv (\hat{\sigma}_{TP}, \hat{\epsilon}_{TP}) = (0.3153(\text{nm}), 0.646(\text{kJ mol}^{-1}))$ (black star), and true value $\boldsymbol{\phi}^0 \equiv (\sigma^0_{TP}, \epsilon^0_{TP}) = (0.315(\text{nm}), 0.648(\text{kJ mol}^{-1}))$ (blue cross) used to generate the simulated dataset. A strong negative posterior correlation of −0.994 between the parameters is present. The posterior distribution is obtained using a Gaussian kernel density estimator with bandwidth 0.7.

2. Experimental data

We now illustrate the performance of the inference scheme for the experimental dataset of water molecules under ambient conditions (temperature and pressure being fixed at 298 K and 1 atm), assuming TIP4P force-field formalism of water. Though the exact coordinates over time of water molecules in an enclosed space cannot be observed, the radial distribution functions of different molecular bonds and self-diffusion coefficient can be learned by experimental studies.[30,31] As our inference scheme only depends on the summary statistics extracted from the radial distribution functions of O–O, O–H, and self-diffusion coefficient, provided we have access to the experimentally obtained radial distribution functions and self-diffusion coefficient, we can compute the summary statistics and thus infer the non-bonded force-field parameters.

a. Neutron diffraction dataset. First we consider the neutron diffraction derived radial distribution function of water from Ref. 30. The experimentally obtained radial distribution function of O–O and O–H is shown in Figs. 7(a) and 7(b), for further details we point readers to Ref. 30. Additionally, we use the value of the self-diffusion coefficient, $1.3\, e^{-5}$ cm^2 s^{-1}, reported in Ref. 34. The experimental values of the other summary statistics calculated are $S^e_1 = 0.9628$, $S^e_2 = 0.2432$, $S^e_3 = 0.059$, $S^e_4 = 0.8634$, $S^e_5 = 0.33$, $S^e_6 = 0.130$, $S^e_7 = 2.69$, and $S^e_8 = 0.275$(nm).

In Fig. 7(c), we plot the posterior distribution of $\boldsymbol{\phi}$ and the Bayes estimates $\widehat{\boldsymbol{\phi}} \equiv (\hat{\sigma}_{TP}, \hat{\epsilon}_{TP}) = (0.319(\text{nm}), 0.614(\text{kJ mol}^{-1}))$ obtained using the proposed inference scheme. The posterior correlation between σ_{TP} and ϵ_{TP} has a negative correlation of −0.95 as in the case of the simulated dataset.

b. X-ray diffraction dataset. Next we consider an experimental dataset where only the radial distribution function of O–O of water using the X-ray diffraction is obtained.[31] The experimentally obtained radial distribution function of O–O is shown in Fig. 8(a), for further details we point readers to Ref. 31. Additionally, we use the value of the self-diffusion

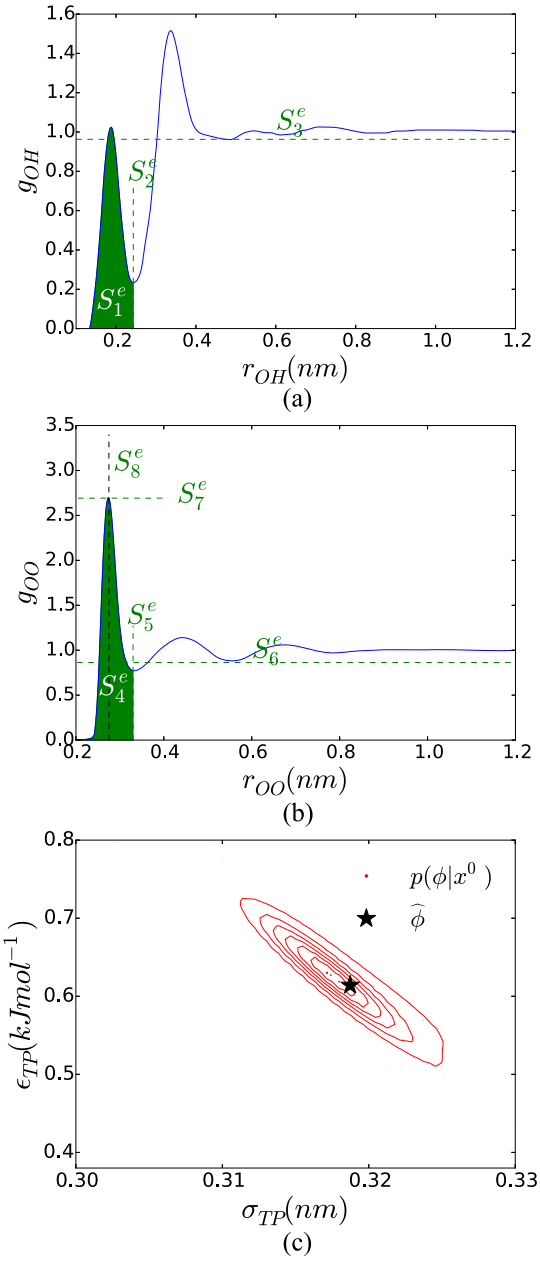


FIG. 7. **Neutron diffraction dataset (water)**: (a) and (b) The radial distribution functions of O–H and O–O of water obtained by neutron diffraction and reported in Ref. 30. (c) Contour plot of posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x^0})$ (red) and Bayes estimate $\widehat{\boldsymbol{\phi}} \equiv (\hat{\sigma}_{TP}, \hat{\epsilon}_{TP}) = (0.319(\text{nm}), 0.614(\text{kJ mol}^{-1}))$ (black star). A strong negative posterior correlation of −0.95 between the parameters is present. The posterior distribution is obtained using a Gaussian kernel density estimator with bandwidth 0.7.

coefficient, 1.3 e^{-5} cm^2 s^{-1}, reported in Ref. 34. The experimental values of the other summary statistics calculated are $S_4^e = 0.98$, $S_5^e = 0.347$, $S_6^e = 0.141$, $S_7^e = 2.57$, and $S_8^e = 0.281(\text{nm})$. In the absence of radial distribution function of O–H, we only consider these 5 summary statistics and the following distance function between them as the discrepancy measure:

$$\tilde{d}_{TP}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := \tilde{d}_{TP}\left(\mathcal{F}_{TP}(\boldsymbol{x}^{(1)}), \mathcal{F}_{TP}(\boldsymbol{x}^{(2)})\right) = \frac{1}{6}\sum_{i=4}^{9} |S_i^{(1)} - S_i^{(2)}|. \quad (8)$$

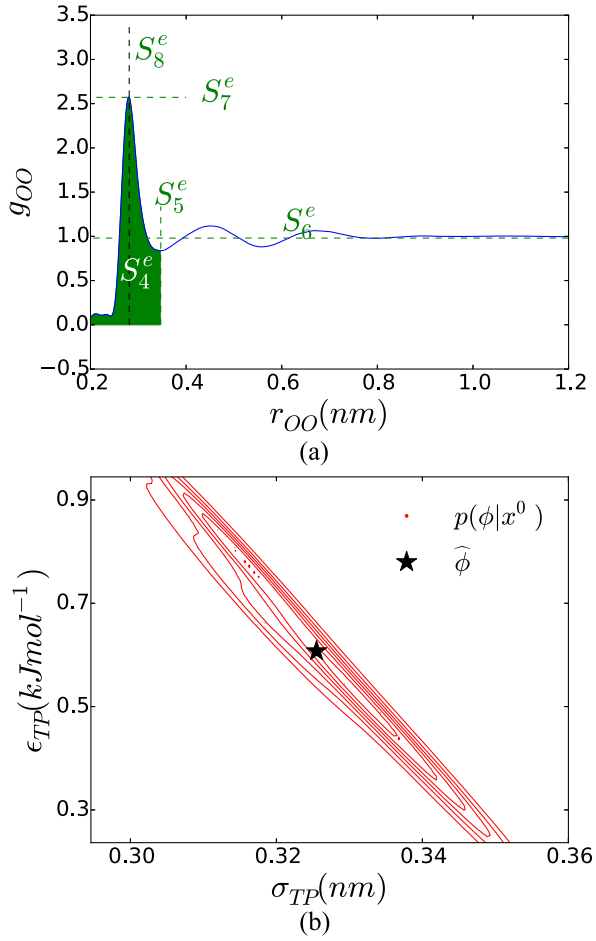


FIG. 8. **X-ray diffraction dataset (water)**: (a) The radial distribution function of O–O of water obtained by X-ray diffraction and reported in Ref. 31. (b) Contour plot of posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x^0})$ (red) and Bayes estimate $\widehat{\boldsymbol{\phi}} \equiv (\hat{\sigma}_{TP}, \hat{\epsilon}_{TP}) = (0.326(\text{nm}), 0.607(\text{kJ mol}^{-1}))$ (black star). A strong negative posterior correlation of −0.99 between the parameters is present. The posterior distribution is obtained using a Gaussian kernel density estimator with bandwidth 0.7.

In Fig. 8(b), we plot the posterior distribution of $\boldsymbol{\phi}$ and the Bayes estimates $\widehat{\boldsymbol{\phi}} \equiv (\hat{\sigma}_{TP}, \hat{\epsilon}_{TP}) = (0.326(\text{nm}), 0.607(\text{kJ mol}^{-1}))$ obtained using the proposed Bayesian inference scheme. Also in this case, we see a strong negative posterior correlation of −0.99 between σ_{TP} and ϵ_{TP}.

V. MODEL PREDICTION AND VALIDATION

A. Model prediction

As mentioned before, we can quantify the model prediction uncertainty, by simulating the dataset from the MD simulation model using values of the parameters randomly drawn from the inferred posterior distribution. Using the posterior distribution inferred and illustrated in Figs. 7(c) and 8(b), we provide posterior prediction for the experimentally obtained radial distribution function of O–O in Figs. 9 and 10. We also illustrate the area between the minimum and maximum of the predicted datasets and the $\frac{1}{4}$th and the $\frac{3}{4}$th quantile of the predicted datasets with the light and dark gray color, correspondingly, to assess the quality of the fit and the prediction uncertainty.

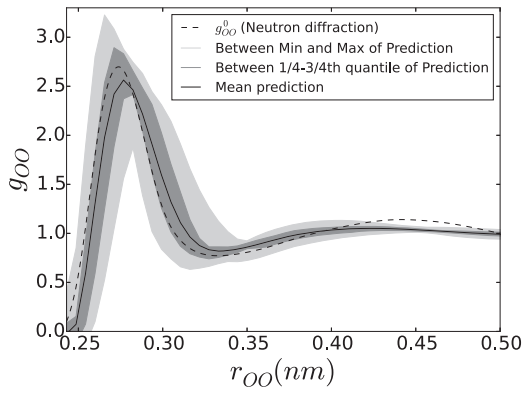


FIG. 9. **Prediction of the radial distribution function (neutron diffraction)**: Radial distribution function of O–O obtained using neutron diffraction (dashed), mean posterior prediction (solid), and the area between the minimum and maximum and the $\frac{1}{4}$th and the $\frac{3}{4}$th quantile of the posterior prediction using posterior samples reported in Fig. 7(c). We notice a close similarity between experimentally obtained and mean predicted data with reference to the properties of the first hydration shell.

Here we note that the experimental dataset is mostly within the prediction band (indeed, a large portion of values are inside the $\frac{1}{4}$th and the $\frac{3}{4}$th quantile of the predicted datasets), indicating a good predictive performance, except near the second maximum of g_{OO}. The mismatch near g_{OO} can be explained by the absence of the location and height of the second maximum of g_{OO} among the chosen summary statistics as reported in Sec. III C. This mismatch can be resolved by including those summary statistics and modifying the discrepancy measure accordingly, which illustrates the flexibility of ABC algorithms.

B. Validation

For further validation of our inferred TIP4P force-fields from the experimental dataset (X-ray and neutron diffraction), in Table II we compare values of a set of properties, which have not been used for parameterization. The properties considered here for comparison are the heat capacity (Cp cal mol^{-1} K^{-1}) and the density (ρ g cm^{-3}) of liquid water at 298 K and of ice at

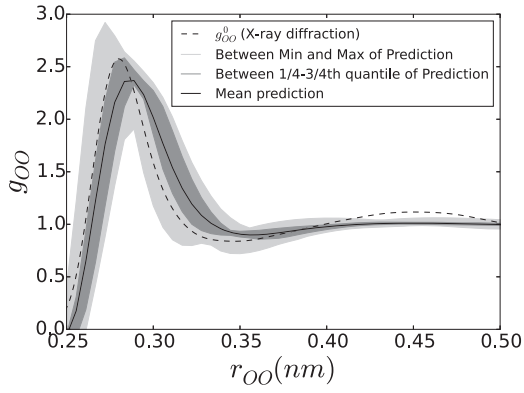


FIG. 10. **Prediction of the radial distribution function (X-ray diffraction)**: Radial distribution function of O–O obtained using X-ray diffraction (dashed), mean posterior prediction (solid), and the area between the minimum and maximum and the $\frac{1}{4}$th and the $\frac{3}{4}$th quantile of the posterior prediction using posterior samples reported in Fig. 8(b). We notice a close similarity between experimentally obtained and mean predicted data with reference to the properties of the first hydration shell.

TABLE II. **Comparison of properties (TIP4P water)**: Values of heat capacity (Cp), density (ρ), isothermal compressibility (κ_T), and the dielectric constant (ξ), experimentally obtained (Expt.) under atmospheric pressure and simulated using estimated $\widehat{\boldsymbol{\phi}} \equiv (\widehat{\sigma}_{TP}, \widehat{\epsilon}_{TP})$ from the two datasets (neutron diffraction and X-ray diffraction) and the original TIP4P parameterization $\boldsymbol{\phi}^0 \equiv (\sigma^0_{TP}, \epsilon^0_{TP})$. These properties were not used in the ABC inference procedure to calibrate/estimate the parameters.

	Properties	Expt.	TIP4P	Neutron diffraction	X-ray diffraction
Ice (250 K)	Cp	8.3	14.7	12.47	20.02
	ρ	0.92	0.937	0.913	1
Water (298 K)	Cp	18	20	20.1	18.3
	ρ	0.997	0.988	0.958	0.854
	κ_T	45.3	59	57.5	79.1
	ξ	78.5	50	47	43

250 K, as well as the isothermal compressibility (κ_T 10^{-6}/bar) and the dielectric constant (ξ) of water at 298 K. The values of these properties were simulated (by performing 5 ns NPT simulations at 298 K and 250 K using the forward simulation models described in Sec. II) using estimated $\widehat{\boldsymbol{\phi}} \equiv (\widehat{\sigma}_{TP}, \widehat{\epsilon}_{TP})$ from both the dataset and the original TIP4P parameterization $\boldsymbol{\phi}^0 \equiv (\sigma^0_{TP}, \epsilon^0_{TP})$ and then compared with their experimental values obtained under atmospheric pressure.

The simulated values of the properties obtained using the estimated parametrization from the neutron diffraction dataset are closer to the experimental values than the ones simulated using the estimated parametrization from the X-ray diffraction dataset, in all cases but for the heat capacity of the liquid water. We notice that the X-ray diffraction dataset did not have the radial distribution function of O–H and the predictive performance of the inference scheme suffers because of the absence of this summary statistics. Remarkably, the parametrization based on the neutron diffraction dataset is able to better predict the values of heat capacity (Cp) and density (ρ) of ice and the isothermal compressibility (κ_T) of water than the normal TIP4P parametrization. However, it predicts worse the dielectric constant (ξ) and the density (ρ) of water at 298 K.

VI. DISCUSSION

We propose a Bayesian inference framework to calibrate the force-field parameters of molecular systems, without a Gaussian model for the system uncertainty. This is achieved using approximate Bayesian computation, specifically with the help of the adaptive population Monte Carlo ABC algorithm and High Performance Computing. The imbalance in the simulation time of MD simulations for different parameter values was resolved using a dynamic-allocation scheme for MPI. The methodology is illustrated by learning the force-field parameters of Lennard-Jones potential of helium and the TIP4P system of water, for both simulated and experimental datasets. The proposed methodology has faster convergence and better scale-up performance compared to the ABCsubsim algorithm previously used to calibrate force-fields, thus contributing to a more efficient uncertainty quantification and parameter estimation of the force-fields.

Bayesian inference provides posterior distribution of the parameters; hence we are able to compute their posterior correlation in a data-driven manner, which turns out to be strongly negative. Furthermore, the negative correlation structure between non-bonded force-fields has been observed across two different experimental datasets, obtained using neutron and X-ray diffraction. This points toward a strong underlying mechanism. Though further investigation is necessary to understand this mechanism, we hypothesize that at smaller values of σ, the repulsion is present at shorter distances and oxygen atoms can come closer. For oxygen atoms to come closer, while maintaining the same summary statistics (i.e., the same number of first neighbors), without creating vacuum vacancies, there should be more attraction between each other. This scenario is equivalent to a bigger ϵ. The opposite holds for higher values of σ. Further using the posterior distribution, we were also able to quantify the model prediction uncertainty, which gives us an assessment on how good the model fit and prediction uncertainty are.

We foresee that, for the calibration of challenging force-fields, more sophisticated ABC algorithms might be needed, possibly using surrogate and less computationally intensive models.[50] Additionally, ABC model selection, without the need of Gaussian modeling of the uncertainty, can be used to learn the most suitable force-field formalism for an experimental dataset. This will be a line of future investigation. Finally, in more complex systems and/or in a more complicated parameter space, reaching ergodicity is often challenging due to high free energy barriers, hampering a fast exploration of the free energy landscape, hence biasing the summary statistics. In the future, we foresee combining our APMCABC framework to state of the art enhanced sampling methods such as Metadynamics, Umbrella Sampling, and Transition Path Sampling.[51–53]

ACKNOWLEDGMENTS

We thank Alessandro Laio and Michele Parrinello for suggesting the specific force-field of TIP4P, for insightful comments on the choice of the summary statistics, and for discussing earlier drafts of this manuscript. This research was supported by the Swiss National Science Foundation (Grant No. 105218-163196) (Statistical Inference on Large-Scale Mechanistic Network Models). We also thank Marcel Schoengens, CSCS, ETH Zurich for help regarding HPC and the Swiss National Super Computing Center for providing computing resources.

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, New York, NY, USA, 1989).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, Inc., Orlando, FL, USA, 2001).
[3] M. Karplus and R. Lavery, "Significance of molecular dynamics simulations for life sciences," Isr. J. Chem. **54**(8-9), 1042–1051 (2014).
[4] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**(2), 926–935 (1983).
[5] J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: Tip4p/2005," J. Chem. Phys. **123**(23), 234505 (2005).
[6] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," J. Chem. Phys. **112**(20), 8910–8922 (2000).
[7] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate Bayesian computational methods," Stat. Comput. **22**(6), 1167–1180 (2012).
[8] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework," J. Chem. Phys. **137**(14), 144103 (2012).
[9] F. Cailliez and P. Pernot, "Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation," J. Chem. Phys. **134**(5), 054124 (2011).
[10] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, "Uncertainty quantification in MD simulations. Part I: Forward propagation," Multiscale Model. Simul. **10**(4), 1428–1459 (2012).
[11] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, "Data driven, predictive molecular dynamics for nanoscale flow simulations under uncertainty," J. Phys. Chem. B **117**(47), 14808–14816 (2013).
[12] A. Chernatynskiy, S. R. Phillpot, and R. LeSar, "Uncertainty quantification in multiscale simulation of materials: A prospective," Annu. Rev. Mater. Res. **43**, 157–182 (2013).
[13] F. Cailliez, B. Arnaud, and P. Pascal, "Calibration of forcefields for molecular simulation: Sequential design of computer experiments for building cost-efficient kriging metamodels," J. Comput. Chem. **35**(2), 130–149 (2014).
[14] K. Farrell, J. T. Oden, and D. Faghihi, "A bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems," J. Comput. Phys. **295**, 189–208 (2015).
[15] K. Sargsyan, H. N. Najm, and R. Ghanem, "On the statistical calibration of physical models," Int. J. Chem. Kinet. **47**(4), 246–276 (2015).
[16] K. Farrell-Maupin and J. T. Oden, "Adaptive selection and validation of models of complex systems in the presence of uncertainty," Res. Math. Sci. **4**(1), 14 (2017).
[17] R. A. Messerly, T. A. Knotts IV, and W. V. Wilding, "Uncertainty quantification and propagation of errors of the Lennard-Jones 12-6 parameters for n-alkanes," J. Chem. Phys. **146**(19), 194110 (2017).
[18] P. Pernot and F. Cailliez, "A critical review of statistical calibration/prediction models handling data inconsistency and model inadequacy," AIChE J. **63**(10), 4642–4665 (2017).
[19] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel *et al.*, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics **29**(7), 845–854 (2013).
[20] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comput. Phys. **117**(1), 1–19 (1995).
[21] L. Kulakova, P. Angelikopoulos, P. E. Hadjidoukas, C. Papadimitriou, and P. Koumoutsakos, "Approximate Bayesian computation for granular and molecular dynamics simulations," in *Proceedings of the Platform for Advanced Scientific Computing Conference* (ACM, 2016), p. 4.
[22] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Fundamentals and recent developments in approximate Bayesian computation," Syst. Biol. **66**(1), e66–e82 (2017).
[23] M. Chiachio, J. L. Beck, J. Chiachio, and G. Rus, "Approximate Bayesian computation by subset simulation," SIAM J. Sci. Comput. **36**(3), A1339–A1358 (2014).
[24] R. Dutta, M. Schoengens, J.-P. Onnela, and A. Mira, "ABCpy: A user-friendly, extensible, and parallel library for approximate Bayesian computation," in *Proceedings of the Platform for Advanced Scientific Computing Conference* (ACM, 2017), p. 8.
[25] R. Dutta, M. Schoengens, A. Ummadisingu, N. Widmer, J.-P. Onnela, and A. Mira, "ABCpy: A high-performance computing perspective to approximate Bayesian computation," preprint arXiv:1711.04694 (2017).
[26] Z. F. Brotzakis, C. C. M. Groot, W. H. Brandeburgo, H. J. Bakker, and P. G. Bolhuis, "Dynamics of hydration water around native and misfolded α-lactalbumin," J. Phys. Chem. B **120**(21), 4756–4766 (2016).
[27] H. Ohtaki and T. Radnai, "Structure and dynamics of hydrated ions," Chem. Rev. **93**(3), 1157–1204 (1993).
[28] R. Abel, T. Young, R. Farid, B. J. Berne, and R. A. Friesner, "Role of the active-site solvent in the thermodynamics of factor xa ligand binding," J. Am. Chem. Soc. **130**(9), 2817–2831 (2008).
[29] Z. F. Brotzakis, I. K. Voets, H. J. Bakker, and P. G. Bolhuis, "Water structure and dynamics in the hydration layer of a type iii anti-freeze protein," Phys. Chem. Chem. Phys. **20**(10), 6996–7006 (2018).
[30] A. K. Soper, "The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa," Chem. Phys. **258**(2-3), 121–137 (2000).

[31]L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore, "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide q-range," J. Chem. Phys. **138**(7), 074506 (2013).

[32]Note that in this paper, for the helium system simulated with LAMMPS we use zJ and nm units for the σ_{LJ} and ϵ_{LJ}, whereas for the TIP4P water simulated with GROMACS, we used the GROMACS conventional units of kJ mol^{-1} and nm for the σ_{TP} and ϵ_{TP}, respectively. We select these units to be consistent with the papers we have chosen as our benchmark.

[33]W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," Phys. Rev. B **69**(13), 134103 (2004).

[34]C. Vega and J. L. F. Abascal, "Simulating water with rigid non-polarizable models: A general perspective," Phys. Chem. Chem. Phys. **13**(44), 19663–19688 (2011).

[35]B. J. Jaidhan, P. S. Rao, and A. Apparao, "Energy minimization and conformation analysis of molecules using steepest descent method," Int. J. Comput. Sci. Inf. Technol. **5**(3), 3525–3528 (2014), ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503193.pdf.

[36]B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije *et al.*, "LINCS: A linear constraint solver for molecular simulations," J. Comput. Chem. **18**(12), 1463–1472 (1997).

[37]T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems," J. Chem. Phys. **98**(12), 10089–10092 (1993).

[38]G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, "Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics," J. Am. Chem. Soc. **128**(41), 13435–13441 (2006).

[39]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**(12), 7182–7190 (1981).

[40]C. P. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer-Verlag, New York, 2005).

[41]E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, "Real-time dynamics of lattice gauge theories with a few-qubit quantum computer," Nature **534**(7608), 516–519 (2016).

[42]P. Turchin, T. E. Currie, E. A. L. Turner, and S. Gavrilets, "War, space, and the evolution of old world complex societies," Proc. Natl. Acad. Sci. U. S. A. **110**(41), 16384–16389 (2013).

[43]J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford, "The EAGLE project: Simulating the evolution and assembly of galaxies and their environments," Mon. Not. R. Astron. Soc. **446**(1), 521–554 (2015).

[44]M. Lenormand, F. Jabot, and G. Deffuant, "Adaptive approximate Bayesian computation for complex models," Comput. Stat. **28**(6), 2777–2796 (2013).

[45]P. Fearnhead and D. Prangle, "Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation," J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **74**(3), 419–474 (2012).

[46]P. Pudlo, J.-M. Marin, A. Estoup, J.-M. Cornuet, M. Gautier, and C. P. Robert, "Reliable ABC model choice via random forests," Bioinformatics **32**(6), 859–866 (2015).

[47]B. Jiang, T.-yu Wu, C. Zheng, and W. H. Wong, "Learning summary statistic for approximate Bayesian computation via deep neural network," Stat. Sin. **27**, 1595–1618 (2017); preprint arXiv:1510.02175 (2015).

[48]M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Likelihood-free inference via classification," Stat. Comput. **28**(2), 411–425 (2018).

[49]R. L. Graham, "Bounds for certain multiprocessing anomalies," Bell Labs Tech. J. **45**(9), 1563–1581 (1966).

[50]E. Meeds and M. Welling, "GPS-ABC: Gaussian process surrogate approximate Bayesian computation," in *UAI'14 Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, Quebec City, Quebec, Canada, 23-27 July 2014* (AUAI Press, Arlington, VA, 2014), pp. 593–602; preprint arXiv:1401.2838 (2014).

[51]A. Laio and M. Parrinello, "Escaping free-energy minima," Proc. Natl. Acad. Sci. U. S. A. **99**(20), 12562–12566 (2002).

[52]G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**, 187–199 (1977).

[53]Z. F. Brotzakis and P. G. Bolhuis "A one-way shooting algorithm for transition path sampling of asymmetric barriers," J. Chem. Phys. **145**(16), 164112 (2016).