Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework

Panagiotis Angelikopoulos,1,a) Costas Papadimitriou2,b) and Petros Koumoutsakos1,c)
1 Computational Science and Engineering Laboratory, ETH Zürich, CH-8092 Zurich, Switzerland
2 Department of Mechanical Engineering, University of Thessaly, Pedion Areos, GR-38334 Volos, Greece
(Received 5 June 2012; accepted 20 September 2012; published online 10 October 2012)

We present a Bayesian probabilistic framework for quantifying and propagating the uncertainties in the parameters of force fields employed in molecular dynamics (MD) simulations. We propose a highly parallel implementation of the transitional Markov chain Monte Carlo for populating the posterior probability distribution of the MD force-field parameters. Efficient scheduling algorithms are proposed to handle the MD model runs and to distribute the computations in clusters with heterogeneous architectures. Furthermore, adaptive surrogate models are proposed in order to reduce the computational cost associated with the large number of MD model runs. The effectiveness and computational efficiency of the proposed Bayesian framework is demonstrated in MD simulations of liquid and gaseous argon. © 2012 American Institute of Physics. [http://dx.doi.org/10.1063/1.4757266]

I. INTRODUCTION

Molecular dynamics (MD) is a powerful computational tool for investigating atomistic phenomena in areas ranging from biology to materials science.1–3 In recent years, several algorithmic developments and the availability of massively parallel computer architectures has extended the applicability and time scales that can be addressed by MD simulations.4 Albeit these extensions, a critical aspect of MD simulations is that the physics of the system being simulated hinges on the selected interaction potentials.

The parameters of MD simulation potentials are often obtained through careful validations with experimental works.5 However, when these potentials are implemented for simulating systems beyond their calibration envelope, the interpretation and validity of the simulated predictions are questionable. Examples include the water-carbon interaction potentials, as quantified by Lennard-Jones (LJ) potentials between carbon and oxygen atoms. A large number of simulations have employed different parameters for these interaction potentials, reporting from strongly hydrophobic to strongly hydrophilic behavior. The availability of experimental results can help specify the suitable binding energy,6 but its value can be sensitive to experimental variations. Furthermore, the enhancement of carbon-oxygen interaction models with polarization effects, leads to other differences such as the filling of carbon nanotubes by water molecules.7,8 In another example, effective pairwise potentials are often incorporated in the MD simulations of macromolecules via some mixing or phenomenological rules.9 In coarse grained MD simulations, as for example in the well established MARTINI force field,10 the choice of cutoff length has been shown to significantly affect the permeation free energy of gold nanoparticles interacting with a lipid bilayer.11 A number of studies evaluating the predictive capabilities of popular force fields describe the “forward” problem. That is, given an already calibrated force field against some quantity, measure the predictive capabilities of this force field to another quantity. In a notable example,5 the authors estimate the goodness of fit for biomolecular force fields based on neutron magnetic resonance data.

It is important to note that albeit the widespread use of MD simulations, there have been limited efforts to implement a systematic Uncertainty Quantification and Propagation (UQ+P) framework to assess the functional form of potentials, their parameters as well as the effects of computational parameters in the results of these simulations. The influence of parametric uncertainty in predictive simulations is gaining recognition across scientific disciplines.12–15 In the context of MD simulations Bayesian statistics were first employed to evaluate interatomic potentials and to assess related error bars in predictions of atomistic simulations of molybdenum.16 Cooke and Schmidler17 used full replica exchange MD simulations to perform Bayesian calibration of the internal dielectric value, against experimentally measured helicity of peptide molecules. A recent work18 has developed a UQ+P Bayesian framework focused on simplified models of atomistic systems for which analytic expressions exist that relate thermodynamic properties with force-field parameters. The use of these analytic expressions bypasses the expensive runs of the MD simulations, required in Bayesian frameworks, to compute the uncertainty in the force-field model parameters. However, in most complex molecular systems, such analytic expressions are not available and a large number of computationally demanding simulation runs have to be performed. In Ref. 19, uncertainty quantification has been implemented in a multiscale model of single crystal plasticity with the uncertainty in the initial conditions and parameters of a MD model propagated to a coarse grain phase field model.

Bayesian inference approaches involve Monte Carlo simulations, a versatile approach for stochastic analysis, as they do not postulate any assumption on the system response function. A drawback of these approaches is the high
computational demand required for the repeated system runs sampling the parameter space. Alternatively, deterministic techniques with reduced computational cost, such as perturbation and spectral methods, and model reduction schemes, notably the polynomial chaos projection techniques, can be applied for assessing the variability of the response of interest. However, the fluctuations inherent to quantities measured in MD simulations is not compatible with the assumptions of the functional approximations of these approaches.

In this work, we develop a high performance computing (HPC) implementation to reduce time to solution and overall computational cost of applying the Bayesian framework for UQ+P of MD simulations. The present work is based on the Bayesian probabilistic framework established in structural dynamics for quantifying and calibrating uncertainty models based on measured data, as well as propagating these modeling uncertainties in simulations to achieve updated or posterior robust predictions of system performance, reliability, and safety. The proposed Bayesian framework exploits the available thermodynamic and kinetic experimental data for a system described by MD simulations, to perform the following tasks:

- Identify the uncertainties in the parameters of force-field models and suitable thermodynamic ensembles representing the behavior of the components of the system.
- Identify probabilistic models that best account in predictions for the mismatch between MD based predictions and measurements.
- Propagate uncertainties in MD simulations for robust predictions that take into account the validated models and the calibrated uncertainties.

We employ a variant of the transitional Markov chain Monte Carlo (TMCMC) algorithm with highly parallel performance along with efficient surrogate models to reduce the number of full MD simulation runs. Theoretical, algorithmic, and computational developments are demonstrated, using available analytical expressions as well as full molecular simulations for the prediction of thermodynamic and other quantities of interest (QoI) for noble gases. To the best of our knowledge, no such systematic Bayesian inference framework has been applied to MD simulations.

The paper is structured as follows: In Sec. II, we outline sources of uncertainty in MD simulations, followed by the background of the Bayesian framework in Sec. III. We describe Markov chain Monte Carlo (MCMC) and TMCMC algorithms in Sec. IV, surrogate models in Sec. V, TMCMC parallel considerations in Sec. VI, and conclude in Sec. VII with illustrative examples of MD simulations combined with UQ+P algorithms.

II. UNCERTAINTIES IN MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics simulations integrate Newton’s equation of motion for interacting atoms:

\[ m_i \ddot{q}_i = F(q), \quad q(0) = q_0, \quad \dot{q}(0) = p_0. \]  

where \( m_i, q_i(t) \) denote the mass and locations of \( i = 1 \ldots N \) atoms, and \( F(q) \) and \( F(q) = -\nabla_q \phi(q) \) the a priori specified potential and corresponding force field.

Statistical estimates of various thermodynamic and kinetic QoI of the system, such as the temperature \( T(\dot{q}^2) \) or the self-diffusion coefficient \( D_{ij} = \langle \dot{q}_i^2(t) \rangle / 6t \) can be obtained by processing the particle trajectories. This ensemble averaging, denoted by \( \langle \cdot \rangle \), is prone to statistical error due to the finite integration time and number of particles, as well as due to the numerical quadrature.

At the core of any MD simulation lies the functional form and parameters of the chosen interaction potentials. These potentials encompass a priori knowledge concerning the nature of interactions between atoms, obtained for example by quantum mechanics simulations, analytical observation, and validations with experiments.

The functional form of the intermolecular potential is often based on rather empirical considerations. For example in the commonly used Lennard-Jones potential

\[ \phi_{\text{LJ}}(r_{ij}) = 4\epsilon_{\text{LJ}} \left[ \left( \frac{\sigma_{\text{LJ}}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{\text{LJ}}}{r_{ij}} \right)^{6} \right], \]  

while the term to the power 6 was chosen to reflect the van der Waals interactions of monatomic atoms, the power 12 term is chosen simply as a multiple of the power 6 term for computational efficiency. We note that there is no rigorous justification for the prevalence of the Lennard-Jones (LJ) potential over other suggested pairwise short range interaction potentials (such as the exp-six potential or the Buckingham potentials). In the context of the LJ potential the well depth \( \epsilon_{\text{LJ}} \) and the length parameter \( \sigma_{\text{LJ}} \) of the LJ potential are usually calibrated to re-produce one or more available experimental measurements of system properties (second virial coefficient, viscosity, diffusion coefficient, liquid/vapor equilibrium, enthalpy of vaporization, free energy of mixing, etc.). In most cases, only a single best value of the model parameters is used in the simulations, while the uncertainty in the model parameters is largely ignored.

In this work, we use the force shifted LJ potential and quantify uncertainties in its parameters and estimate the confidence associated with the predictions obtained by using this potential in MD simulations. We employ a LJ potential \( \phi(r_{ij}) \),

\[ \phi(r_{ij}) = \begin{cases} \phi_{\text{LJ}}(r_{ij}) - \phi_{\text{LJ}}(r_{\text{cut}})(r_{ij} - r_{\text{cut}}), & r_{ij} < r_{\text{cut}} \\ 0, & r_{ij} \geq r_{\text{cut}} \end{cases} \]  

where \( r_{\text{cut}} = r_c \) is the cut-off-radius, \( \phi_{\text{LJ}}(r) = d\phi_{\text{LJ}}(r)/dr \), \( r_{ij} = r_j - r_i \), and \( r_i \) and \( r_j \) are the positions of particles \( i \) and \( j \), respectively. All pairwise interactions are truncated to speed up the computations.

The sources of uncertainties in MD simulations can be distinguished as

- **Modeling uncertainties:** pertain to the functional form of the force field models used in MD simulations. Such uncertainties entail also the use of truncation in the calculation of short and long range interactions.
• **Parametric uncertainties**: represent the lack of knowledge for the precise values of the parameters involved in the chosen force-field models. For example in LJ potentials these include $\epsilon_{1J}$ and $\sigma_{1J}$. Other parameters relate to the atomistic simulation protocol, such as the biasing force constants in free energy calculations.

• **Computational uncertainties**: originate from the particular computational setup. Examples include the finite time of sampling, the size of the simulation box, the number of molecules included in the simulation box, and the time integrators. These parameters affect the variability in the predictions of the output QoIs that are obtained through statistical ensembles. Output quantities in atomistic simulations are inherently noisy due to the way of calculating them under assumptions of statistical thermodynamics. However, since thermodynamic equilibrium only occurs in the limit of infinite time, the predicted quantities always suffer from finite sampling and have to be treated probabilistically.

• **Measurement uncertainties**: refer to experimental or observational errors, due to variabilities in the experimental setup, errors in the measuring equipment, and inaccuracies in the data acquisition system.

III. BAYESIAN UNCERTAINTY QUANTIFICATION AND PROPAGATION

A. Formulation

Consider a parameterized class $M_m$ of molecular models that are used to predict various output QoIs of a molecular system. Let $\theta_m \in \mathbb{R}^N_m$ be a set of parameters (e.g., force-field parameters, cut-off radius) in this model class to be estimated using experimental data. Let $q(\theta_m|M_m) \in \mathbb{R}^N_m$ be the output QoI in molecular systems (e.g., virial coefficients, viscosity, self-diffusion, heat capacities) given the model class $M_m$ and the value of the parameter set $\theta_m$. The model parameters $\theta_m$ are considered to be uncertain and a probability distribution quantifies their plausible values. One may assign an *a priori* probability distribution function (PDF) $\pi_m(\theta_m|M_m)$ to the model parameters incorporating prior information that is usually subjective based on previous knowledge, experience, or physical limitations.

In Bayesian inference, the probability distribution of the model parameters $\theta_m$ is updated based on measurements available either at molecular component or system level. For this, let $D \equiv \hat{y} = \{\hat{y}_r, r = 1, \ldots, n_\hat{y}\} \in \mathbb{R}^{n_\hat{y}}$ be a set of observations (data) available from experiments, where $n_\hat{y}$ is the number of observations. These data may correspond to various measured quantities obtained under different states of the system (e.g., different temperature and pressure conditions). We wish to:

1. first, estimate the uncertainty in the model parameters using the available experimental information and
2. subsequently, propagate the calibrated uncertainties through the MD model to quantify the uncertainty of the output QoI.

The UQ+P formulation starts by building a probabilistic model that characterizes the discrepancy between the model predictions $g(\theta_m|M_m) \in \mathbb{R}^N_m$ obtained from a particular value of the model parameters $\theta_m$ and the corresponding data $\hat{y}$, available from experiments. We assume that the observation data and the model predictions satisfy the prediction error equation

$$\hat{y} = g(\theta_m|M_m) + e,$$

where the prediction error specified as

$$e = e^d + e^c + e^m$$

is composed of three parts, accounting for the measurement ($e^d$), computational ($e^c$), and modeling ($e^m$) uncertainties, respectively.

To complete the Bayesian formulation, a probabilistic structure for the prediction error terms is postulated. Let $M_r$ be a family of probability model classes for the prediction error terms, which depend on a set of prediction error parameters $\theta_e$ to be determined using the experimental data. A prior probability distribution $\pi_e(\theta_e|M_m)$ is assigned to quantify the uncertainties of the prediction error parameters $\theta_e$ based on subjective information.

The Bayesian approach to model calibration deals with updating the combined parameter set $\theta = (\theta_m, \theta_e)$. The parameters $\theta_m$ and $\theta_e$ can be considered as independent with prior probability distribution for the combined set given by

$$\pi(\theta|M) = \pi_m(\theta_m|M_m) \cdot \pi_e(\theta_e|M_e),$$

where $M = \{M_m, M_e\}$. The updated distribution $f(\theta|D, M)$ of the model parameters $\theta$, given the data $D$ and the model class $M$, results from the Bayes theorem:

$$f(\theta|D, M) = \frac{f(D|\theta, M) \cdot \pi(\theta|M)}{f(D|M)},$$

where $f(D|\theta, M)$ is the *likelihood* of observing the data $D$ from a model corresponding to a value $\theta$ of the model class $M$ and $f(D|M)$ is the *evidence* of the model class $M$, given by the multi-dimensional integral

$$f(D|M) = \int_{\Theta} f(D|\theta, M) \cdot \pi(\theta|M) d\theta$$

over the space $\Theta$ of the uncertain model parameters. The constant $f(D|M)$ is selected such that the posterior distribution $f(\theta|D, M)$ of the model parameters integrates to one.

B. Prediction error model classes

The posterior probability distribution of the model parameters depends on the selection of the prediction error model classes. A normal distribution, with zero mean, is a reasonable choice for the prediction error terms involved in (5), consistent with the maximum entropy principle, \(^{35}\)

$$e^d \sim N(0, \Sigma^d), e^c \sim N(0, \Sigma^c), \text{ and } e^m \sim N(0, \Sigma^m).$$

For the measurement error, we assume a diagonal covariance matrix $\Sigma^d = \text{diag}(\nu_1^2, \ldots, \nu_n^2)$, where $\nu_i^2$ is the normalized variance of the $i$th observation. Similarly, for the computational error, the covariance matrix can be assumed to be diagonal ($\Sigma^c = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$). The normalized variance parameters $\sigma_i^2$ are measures of the variability in the MD predictions.
for the corresponding measured quantities. The variance parameters \( \sigma_i^2 \) can be assumed as known constants, that can be controlled by imposing that the MD simulation terminates after the variance of the response quantity of interest has fallen below the \( \sigma_i^2 \) value.

Finally, a diagonal covariance matrix can be a reasonable choice for the modeling error (\( \Sigma^m = diag(s_1^2, s_2^2) \)), where the normalized variance parameters \( s_i^2 \) are unknown constants to be determined by the Bayesian estimation. The homoscedastic case assumes \( s_i^2 \) to be equal for all \( r \). Non-zero mean normal models for \( e^{\text{md}} \) could also be introduced to account for the bias, reconciling conflicting model predictions of various response quantities. The prediction error parameter set \( \eta \) may include the parameters that are involved in the structure of the covariance matrices \( \Sigma^d, \Sigma^t, \Sigma^m \) of the prediction error terms.

### C. Formulation of likelihood

Using the prediction error equation (4) and assuming that the error term in (5) are independent, the measured quantities follow the normal distribution \( \tilde{y} \sim N(g(\theta_m|D), \Sigma(\theta_e)) \), where the covariance matrix \( \Sigma(\theta_e) \) takes the form \( \Sigma^d + \Sigma^t + \Sigma^m \).

Consequently, the likelihood \( f(D|\theta, M) \) of observing the data follows the multi-variable normal distribution given by

\[
f(D|\theta, M) = \frac{1}{(2\pi)^{N/2}} \exp \left[ -\frac{1}{2} J(\theta; \tilde{y}) \right] \pi(\theta|M), \quad (8)
\]

where

\[
J(\theta; \tilde{y}) = [\tilde{y} - g(\theta_m|M)]^T \Sigma^{-1}(\theta_e) [\tilde{y} - g(\theta_m|M)] \quad (9)
\]

is the weighted measure of fit between the MD model predictions and the measured data, and \( | \cdot | \) denotes a determinant.

For relatively large number of experimental data, the posterior distribution of the model parameters can be asymptotically approximated by a Gaussian distribution. This Gaussian is centered at the most probable value of the model parameters with covariance equal to the inverse of the Hessian of the function \( -\ln f(\theta|D, M) \) evaluated at the most probable value.\(^{35}\) For full MD simulations, analytic, differentiable expressions for the model predictions \( g(\theta_m|M) \) involved in (9) are not available. Hence, the estimation of the most probable value requires the use of efficient stochastic optimization algorithms such as the covariance matrix adaptation.\(^{36}\)

For relatively small number of experimental data, the asymptotic estimate may fail to provide an adequate approximation. Moreover, even for large number of experimental data, the asymptotic estimate may fail to give a good representation of the posterior probability distribution in the cases of multimodal distributions, multiple global most probable models, or unidentifiable cases manifested for relatively large number of model parameters in relation to the information contained in the experimental data.\(^{37}\)

Given these drawbacks of the asymptotic approximation and the fact that analytical expressions are not readily available for the prediction of QoI, stochastic simulation algorithms (e.g., variants of MCMC algorithms) represent the only viable solution algorithms. These algorithms are used to populate the important region in the uncertain model parameter space and propagate the uncertainty in the model parameters through the MD model in order to compute the uncertainty in the output QoIs. The algorithms used in this work to generate samples \( \theta^{(i)}, i = 1, \ldots, N \), from the posterior PDF, will be discussed in detail in Sec. IV.

### D. Propagation of uncertainties for robust posterior predictions

Robust posterior predictions of an output QoI \( q \) are obtained by taking into account the updated (posterior) uncertainties in the model parameters given the measured data \( D \).\(^{38}\) Functions of \( q \) can also be predicted by the following methodology.

Let \( f(q|\theta, M) \) be the posterior conditional probability distribution of \( q \) given the model parameters \( \theta \) and the model class \( M \). The posterior robust probability distribution \( f(q|D, M) \) of the output quantity \( q \), taking into account the model \( M \) and the data \( D \), is given by

\[
f(q|D, M) = \int_{\theta} f(q|\theta, M) f(\theta|D, M) d\theta. \quad (10)
\]

This represents an average of the conditional probability distribution weighted by the posterior probability distribution \( f(\theta|D, M) \) of the model parameters. Consistent with the introduction of a Gaussian model for the prediction errors, the conditional PDF \( f(q|\theta, M) \) follows a Gaussian distribution \( N(\mu_q(\theta; M), \sigma_q^2(\theta; M)) \), where \( \mu_q(\theta; M) \) and \( \sigma_q^2(\theta; M) \) are, respectively, the conditional mean and variance of the output quantity \( q \) obtained from the model class \( M \) given a particular value of the model parameters \( \theta \). These mean and variance are computed from the MD simulations for a given \( \theta_m \) value. Using the samples \( \theta^{(i)}, i = 1, \ldots, N \) drawn from the posterior probability distribution \( f(\theta|D, M) \), the integral (10) is approximated by the sample estimate

\[
f(q|D, M) \approx \frac{1}{N} \sum_{i=1}^N f(q|\theta^{(i)}, M). \quad (11)
\]

We define \( G(q) \) as a function of the output QoI \( q \) and its posterior robust performance measure, given the data \( D \), is estimated from:\(^{38}\)

\[
E[G(q)|D, M] = \int E[G(q)|\theta, M] f(\theta|D, M) d\theta. \quad (12)
\]

Setting \( G(q) = q \) and \( G(q) = (q - E[q|D, M])^2 \) in (12), the following sample estimates are obtained for the posterior robust mean:

\[
\hat{\mu} \equiv E[q|D, M] \approx \frac{1}{N} \sum_{i=1}^N \mu_q(\theta^{(i)}, M) \quad (13)
\]

and the posterior robust variance

\[
\hat{\sigma}_q^2 = E[(q - \hat{\mu})^2|D, M] \approx \frac{1}{N} \sum_{i=1}^N \sigma_q^2(\theta^{(i)}, M) + \frac{1}{N} \sum_{i=1}^N [\mu_q(\theta^{(i)}, M) - \hat{\mu}]^2 \quad (14)
\]

of the output quantity of interest \( q \).
E. Model class selection

The Bayesian framework can be used to select the best model class among a family of alternative model classes. Specifically, let \( M^i \), \( i = 1, \ldots, \kappa \), be competitive model classes, with the \( i \)th model class parameterized by the parameter set \( \theta^i \). Let \( Pr(M^i) \) be the prior probability of the \( i \)th model class. Using the Bayes theorem, the posterior probability \( Pr(M^i|D) \) of the model class \( M^i \) given the data are obtained from \(^{28,39}\)

\[
Pr(M^i|D) = \frac{f(D|M^i)Pr(M^i)}{f(D)}, \quad (15)
\]

where \( f(D|M^i) \) is the evidence of the model class \( M^i \) given by (7), and \( f(D) = \sum_{i=1}^{\kappa} f(D|M^i)Pr(M^i) \) is a normalization constant. Assuming that the model classes are equally probable prior to the use of the data, then the most probable model class based on the data corresponds to the model class with the highest evidence. In Sec. VII, alternative model classes for the prediction errors are introduced and compared using the Bayesian model class selection framework.

IV. MARKOV CHAIN MONTE CARLO AND PARALLEL COMPUTATION

Markov chain Monte Carlo algorithms are used to efficiently draw samples from the posterior distribution. MCMC is a framework of algorithms built around the origin Monte Carlo sampling method by Metropolis\(^{40}\) for creating samples from complex probability distributions. A standard MCMC algorithm is the Metropolis-Hastings (MH) Markov chain.

To improve the acceptance rate of trial moves involved in MCMC algorithms, Green and Mira\(^{41}\) introduced the delayed rejection MCMC (DR-MCMC) method. Adaptive MCMC methods\(^{42}\) on the other hand try to change the proposal size on the fly, based on the past samples covariance. The delayed rejection adaptive Metropolis (DRAM) method was proposed as a combination bridging pure adaptive-MCMC and DR-MCMC by Haario et al.\(^{43}\)

There are two ways to introduce a HPC framework to accelerate the calibration procedure. One can parallelize the sampling generation processes (high-level parallelism), or alternatively at the likelihood function by accelerating the individual model runs (low level parallelism).\(^{44}\) In MD simulations, low level parallelization can be achieved with the use of Graphical Processing Units (GPUs) and multicore accelerated MD software.\(^{45}\) Parallelization of the sampling scheme heavily relies on the MCMC method used. Standard MCMC algorithms described above are generically serial since the next sample drawn during the MCMC algorithm depends on the prior one. Parallel MCMC variants exist also such as differential evolution MC,\(^{46}\) or differential evolution random sub-sampling MC\(^{17}\) (DREAM). These methods consist of a population of chains that interact by exchanging information but at the same time preserve the MCMC convergence characteristics at the individual chain.

A. Transitional Markov chain Monte Carlo

The TMCMC\(^{48}\) has been proposed to address the problem of choosing the right adaptive proposal PDF in MCMC methods for accelerating convergence to the posterior PDF. This can be a serious problem when the support of the posterior PDF in the parameter space has complex geometry and/or when the posterior PDF is very peaked and isolated in a small region in the parameter space. Due to a large number of independent parallel chains involved, TMCMC is more efficient in terms of parallel efficiency compared to the DRAM and DREAM algorithms. DRAM is essentially serial and DREAM based on differential evolution suggests the use of a few independent parallel chains\(^{49}\) proportional to the parameter space dimensionality. TMCMC also applies to multimodal posterior PDFs as it handles efficiently very peaked or very flat PDFs along certain directions in the parameter space, as well as it estimates the evidence \( f(D|M) \) which can further be used for model selection.

The TMCMC is a generalization of the method proposed by Beck and Au\(^{50}\) extended by notions inherent to simulated annealing algorithms and as such it can be categorized in the framework of evolutionary strategy MCMC methods.\(^{51}\) According to the TMCMC algorithm,\(^{48}\) a series of intermediate probability distributions

\[
f_j(\theta) \sim [f(D|\theta, M)]^{p_j} \cdot \pi(\theta|M), \quad j = 0, \ldots, m
\]

are constructed iteratively. The process starts by generating samples from the prior probability distribution \( f_0(\theta) \sim \pi(\theta|M) \) and continuing with a series of sampling operations for each intermediate stage \( j = 1, \ldots, m \) as follows: given the \( N_j \) samples \( \{\theta_{j,k}, k = 1, \ldots, N_j\} \) from the intermediate probability distribution \( f_j(\theta) \) at stage \( j \), generate \( N_{j+1} \) samples \( \{\theta_{j+1,k}, k = 1, \ldots, N_{j+1}\} \) from the next PDF \( f_{j+1}(\theta) \) at stage \( j + 1 \) based on the plausibility weights of the \( N_j \) samples drawn from \( f_j(\theta) \) with respect to the probability distribution \( f_{j+1}(\theta) \).

\[
\frac{w(\theta_{j,k})}{f_{j+1}(\theta_{j,k})} = \frac{f_{j+1}(\theta_{j,k})}{f_j(\theta_{j,k})} = f(D|\theta_{j,k}, M)^{p_{j+1}-p_j}, \quad (17)
\]

The choice of the exponents \( p_j \) controls the transition between adjacent probability distributions which in turn controls the rate of convergence and effectiveness of TMCMC. The coefficient of variation \( COV_p(j) \) of the plausibility weights \( w(\theta_{j,k}) \) at stage \( j \) is a good indicator of the smoothness of this transition. The choice of the \( p_{j+1} \) value is controlled automatically by the TMCMC algorithm so that the \( COV_p(j) \leq tolCOV \), where \( tolCOV \) is a prescribed tolerance.

The steps involved in the TMCMC algorithm (Algorithm I) are summarized next in order to reveal the parallelization aspects of the algorithm.

In the present work the resampling step is performed as follows:

1. Choose \( l = 1, \ldots, N_{j+1} \) samples from the \( \{\theta_{j,k}, k = 1, \ldots, N_j\} \) sample points, with probability equal to \( \tilde{w}(\theta_{j,k}) \). The same point can be chosen multiple times. A selected sample is called leader, and is denoted by
Algorithm I (TMCMC) Ref. 48.

BEGIN, SET $j = 0, p_0 = 0$

- **Generate** $\{\theta_{0,i}, k = 1, \ldots, N_0\}$ from the prior $f_0(\theta) = \pi(\theta|M)$ and compute the likelihood $f(D|\theta_{0,i}, M)$ for each sample.

WHILE $p_{j+1} \leq 1$ DO:

- **Analyze** the samples $\{ \theta_{j,k}, k = 1, \ldots, N_j \}$ at stage $j$ to estimate the exponent $p_{j+1}$, the normalized plausibility weights $\tilde{w}(\theta_{j,k})$ from $w(\theta_{j,k})$ in (17), the covariance matrix $\Sigma_j$ based on the available samples from $f(\theta)$ and the estimator $S_j$ of $E[w(\theta_{j,k})]$, as follows:

(a) Choose the value of the exponent $p_{j+1}$ so that the covariance of the plausibility weights satisfy $COV_{w}(j) = t_0COV$. Care must be taken so that the calculated $p_{j+1}$ satisfies the conditions in (16), i.e., $0 \leq p_{j+1} \leq 1$. Therefore, finding of the root can be translated to a constrained minimization problem,52

(b) Compute the plausibility weight $w(\theta_{j,k}) = \prod_{l=0}^{\infty} f(|D_1, \ldots, D_l; \theta_{j,l}|) f_{j+1}^{p_{j+1}}$ for all the samples $\{\theta_{j,k}, k = 1, \ldots, N_j\}$ in the stage $j$. Also estimate $S_j = \frac{1}{N_j} \sum_{k=1}^{N_j} w(\theta_{j,k})$ which is an asymptotically unbiased estimator for $E[w(\theta_{j,k})]$.

(c) Estimate the normalized plausibility weights at stage $j$,

$$\tilde{w}(\theta_{j,k}) = w(\theta_{j,k}) / \sum_{i=1}^{N_j} w(\theta_{j,i}) = w(\theta_{j,k})/(S_j N_j)$$ (18)

for each sample $\theta_{j,k}$ and then estimate the interchain covariance matrix:

$$\Sigma_j = \beta^2 \sum_{k=1}^{N_j} \tilde{w}(\theta_{j,k}) [\theta_{j,k} - \mu_j] [\theta_{j,k} - \mu_j]^T$$ (19)

based on the samples $\{\theta_{j,k}, k = 1, \ldots, N_j\}$, where $\beta^2$ is a given scaling parameter and $\mu_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \tilde{w}(\theta_{j,i}) \theta_{j,i}$ is the sample mean.

- **Resample** based on the samples available in stage $j$ in order to generate samples for stage $j + 1$ distributed as $p_{j+1}(\theta)$:

**IF** $p_{j+1} > 1$ **BREAK, ELSE** $j = j + 1$

END

The computation of a sample $\theta_{j+1,k}$ in any chain requires the estimation of the likelihood $f(D|\theta_{j+1,k}, M)$ needed in the intermediate probability distribution $f_{j+1} = \prod_{l=0}^{\infty} f_{j+1}^{p_{j+1}}$ that is involved in the MH step performed within a chain. The $N_{j+1}$ Markov chains can run independently in parallel. Efficient parallel implementation strategies are detailed in Sec. VI. An asymptotically unbiased estimate of the evidence $f(D|M)$ of the model class $M$ is readily provided from the TMCMC algorithm as $\prod_{j=0}^{\infty} S_j$.48

In the resampling of the original TMCMC algorithm, the MH with delayed rejection can be used to estimate the samples in each chain, improving the acceptance rate and thus the autocorrelation decay of the generated samples.

V. ADAPTIVE SURROGATE MODELS

The most computationally intensive part of the algorithm lies in the likelihood function evaluation in the resampling step, requiring multiple MD model runs. The use of surrogate models offers the possibility to drastically reduce the computational efforts, while still achieving highly accurate approximations.

Surrogate models are especially well suited for use with the TMCMC algorithm since at each intermediate stage, a large number of samples that sufficiently cover the supports of the intermediate posterior PDFs from the current and previous stages are available to be used as design points for approximating the likelihood estimate at a new sample based on a surrogate technique.

Surrogates are meta-models representing a functional relation between the input parameters and the model output QoI. Response surfaces have been introduced by Ref. 54 to represent the relationship between the parameter space and an output QoI. While there is a plethora of approaches,55 the most commonly used meta-models are based on linear or polynomial regression, on a least-squares formulation or on kriging and radial basis functions. Here we implement kriging and the software DACE56 is integrated in the TMCMC algorithm. Such local interpolation schemes have been shown to be beneficial in related stochastic optimization algorithms,57 gathering information from the region in the parameter space that is closest to the surrogate point. The computational savings are highly dependent on the user requirements for accuracy and the local smoothness of the function to be approximated, which is not known a priori.

Here a surrogate estimate is performed and accepted if it simultaneously obeys the following heuristic rules:

1. The design points used for interpolation correspond to real function evaluations.
2. The surrogate estimate is based on a user-defined minimum number $n_{\text{neigh}}$ of design points $\{\theta_{i,\text{neigh}}, i = 1, \ldots, n_{\text{neigh}}\}$ that are in the neighbor of the point $\theta^C$ This minimum number depends on the dimension of the uncertain parameter space.
3. The surrogate point must belong to the convex hull of the design (real) points. In this way only interpolations are performed, while extrapolations are prohibited by doing a simple neighborhood search around the point to
be approximated, and ensuring that the point lies in the $n$-dimensional convex hull of its basis points.\textsuperscript{58} 

4. A local approximation consistent with the proposal probability distribution at stage $j$ is performed by selecting the neighbor design points within the hyper-ellipse of the TMCMC proposal covariance matrix $\gamma \Sigma_{j+1}$. If the number of points within the hyper-ellipse are equal or greater than $n_{\text{neigh}}$, and $\theta^C$ belongs to the convex hull of the design points $\{\theta^C_{i,\text{neigh}}, \ i = 1, \ldots, n_{\text{neigh}}\}$, then we construct the surrogate approximation. A surrogate estimate is not allowed if $\gamma$ exceeds a pre-specified number so that only local estimates are accepted.

5. The surrogate estimate is accepted only when the variance of error of the surrogate estimate is smaller than a pre-specified value.

Surrogate models are especially attractive for uncertainty propagation in MD simulations. When all trajectories associated with each individual MD run through the course of the calibration procedure are stored, one ends up with all the necessary information around multiple design points. This information can be subsequently used to generate, even offline (without any further real model evaluations), data required for the uncertainty propagation of any output QoI.

VI. PARALLELIZING THE TMCMC

The parallel implementation of the TMCMC is next described, taking into account that the most time consuming operation is the evaluation of the likelihood function. This evaluation involves $n_s$ full MD model simulation runs for a given value of the parameter set, where $n_s$ is the number of different temperature and pressure conditions under which the experimental data are available. The $n_s$ model runs can be performed concurrently with at most $n_c$ compute cores.

At the level of the TMCMC algorithm, a parallel implementation strategy can be activated at every stage $j$ of the algorithm. Specifically, at stage $j$ a number of $i = 1, \ldots, N_{j+1}$ chains have to be generated using the leader samples $\theta_{i,\text{leader}}^0$. Each chain $i$ can be thought as a MH chain starting at the leader sample $\theta_{i,\text{leader}}^0$ and having a fixed length $n_{j+1}$, determined by the number of times the lead sample is selected. In particular, for $j = 0$ the number of chains equals the number of samples for stage $j = 0$ ($N_1 = N_1$), while the number of samples per chain equals $n_{1,j} = 1$. The process of generating the samples $\theta_{j,k}^0, k = 1, \ldots, n_{j+1}$ within the chain $i$, starting from the leader sample $\theta_{j,0}^0 = \theta_{i,\text{leader}}^0$ is in general sequential, as each sample in the chain $i$ depends on the previous. However, the computations within a chain $i$ are completely independent from the computations in the rest of the chain so that their evolution can be performed concurrently.

The length $n_{j+1}$ of a chain $i$ specifies the number of individual likelihood evaluations, and therefore is a direct indication of the computational effort required to form the samples. The length $n_{j+1}$ differs for each chain within a stage or across all stages $j = 1, \ldots, m - 1$. The total computational time for the TMCMC algorithm, including operations for all stages, is proportional to the $\sum_{j=0}^{m-1} \hat{n}_{j+1}$.

The proposed parallelized software implementation uses a master-worker configuration, enabling a multi-host configuration of complete heterogeneous workers. It works in the following way: the master node estimates the distribution of chains per worker according to the weighted round-robin scheduling scheme and sends to each worker the parameters of the lead samples $\theta_{j,0}^{\text{leader}}, i = 1, \ldots, n_{j+1}$, together with the information about the length $n_{j+1}$ of the respective chains. The generation of the Markov chains assigned to each worker is then performed in parallel for all workers. The output of each worker consists of the samples for each chain in the worker and the function/likelihood evaluations which are sent back to the master node. After receiving all samples at the stage $j$, the master node evaluates the probability weights needed for the next stage $j + 1$ based on Algorithm I, and the resampling step 1 of the TMCMC algorithm. The iteration is repeated for the next stages $j + 1, \ldots, m - 1$ until samples of the final stage involving the non-scaled posterior PDF are generated, i.e., until the exponent $p_j = 1$ for $j = m$.

The aforementioned scheduling scheme is referred to as a static scheduling scheme which is computationally efficient as long as: (a) the computational time for a likelihood evaluation is independent of the value of the sample $\theta_{j,0}^0$ in a chain, which is usually the case only for a limited class of systems, (b) surrogate modeling is not activated to replace the time consuming operation of the likelihood evaluation by a much faster approximate surrogate estimate, and (c) delayed rejection is not activated to increase the computational time involved in the chain due to the extra likelihood evaluations that may arise once a sample in the chain is rejected. For the cases where such conditions do not hold, it is more appropriate to use a dynamic scheduling scheme for which each worker is periodically interrogated at pre-specified regular time intervals by the master computer about its availability, and a sample from a chain or a whole chain is sent to the workers on the first come first serve basis to perform the likelihood function evaluations so that the idle time of the workers is minimized during a TMCMC stage.

VII. RESULTS

We demonstrate the present framework by estimating the parameters $\sigma_{11}$, $\epsilon_{11}$, and cut-off radius $r_c$ of a Lennard-Jones force field in MD simulations of argon, using experimentally measured self-diffusion coefficients and radial correlation functions. In Sec. VII A, inspired by the work of Cailliez and Pernot,\textsuperscript{18} we perform the UQ+P analysis using analytical functions for the self diffusion coefficient $D_{11}$, available for argon based on a model for the kinetic theory of gases. In Sec. VII B, we apply the UQ+P framework using full MD simulations.

The self-diffusion data, as taken from Refs. 59–61 for argon at the gas state, are presented in Table I.

A. UQ+P using analytical expressions

Analytical expressions relating the diffusion coefficient $D_{11}$ with the LJ parameters $\sigma_{11}$ and $\epsilon_{11}$ at different
temperatures and pressures,\textsuperscript{53, 62} can be used to carry out the likelihood evaluation \( f(\theta | D, M) \), allowing for rapid evaluation of a very large number of samples.

We introduce the simplest prediction error model class, denoted by PEM\(_1\), which includes measurement and model errors. The covariance matrix \( \Sigma^d \) of the measurement error is assumed to be diagonal with elements the variances \( \hat{v}_i^2 \) of the measurements as given in the last column of Table I. A diagonal covariance matrix \( \Sigma^m = s^2 \text{diag}(\hat{v}_i^2) \), is assumed for the model error, where \( s \) is the relative error. Hence the covariance matrix of the total prediction error is \( \Sigma = \text{diag}(\hat{v}_i^2 + s^2 \hat{v}_i^2) \). The standard deviation \( s \) is included in the parameter set \( \theta = (\sigma_{\text{LJ}}, \epsilon_{\text{LJ}}, s) \) to be estimated by the Bayesian methodology. A second prediction error model class, denoted by PEM\(_2\), is introduced to account for different variances in the model errors of the predictions for two different temperature regimes appearing in the experimental data. Specifically, the variance of the model error for all observations with \( T \leq 120 \) is introduced to be \( s_{\text{low}}^2 \), while the variance of the model error for all observations with \( T > 120 \) is introduced to be \( s_{\text{high}}^2 \). This augments our parameter space dimension to four: \( \theta = (\sigma_{\text{LJ}}, \epsilon_{\text{LJ}}, s_{\text{low}}, s_{\text{high}}) \). Note that the PEM\(_2\) model contains the PEM\(_1\) in the sense that the PEM\(_2\) can provide the same predictions as the PEM\(_1\) with an additional degree of freedom that gives flexibility to improve the fit in the experimental data. The selection of the prediction error model PEM\(_2\) is arbitrary, depends on the experimental data set used and it is not meant to be justified on rational modeling assumptions. It is introduced as an example to demonstrate the capabilities of the proposed framework to carry it with alternative prediction error models for further improving the fit to the experimental data.

We use the setup for the TMCMC method as shown in Table II. The prior probability distribution \( \pi(\theta | D) \) of the model parameters is considered to be uniform on the support \([0.2, 0.5] \times [50, 350] \times [0.001, 1.0] \) in the three-dimensional parameter space for PEM\(_1\), and on the support \([0.2, 0.5] \times [50, 350] \times [0.001, 1.0] \times [0.001, 1.0] \) in the four-dimensional parameter space for PEM\(_2\). The units for \( \sigma_{\text{LJ}} \) is nm and for \( \epsilon_{\text{LJ}} \) is K.

### Table I. Data set for kinetic calibration using the self-diffusion coefficient \( D_{11} \) for various temperature and pressures. The last column \( \nu_i \) refers to the measurement uncertainties reported.

<table>
<thead>
<tr>
<th>( \nu_i ) ( \propto D_{11} \times 10^5 )(\text{cm}^2/\text{s})</th>
<th>( T )(K)</th>
<th>( P )(MPa)</th>
<th>( \nu_i \times 10^3 )(\text{cm}^2/\text{s})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0619</td>
<td>78.0</td>
<td>0.0212</td>
<td>0.00183</td>
</tr>
<tr>
<td>0.1120</td>
<td>85.1</td>
<td>0.1553</td>
<td>0.00066</td>
</tr>
<tr>
<td>0.0320</td>
<td>90.0</td>
<td>0.0574</td>
<td>0.00096</td>
</tr>
<tr>
<td>0.0484</td>
<td>91.8</td>
<td>0.0382</td>
<td>0.00014</td>
</tr>
<tr>
<td>0.0337</td>
<td>94.6</td>
<td>0.0589</td>
<td>0.00011</td>
</tr>
<tr>
<td>0.0120</td>
<td>100.7</td>
<td>0.0653</td>
<td>0.00100</td>
</tr>
<tr>
<td>0.0456</td>
<td>120.7</td>
<td>0.0735</td>
<td>0.00010</td>
</tr>
<tr>
<td>0.08300</td>
<td>194.5</td>
<td>0.1013</td>
<td>0.001</td>
</tr>
<tr>
<td>0.1560</td>
<td>273.0</td>
<td>0.1013</td>
<td>0.002</td>
</tr>
<tr>
<td>0.1780</td>
<td>295.0</td>
<td>0.1013</td>
<td>0.002</td>
</tr>
<tr>
<td>0.2120</td>
<td>326.0</td>
<td>0.1013</td>
<td>0.002</td>
</tr>
<tr>
<td>0.2490</td>
<td>353.0</td>
<td>0.1013</td>
<td>0.003</td>
</tr>
</tbody>
</table>

### Table II. TMCMC settings for the estimation of the parameters of the analytical model.

<table>
<thead>
<tr>
<th>TMCMC parameter</th>
<th>No surrogates</th>
<th>Surrogates</th>
</tr>
</thead>
<tbody>
<tr>
<td>tolCov</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Samples per stage ( N_j )</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>( N_j ) (last stage)</td>
<td>5000</td>
<td>5000</td>
</tr>
<tr>
<td>Number of stages</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Surrogate model</td>
<td>...</td>
<td>Kriging</td>
</tr>
</tbody>
</table>

### 1. Parameter estimation

The TMCMC samples drawn from the posterior distribution of the model parameters are presented as a \(3 \times 3\) plot matrix in Figure 1 for the model class PEM\(_1\). In Figure 1, we present the projections of the samples in the two-dimensional spaces \((\sigma_{\text{LJ}}, \epsilon_{\text{LJ}}), (\sigma_{\text{LJ}}, s),\) and \((\epsilon_{\text{LJ}}, s)\). Results for the marginal distributions of the LJ model parameters \(\sigma_{\text{LJ}}, \epsilon_{\text{LJ}},\) and the prediction error parameter \( s \) are also shown in the diagonal of the plot matrix. From the projections in the \((\sigma_{\text{LJ}}, \epsilon_{\text{LJ}})\) space, it is observed that the important region of high probability volume of the posterior PDF is directed along preferred elongated directions in the parameter space. This strong correlation, also reported in Ref. 18, is expected to affect predictions of uncertainties in the thermodynamic and other output QoI.

Table III summarizes the results of the parameter estimation obtained for the model classes PEM\(_1\) and PEM\(_2\) using the TMCMC algorithm. Results are reported for the most

---

The content includes a table for the data set used in kinetic calibration, settings for the TMCMC method, and a discussion on parameter estimation. The table and figures provide insights into the prediction error model classes and the TMCMC setup.
The effectiveness of the surrogate models to accurately quantify and propagate uncertainties is investigated next. The optimal values of the calibrated parameters and the standard deviation of the marginal distributions of the model parameters, obtained by activating the kriging surrogate approximations for the model class PEM1, are given in Table III. The marginal distribution of the model parameters are also drawn in the diagonal plots of Figure 1. Comparing these results with the ones obtained without activating surrogate estimates, it can be concluded that the surrogate estimates provide a very good estimate, making them suitable for using with the TMCMC algorithm. Discrepancies are manifested mainly in the standard deviations of the marginal distributions of the model parameters, which are reported to be slightly higher for the surrogate approximations. Note that these small discrepancies are attributed to the statistical nature of the TMCMC algorithm.

In order to further quantify the discrepancy between the two obtained posterior distributions \( \tilde{f}(\theta|D, M) \) and \( f(\theta|D, M) \) for the PEM1 model, estimated with and without the use of surrogates, respectively, we employ the symmetric Kullback-Leibler (KL) divergence metric.\(^{67}\) This distance is found \( KL = 0.0016 \) which is very small. The number of surrogate evaluations in relation to the total number of function evaluations is also shown as points in Figure 1 in the \((\sigma_{LJ}, \epsilon_{LJ})\) plot. Despite the fact that the model and/or the experimental data used in this study and the studies reported in the literature are different, it can be seen that these values fall close to the correlated uncertainty region estimated by the Bayesian methodology.

2. Use of surrogate models

The optimal value of the parameter \( \epsilon_{LJ} \) in Table III is very close to the value \( \epsilon_{LJ} = 0.33605 \) reported in Ref. 66 which calibrated the parameters using second virial values. Moreover, the parameters \( \sigma_{LJ} = 142.095 \) and \( \epsilon_{LJ} = 125.7 \) of Refs. 65 and 66, which are calibrated using second virial coefficients, are contained within the uncertainty regions of \( \epsilon_{LJ} \) reported in Table III for PEM1 and PEM2, respectively. For a more fair comparison that takes into account the correlation of the uncertainty region in the \((\sigma_{LJ}, \epsilon_{LJ})\) plane, values from the literature\(^{18,65–66}\) are also shown as points in Figure 1.
is 72% which results in a reduction of the computational time due to surrogate estimation by a factor of three. A similar strategy evaluation but with relaxed rules regarding the local surrogate approximation, allowing more than 86% of the total evaluations to be evaluated by a surrogate, lead to a value of KL = 0.0121 and a reduction of the computational time by a factor of six. The Q-Q plots between the posterior distribution \( \hat{f}(\theta|D, M) \) and \( f(\theta|D, M) \), shown in Figure 2 for the three model parameters, quantify the discrepancies between the two distributions. These Q-Q plots suggest that there are slight differences in the two distributions. In particular, the steeper trend of the Q-Q plot than the line with slope 1 suggests that the standard deviation of the distributions based on the surrogate estimates is higher than the standard deviation of the distribution obtained without activating surrogates. This is consistent with the results obtained in Table III or shown in the diagonal plots in Figure 1.

### 3. Uncertainty propagation

Following the estimation of the posterior distribution of the model parameters using the PEM1, we propagate the uncertainty in the parameter estimates through the same model to compute the uncertainty in output quantities. Figure 3(a) presents the probability distributions for the self-diffusion coefficients \( D_{11} \) for the 12 sets of temperatures and pressures reported in Table I. The estimation of the posterior probability distribution for \( D_{11} \) is done using a Gaussian kernel estimate applied on the sample values \( D_{11}(\theta^{(i)}), i = 1, \ldots, N \), obtained from the samples \( \theta^{(i)} \) that are generated at the last stage of the TMCMC algorithm. The results in Figures 3(a) are based on both real function evaluations and surrogate evaluations, demarcated with different colors. The uncertainties in the experimental measurements of \( D_{11} \) are also drawn in these figures. Specifically, the normal distributions corresponding to the experimental mean and standard deviation reported in Table I, are drawn in Figure 3 with the green lines. Similar comparisons for the viscosity predictions for the 12 temperature and pressure sets are shown in Figure 3(b). The values for the viscosity and their uncertainty are taken from NIST databases,68,69 while the analytical expressions used to predict the viscosity are taken from Ref. 62.

The experimental measurements and their associated uncertainties of \( D_{11} \) is captured well by the model for almost all experimental conditions. The uncertainty region (yellow normal distribution) as was identified experimentally for the self-diffusion coefficient is, in most cases, contained within the uncertainty region predicted by the model. It is observed that the fit is not good for the first two experimental measurements where the experimental uncertainty region fall outside the uncertainty region predicted by the analytical expressions. This is due to the model inadequacy to simultaneously fit all the 12 experimental measurements and uncertainties. The inadequacy of the analytical expressions and prediction error model may be compensated by introducing prediction error models with more parameters such as PEM2. In general, it can be stated that the analytical expressions are good prediction models for the \( D_{11} \) of the argon system in the temperature and pressure regime used for calibration of the model parameters. However, the uncertainty in the viscosity predictions, propagated through the model using the identified uncertainties in the LJ model parameters, do not adequately capture the significantly higher experimental uncertainty in viscosity. Analytical expressions for the self-diffusion coefficients and viscosity can be simultaneously used with the same LJ model parameters to predict the available data set measurements. Finally, the results in Figure 3 for the surrogate estimates illustrate their effectiveness in accurately modeling and propagating uncertainty.

### B. UQ+P using full molecular dynamics simulations

In the absence of analytical expressions relating various thermodynamic quantities with model parameters, as is commonly the case, full MD simulations are necessary. To demonstrate the effectiveness of the proposed HPC approach for Bayesian UQ+P, the uncertainties in the parameter estimates are next estimated using the MD simulations for argon. The MD simulations protocol used for predictions can be found in the Appendix. Unless specified otherwise, all results are obtained using the prediction error model PEM1.

#### 1. Parameter estimation

The results presented in this section were obtained with the setup for the TMCMC method as shown in Table IV. To illustrate accuracy issues related with the MD simulation protocol, the parameter estimation is based on the same measured self-diffusion data used with the analytical expressions (see Table I). Results are obtained from the TMCMC algorithm using the prediction error model PEM1. We remark that the MD-simulation runs are terminated when the standard deviation in the predictions of the self-diffusion coefficients, computed from the simulation protocol for various temperature and pressure conditions, fall below the pre-selected values of \( \sigma_i = 5\% \). Thus the computational error is expected to be of the order of 5\%. The covariance matrix of the prediction error is \( \Sigma = \text{diag}(\hat{\sigma}_r^2 + s^2 \hat{\sigma}_e^2) \), where the normalized variance \( s^2 \) in the MD simulation case takes into account both the computational and model error.

The computational efficiency of the parallel implementation of the TMCMC algorithm is next demonstrated. Table V presents the number of independent chains and the maximum chain length per stage generated. The results indicate that at least 300 or more independent chains can run in parallel, efficiently exploiting our available computational resources containing 48 nodes of 16 cores each. The computational cost for the simulation campaign associated with this system, containing 800 atoms, was 3 CPUhrs per likelihood evaluation, while the total cost for all the runs was \( \approx 21\,600 \) CPUhrs. Without the use of surrogate estimates approximately 420,000 CPUhrs would have been required. More importantly, had we used a generically serial algorithm like DRAM, the time-to-solution would then be approximately 3 and 1/2 months for the same computational resources. Using the proposed parallelization of TMCMC, the time-to-solution is down to 3 days using 48 nodes of 16 cores each. This shows the superiority of TMCMC in terms of time-to-solution, making it a key tool for UQ+P in HPC facilities.
FIG. 3. Propagation of uncertainty for PEM1: Probability distribution function of (a) the self-diffusion coefficient and (b) the viscosity for the temperatures and pressures presented in Table I. Solid curves with red filling indicate model prediction uncertainties. Solid curves with yellow filling indicate uncertainties based on experimental data. Dashed curves with blue filling indicate surrogate estimations.
The probability distribution of the model parameters are shown as a $3 \times 3$ plot matrix in Figure 4. The numbers in the plots above the diagonal in Figure 4 indicate the Pearson correlation number for the scatter pairs which give a measure of correlation between the two parameters. From the projections in the $(\sigma_{ij}, \epsilon_{ij})$ space, the uncertainties in the LJ parameters are strongly correlated. This result is consistent with the one obtained using the analytical expressions. Comparing the marginal distributions of the LJ model parameters in Figure 4 with the corresponding marginal distribution in Figure 1, it is noted that the region of uncertainty of the LJ model parameters estimated from the full MD simulations contains the region of uncertainty estimated from the analytical expressions. Overall, it can be concluded that estimation of the LJ parameters based on the full MD simulations and the analytical expressions are consistent. The 95% confidence interval for the prediction error parameter $s$ in Figure 4 is approximately [3.0%, 6.5%] with most probable value equal to 4.6%. These values are significantly higher than those obtained from the model based on the analytical expressions (see Figure 1) due to the fact that the prediction error for full MD simulations includes also the computational error of 5%. As expected, the most probable value of $s$ is close to the value 5% of the assumed computational error. The confidence interval includes values higher than 5%, attributed to the extra model error involved in the MD model class simulations.

The most probable values and standard deviations of the marginal posterior probability distribution of each one of the model parameters are reported in Table III for both the “no surrogates” and “surrogates” cases. The differences observed with the ones obtained based on the empirical relations in Sec. VII A are due to two different sources of the model errors. First, the model error arising from the use of full MD model is qualitatively different from the model error arising from the use of the available analytical expressions for predicting the self-diffusion coefficient. The model error associated with the MD simulations arises from several factors, including the selection of the size of the reference finite volume, the number of particle used within the volume for predictions, the potential cut-off value used and the finite time sampling of the diffusion coefficient. Second, the full MD model runs contain an extra computational error due to the variability in the estimate of the self-diffusion coefficient provided by the MD simulation. Such model errors and variabilities are not relevant in the case of analytical expressions.

The log evidence of the full MD model is also evaluated from the TMCMC algorithm and reported in Table III. The log evidence of the model class based on full MD simulations is 19.32 which is slightly higher than the log evidence 19.02 of the model class based on the analytical expressions. These small differences indicate that the two model classes are almost equally probable with the model class based on full MD simulations to have slightly higher preference than the model class based on the analytical expressions. These results also demonstrate the adequacy of the analytical expressions to provide predictions of the diffusion coefficient that are almost as good as the ones provided by the full MD simulations. The log evidence value predicted by activating the surrogates is sufficiently close to the log evidence value predicted without the surrogates, demonstrating the high accuracy of the surrogate estimates. In contrast, the most probable value of the parameter set predicted for the surrogate case, compared to the one predicted by the no surrogate case, might not be a good indicator of the accuracy of the surrogate estimates to approximate the posterior PDF since a surrogate sample may involve large error that can cause a shift in the location of the optimum of the posterior PDF. This explains the discrepancy of the optimal value of $\epsilon_{ij}$ observed in Table III for the surrogate and no surrogate cases.

### 2. Uncertainty propagation

The uncertainty in the model parameters is propagated through the full MD model to predict the posterior uncertainty in the following properties: Density $\rho$, thermal expansion coefficient $\alpha$, isothermal compressibility $\beta$, enthalpy $(\Delta H)$ as well as the viscosity $\eta$. The results are presented in Table VI for “no surrogate” runs. The confidence intervals of these properties, quantified by the mean plus/minus one standard deviation, are shown in Table VI for two $(T, P)$ pairs. Note that the information required for the calculation of these properties is readily available during the calibration process, so essentially they can be computed at negligible extra cost. The experimentally observed values along with their uncertainties for the two $(T, P)$ pairs as determined from NIST database, or other works are also reported in Table VI.

Highly accurate robust predictions of the density $\rho$ and the enthalpy $\Delta H$ are provided that are very close to the corresponding experimental measurements. Large uncertainty are associated with the predictions of the thermal expansion coefficient $\alpha$, isothermal compressibility $\beta$ and the viscosity $\eta$. The high level of uncertainties in relation to the uncertainties observed in the experimental quantities is partly due to the large fluctuation of these properties arising from short time intervals used in MD simulations. The uncertainties in the experimental measurements is within the uncertainties predicted by the Bayesian formulation for most properties in Table VI.

---

**TABLE IV. TMCMC settings for the MD-driven calibration.**

<table>
<thead>
<tr>
<th>TMCMC parameter</th>
<th>No surrogates</th>
<th>Surrogates</th>
</tr>
</thead>
<tbody>
<tr>
<td>tolCov</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Number of samples $N_j$</td>
<td>980</td>
<td>3000</td>
</tr>
<tr>
<td>Number of stages</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Surrogate model</td>
<td>…</td>
<td>Kriging</td>
</tr>
<tr>
<td>Surrogate ratio</td>
<td>…</td>
<td>61%</td>
</tr>
</tbody>
</table>

**TABLE V. TMCMC-parallel performance for full, no surrogates, MD runs, showing the number of independent chains and the maximal chain length per stage.**

<table>
<thead>
<tr>
<th>TMCMC-stage $j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{j+1}^{\text{max}}$</td>
<td>980</td>
<td>817</td>
<td>712</td>
<td>713</td>
<td>701</td>
<td>502</td>
<td>427</td>
<td>421</td>
<td>412</td>
<td>316</td>
<td>309</td>
</tr>
<tr>
<td>$\hat{n}_{j+1}$</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>15</td>
<td>17</td>
<td>20</td>
</tr>
</tbody>
</table>
The uncertainties in the predictions can be reduced by using longer time intervals in MD simulations as seen in Ref. 18 and/or using better MD model classes involving a larger system size, higher order integration schemes, etc.

Figure 5(a) presents the probability distributions for the self-diffusion coefficients $D_{11}$ for the 12 sets of temperatures and pressures reported in Table I. The results in Figure 5(a) are based on both real function evaluations and surrogate evaluations. The uncertainties in the experimental measurements of $D_{11}$ are also drawn in these figures with green lines. Similar comparisons for the viscosity predictions for the 12 temperature and pressure sets are shown in Figure 5(b).

The experimental measurements and their associated uncertainties of $D_{11}$ are well captured by the model for almost all experimental temperatures and pressures. The uncertainty region (yellow normal distribution) as was identified experimentally for the self-diffusion coefficient is contained well within the uncertainty region predicted by the model. This implies that the full MD simulations are good prediction models for the $D_{11}$ property of the argon system in the temperature and pressure regime used for calibration of the model parameters. The uncertainty in the viscosity predictions, propagated through the full MD model using the identified uncertainties in the LJ model parameters, also capture well the experimental viscosity and its uncertainty. Comparing the results in Figure 5(b) with those obtained in Figure 3(b) based on the analytical expressions, it can be concluded that the full MD model is an adequate model for providing reliable robust

TABLE VI. Uncertainty propagation for various QoI based on the kinetic calibration. Values reported are the most probable value plus/minus ($\pm$) a standard deviation of the PDF of the output quantities.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$P$ (MPa)</th>
<th>$\rho$ (kg/m$^3$)</th>
<th>$\alpha$ (1/K) $\times 10^2$</th>
<th>$\beta$ ($\text{MPa}^{-1}$) $\times 10^2$</th>
<th>$\Delta H$ (J/mol K)</th>
<th>$\eta(\text{cP})$ $\times 10^{-2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt. 91.8</td>
<td>0.0382</td>
<td>2.0210 $\pm$ 0.0100</td>
<td>1.124 $\pm$ 0.034</td>
<td>2.647 $\pm$ 0.79</td>
<td>1852.91 $\pm$ 21.985</td>
<td>0.752 $\pm$ 0.070</td>
</tr>
<tr>
<td>MD 91.8</td>
<td>0.0382</td>
<td>2.0357 $\pm$ 0.0160</td>
<td>2.714 $\pm$ 1.2515</td>
<td>6.211 $\pm$ 2.92</td>
<td>1875 $\pm$ 21.8</td>
<td>0.65 $\pm$ 0.13</td>
</tr>
<tr>
<td>Expt. 353.0</td>
<td>0.1013</td>
<td>1.3791 $\pm$ 0.0069</td>
<td>0.2838 $\pm$ 0.0085</td>
<td>0.987 $\pm$ 0.030</td>
<td>7331.73 $\pm$ 36.5</td>
<td>2.60 $\pm$ 0.26</td>
</tr>
<tr>
<td>MD 353.0</td>
<td>0.1013</td>
<td>1.3823 $\pm$ 0.0014</td>
<td>0.1832 $\pm$ 0.0297</td>
<td>0.662 $\pm$ 0.1051</td>
<td>7322 $\pm$ 35</td>
<td>2.23 $\pm$ 0.421</td>
</tr>
</tbody>
</table>
FIG. 5. Propagation of uncertainty for PEM1: Probability distribution function of (a) the self-diffusion coefficient and (b) the viscosity for the temperatures and pressures presented in Table I. Solid curves with red filling indicate model prediction uncertainties. Solid curves with yellow filling indicate experimental measurements with their associated uncertainty. Dashed curves with blue filling indicate surrogate estimations.
predictions for the viscosity, while the analytical expressions fail to provide reliable predictions consistent with uncertainties identified from experimental data. The results in Figure 5 suggest that surrogate models are quite effective in modeling and propagating uncertainty in MD.

3. Structural calibration

The parameter estimation in this section is based on the experimentally measured radial distribution function $g(r)$ from x-ray scattering experiments, shown by red dots in Figure 6 (a total of 338 experimental observations).

This example demonstrates the Bayesian parameter estimation for a molecular simulation force field based on medium-scale simulations. There are multiple formulation linking the LJ potential parameters via an integral equation to the liquid state radial distribution function, via Meyer function expansions.72, 73 These formulations treat the particle interactions as short range, whereas the structural behavior changes when they allow for long range interactions. In order to mimic this behavior in the MD model, in addition to the $\sigma_{\text{LJ}}, \epsilon_{\text{LJ}}$, an extra free parameter is introduced by letting the potential cut-off value $r_c$ to lie within 0.8 nm and 1.5 nm. This introduces one more degree of freedom to the MD model since $r_c$ is another parameter for the potential of model $M$, using the force-shifted LJ potential.74

The Bayesian estimation is performed by setting the TMCMC parameters to $t_{\text{ol Cov}} = 1$, $\beta = 0.15$, and $N_{j} = 560$, resulting in 8 TMCMC stages. Surrogate estimates are not activated. The computational cost associated with this model is significantly higher, since the system contains 51 200 atoms and an extra variable, the cut-off distance $r_c$. The variation in the cut-off distance $r_c$ led also to a non-fixed turn-around time of each likelihood evaluation, varying between approximately 1 h until 8 h depending on the conditions and the cut-off simulated. The nature of the problem was thus ideally suited for the use of the dynamic scheduling, as well as the use of heterogeneous resources based on availability. The total campaign costed $\approx 100 000$ CPUhrs, not including the use of the GPUs, while the turn around time using 2520 cores was 1 and 1/2 days. These times illustrate the excessive computational demands and the drastic reductions that can be achieved using the proposed TMCMC parallel capabilities.

Since the radial distribution function indicates the three-dimensional layering around a particle, the values in close spatial proximity are expected to be correlated. The prediction errors are thus expected to have a certain degree of correlation along the direction $r$. An exponential correlation for the prediction error along the direction $r$ is postulated by selecting the covariance function of the Gaussian prediction error model to be of the form $\Sigma_{i,j} = \sigma_i^2 e^{-\text{dist}(i,j)/\lambda}$, where $\lambda$ is the correlation length, and $\text{dist}(i,j)$ is the distance between the $i$th
and $j$th experimental measurements of the $g(r)$. This prediction error model is denoted as PEM$_3$ and is a generalization of PEM$_1$ to include correlation along the distance $r$. The parameters $s$ and $\lambda$ are left as unknowns to be determined from the Bayesian estimation procedure, i.e., the parameter set is $\theta = (\sigma_{LJ}, \epsilon_{LJ}, r_c, s, \lambda)$.

Using the TMCMC algorithm for Bayesian estimation, results for the optimal model parameters are presented in Table VII, while results for the posterior distribution of the model parameters are shown in Fig. 7. Very small uncertainty (less than 0.2%) is estimated for the LJ parameters $\sigma_{LJ}$ and $\epsilon_{LJ}$. A relative large uncertainty, of the order of 9.5% is observed for the cut-off radius $r_c$. The most probable value of the cut-off radius is approximately 1.29 nm with 95% confidence interval contained within the [1.16, 1.42] nm, slightly higher than the usually preferred cut-off radius of 0.9–1.2 nm. The higher uncertainty in $r_c$ indicates that MD predictions of the distribution function $g(r)$ are less sensitive to the $r_c$ within the confidence interval. Relatively large uncertainty is estimated for the standard deviation $s$ and correlation length $\lambda$ of the prediction error parameters. From the projection in $(s, \lambda)$ space shown in Figure 7, these two parameters are shown to be highly correlated along certain elongated directions in the five-dimensional parameter space.

We remark that the estimated cut-off radii are smaller than the highest value of 1.5 nm allowed in the simulations. This is due to the model and computational errors arising from the selection of the force-field potential, the finite size of the simulation box, etc. Specifically for a force-field model that accurately represents reality, it is expected (neglecting computational errors) that the estimated, most probable value of the cutoff, would be infinite or the highest value allowed in the simulations. The present LJ model is an approximation to reality. Hence, the smaller than expected optimal cutoff

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>$\sigma_{LJ}$(nm)</th>
<th>$u_{\sigma_{LJ}}$</th>
<th>$\epsilon_{LJ}$(K)</th>
<th>$u_{\epsilon_{LJ}}$</th>
<th>$r_c$(nm)</th>
<th>$u_{r_c}$</th>
<th>$s$</th>
<th>$u_s$</th>
<th>$\lambda$(nm)</th>
<th>$u_\lambda$</th>
<th>Log evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEM$_3$</td>
<td>0.3388</td>
<td>0.04%</td>
<td>138.98</td>
<td>0.2%</td>
<td>1.29</td>
<td>9.54%</td>
<td>0.0185</td>
<td>10.03%</td>
<td>0.264</td>
<td>24.7%</td>
<td>959.42</td>
</tr>
</tbody>
</table>
TABLE VIII. Uncertainty propagation for the structural calibration based on the radial distribution function. Values reported are the most probable value plus/minus (±) a standard deviation of the PDF of the output quantities. The standard deviation is estimated using (14).

<table>
<thead>
<tr>
<th>$T$(K)</th>
<th>$P$(MPa)</th>
<th>$\rho$(kg/m$^3$)</th>
<th>$\alpha$(1/K)</th>
<th>$\beta$(1/kPa)</th>
<th>$\Delta H$(J/mol K)</th>
<th>$C_p$(J K mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt.</td>
<td>85.1</td>
<td>0.101</td>
<td>1409.0 ± 7</td>
<td>0.00434 ± 0.00013</td>
<td>(1.996 ± 0.06) × 10$^{-6}$</td>
<td>6493 ± 6.69</td>
</tr>
<tr>
<td>MD</td>
<td>85.1</td>
<td>0.101</td>
<td>1426.58 ± 18.2</td>
<td>0.0245 ± 0.006</td>
<td>(3.54 ± 0.96) × 10$^{-6}$</td>
<td>6568 ± 16.24</td>
</tr>
</tbody>
</table>

estimated by the Bayesian methodology, compensate for the LJ model inadequacy to represent accurately the realistic behavior of the analyzed system.

Figure 6 presents the radial distribution function $g(r)$ predicted by the MD simulations for the most probable model parameters, as well as its uncertainty (depicted as a blue fill) propagated through the model, along with the experimentally measured radial distribution function. The calibration procedure yields an accurate capture of the distribution function, with the experimental data to be contained within the uncertainty region predicted by the proposed Bayesian methodology. The unrealistic negative values of the distribution function predicted by the MD model for $r \in [0, 0.3]$ in Figure 6 are due to the fact that the standard deviation $s$ of the covariance of the model prediction error was assumed to be constant over the whole range of $r$ (from $r = 0.2$ to $r = 2.5$) in the calibration process. This choice is not appropriate for the region of $r$ for which the distribution function is very close to zero. However, such modeling does not affect the Bayesian calibration results due to the very small values of $g(r)$ in the range $r \in [0.2, 0.3]$. It only affects the predictions of the uncertainty of $g(r)$ at the range $r \in [0, 0.3]$ obtained from Eq. (14). Such predictions are influenced by the relative high standard deviation $s$ for the range $r \in [0, 0.3]$, giving unrealistic, albeit small, negative values of uncertainty for the distribution function.

We note that the errors between the experimental and the predicted $g(r)$, change sign at approximately the position of the atomic layers. This is an indication that the prediction errors are correlated along the distance $r$, justifying the use of a spatially correlated function for describing the prediction error along $r$. The correlation length $\lambda$ is quite uncertain as observed by the mean and standard deviation reported in Table VII. The mean value of $\lambda$ indicates a correlation distance very close to the characteristic length of the layering.

The uncertainties in the parameter estimates are propagated through the MD model to obtain the uncertainties and compare with experimental measurements for various properties of the argon that are not used for parameter calibration. The results are reported in Table VIII. The propagated uncertainties show that the model can capture relatively well the properties of liquid argon. The uncertainties in the MD predictions are significantly higher than the uncertainties in the experimental data for these properties. With the exemption of $\Delta H$, the experimental measurements (and their uncertainties) of all other properties reported in Table VIII are within the 95% confidence intervals predicted by the proposed methodology. However, the predictions based only on the optimal LJ model parameters and ignoring uncertainties, differ significantly from the experimental measurements. This emphasizes the importance of using the Bayesian framework to account for the uncertainties in the parameters and their propagation when in need of robust predictions consistent with experimental data. Instead, using only single optimal values, inadequate predictions are made that do not contain information about confidence intervals and in several cases deviate from experimental measurements.

VIII. CONCLUSIONS

A Bayesian framework for uncertainty quantification and propagation in MD simulations was presented and integrated with efficient parallel algorithms to reduce the large computational demands associated with the MD simulations. The present framework exploits available measurements to identify the uncertainties in the parameters of the potentials used for the intermolecular interactions, as well as to propagate these uncertainties for the prediction of confidence intervals of various thermodynamic quantities of interest. The TMCMC algorithm is chosen among a number of stochastic simulation algorithms as it is shown to be highly parallelized. Static and dynamic scheduling schemes efficiently distribute the computations involved in the TMCMC algorithm to available heterogeneous GPUs and multi-core CPUs. An adaptive surrogate technique based on the kriging algorithm was demonstrated to be sufficiently accurate and well suited to substitute for the computational demanding full MD runs within the TMCMC stochastic simulation algorithm. The effectiveness of the Bayesian framework, in terms of accuracy and computational efficiency, was demonstrated by applying it to a model of argon. Through the use of surrogates computational savings of several orders of magnitude are possible. Results obtained for argon using the available predictive analytical expressions and the full MD simulation protocols, elucidate the importance of incorporating uncertainties in predictions as well as the limitations of available molecular dynamics models, force fields and simulation protocols.

The proposed framework is general and can be useful in order to perform UQ+P for a wide range of MD and coarse grained simulations. Current work involves the extension of the present framework to other particle based methods for simulations of multiscale phenomena in areas ranging from nanotechnology to biology.

APPENDIX: SYSTEM AND SIMULATION PROTOCOL

The first system regarded calibration with dynamical properties data, that is self-diffusion coefficients of argon gas at various pressures and temperatures as seen in Table I. Each system contains 800 argon molecules. At each different parameter evaluation, each simulation was run for 15 ns with a time-step $\Delta t = 2$ fs. The initial 7.5 ns were discarded as they...
allowed the system to be equilibrated into the reference thermodynamic state. Each thermodynamic state had 25 different pre-equilibrated initial configurations with the Lennard-Jones parameters spanning a rectangular grid inside the initial parameter domain (5 different values for each parameter). Then, when every new sample was chosen, the initial configuration corresponding to the closest grid point was taken as an initial point with a different random seed for the velocity generation. This point was then further re-equilibrated using a random velocity rescaling algorithm and a Berendsen barostat as to provide rapid driving towards the reference ($T$, $P$). Subsequent refinement of the pressure and temperature ($T$, $P$) state variables during the sampling period was done using with a Nose-Hoover thermostat with a coupling constant of 0.1 ps and the Parrinello-Rahman barostat with a coupling constant of 0.1 ps. The cutoff was set at 1.2 nm with a shift function forcing it to decay smoothly to zero starting at $r_c = 1.0$ nm. Normal periodic boundary conditions were used throughout. Self-diffusion coefficients were measured via the mean squared displacement of the unwrapped trajectories and the Einstein-stokes equation. Some of the properties such as ($\alpha$, $\beta$) were extracted from MD simulations as fluctuations dependent properties, e.g., see Ref. 78, whereas $\eta$ was calculated using Einstein’s formula.

The second MD system used with calibration of the radial distribution function of liquid argon, with the data tabulated from experiments given in Ref. 71. The system contained 51 200 argon atoms pre-equilibrated at $T = 85.1$ K and $P = 1$ atm. Each parameter set evaluation consisted again of two parts. An initial 500 ps equilibration period, followed by a 500 ps sampling period every 500 fs of the radial distribution function. The algorithms used to equilibrate and control the system are the same as in system 1. The radial distribution function was computed using the GPU implementation of Visual Molecular Dynamics (VMD) with a radial binning of 0.001 nm. The exact evaluation of the computed $g(r)$ at the experimentally controlled points was done using cubic interpolation. A multi-threaded version of GROMACS 4.6 was used throughout.

62 J. Kestin, K. Knierim, E. A. Mason, B. Najafi, S. T. Ro, and M. Waldman,
(1975).
67 P. J. Moreno, P. P. Ho, and N. Vasconcelos, Advances in Neural Information
68 E. Lemmon, M. Huber, and M. McLinden, NIST Standard Reference
Database 23: Reference Fluid Thermodynamic and Transport Properties—
REFPROP (NIST, 2010).
2130 (1973).
74 D. van der Spoel and P. J. van Maaren, J. Chem. Theory Comput. 2, 1
(2005).
76 G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. 126, 014101
(2007).
(2011).