arXiv:1601.01044v1 [cond-mat.stat-mech] 6 Jan 2016

Sampling the isothermal-isobaric ensemble by Langevin dynamics

Xingyu Gao

Laboratory of Computational Physics,

Huayuan Road 6, Beijing 100088, P.R. China

Institute of Applied Physics and Computational Mathematics,

Fenghao East Road 2, Beijing 100094, P.R. China and

CAEP Software Center for High Performance Numerical Simulation,

Huayuan Road 6, Beijing 100088, P.R. China

Jun Fang and Han Wang*

Institute of Applied Physics and Computational Mathematics,

Fenghao East Road 2, Beijing 100094, P.R. China and

CAEP Software Center for High Performance Numerical Simulation,

Huayuan Road 6, Beijing 100088, P.R. China

Abstract

We present a new method of conducting molecular dynamics simulation in isothermal-isobaric ensemble based on Langevin equations of motion. The stochastic coupling to all particle and cell degrees of freedoms is introduced in a correct way, in the sense that the stationary configurational distribution is proved to be in consistent with that of the isothermal-isobaric ensemble. In order to apply the proposed method in computer simulations, a second order symmetric numerical integration scheme is developed by Trotter's splitting of the single-step propagator. Moreover, a practical guide of choosing working parameters is suggested for user specified thermo- and baro-coupling time-scales. The method and software implementation are carefully validated by a numerical example.

*Electronic address: wang˙han@iapcm.ac.cn

I. INTRODUCTION

Molecular dynamics (MD) simulation is a powerful tool for investigating a broad range of systems, from biological to materials sciences. In the equilibrium situation, it is of crucial importance to consider the ensemble that an MD simulation samples, because the quantities to observe is often calculated from the ensemble averages. Moreover, the equilibrium ensemble may serve as the initial condition for non-equilibrium MD simulations [1]. Most of early MD simulations solve the Hamiltonian dynamics, and sample the microcanonical ensemble. In practical applications, alternative ensembles like canonical or isothermal-isobaric (NPT) ensembles are usually more desirable, so various methods have been developed to sample the required ensemble by modifying the Hamiltonian dynamics.

One class of approaches to generate the desired ensemble is the extended phase space methods. For example, the Nosé-Hoover [2, 3], Nosé-Hoover chain [4] and stochastic Nosé-Hoover thermostats are proposed to generate the canonical ensemble; The Andersen [5], Parrinello-Rahman [6, 7] and Martyna-Tuckerman-Klein [8, 9] barostats are proposed to generate the NPT ensemble. These methods share the idea of extending the physical phase-space (positions, velocities of particles and the simulation cell) by extra variables that control the temperature and/or pressure of the system. The dynamics of the extended system is carefully designed to fulfill the condition that if the trajectory is ergodic, then the marginal stationary distribution in the physical phase-space is in consistent with that of the desired ensemble.

Langevin dynamics is an alternative method for generating the canonical ensemble. It has been proved that the ergodicity is guaranteed under mild restrictions [10, 11], therefore the convergence to the canonical distribution is ensured in the limit of infinitely long simulation time. Because of this advantage, Langevin dynamics has attracted increasing attention recently, and various integration schemes were developed in order to improve the accuracy of numerical simulations [12–17]. The first attempt of using the Langevin dynamics in generating the NPT ensemble was from Feller et.al. [18] and Kolb and Dünweg [19], who proposed the Langevin dynamics for both the particle degrees of freedom and the volume of the simulation cell (isotropic cell fluctuation). Quigley and Probert coupled the Parrinello-Rahamn dynamics with Langevin stochastic terms to extend the method to the fully flexible cell motions (anisotropic cell fluctuation) [20, 21]. However, it is not possible,

as the authors stated, to prove that the stationary distribution of the dynamics is subject to the NPT ensemble when the stochastic terms in cell motions do not vanish. Therefore, theoretically, the cell motions in this approach should be deterministic, and the convergence to the NPT distribution is not guaranteed, although the NPT distribution is one of the stationary distributions. In practice, the authors recommended to accept the cell stochastic motions at the cost of rigor in theory.

In this work, we propose a fully-flexible-cell NPT Langevin dynamics that allows explicit stochastic components in both the particle and cell motions, due to which the convergence of configurational distribution to the NPT ensemble is naturally ensured in the infinitely long time limit. We start by defining the Hamiltonian for an extended phase-space composed of scaled coordinates and simulation cell variables. Then the Langevin dynamics of this system can be directly written down, and the Boltzmann stationary distribution is obtained in the extended phase-space. In order to have a direct description of the particle motions in terms of physical coordinates, the scaled coordinates are transformed back, and the Langevin dynamics is reformulated accordingly by using Ito's formula. The configurational stationary distribution is then proved to be consistent with that of the NPT ensemble by transforming the Boltzmann distribution back to the physical coordinates in the same way. To develop the numerical scheme, we start by considering the Fokker-Planck equation that is equivalent to the Langevin equation, then the scheme is formulated by splitting the single-step propagator of the Fokker-Planck equation according to Trotter's theorem. By construction, the proposed scheme is of second order accuracy with respect to the time-step size.

Before moving to the main results of this work, it should be noted that it is in general difficult to check if the convergence to the NPT distribution is achieved in a finite simulation time. One might check the convergence of some properties of the system, for example the free energy profile along a certain reaction coordinate [22], but the choice of indicating properties depends on the nature of the system and what is wanted from the simulation, and they are usually not sufficient to prove the convergence of the distribution. Therefore, we do NOT intend to investigate the ergodicity or the speed of convergence to the desired ensemble distribution in numerical simulations. The significance of this work is to propose a new fully-flexible-cell NPT Langevin dynamics that takes the *theoretical* advantage of ergodicity, and can be used as an alternative to the existing NPT simulation methods.

This paper is organized as following: The development of NPT Langevin dynamics is

discussed in detail in Sec. II. The discretization of Langevin dynamics is provided in Sec. III. In Sec. IV, we validate the NPT Langevin dynamics by a solid argon system of triclinic region cell. This work is concluded in Sec. V.

II. THE LANGEVIN EQUATIONS OF MOTION

We denote the particle positions in the system by $r_1, \cdots, r_N$, where N is the number of particles in the system. The simulation cell matrix is denoted by $h = [h_1, h_2, h_3]$, where $h_\alpha, \alpha = 1, 2, 3$ are cell vectors. The scaled (direct) coordinate of a particle s_i is defined by $r_i = hs_i$. In order to generate the fully-flexible-cell NPT ensemble, all components of the cell vectors are allowed to fluctuate. We define the kinetic energy of the system by

$$K = \sum_i \frac{1}{2} m_i (h\dot{s}_i)^2 + \sum_{\alpha\beta} \frac{1}{2} M_{\alpha\beta} \dot{h}_{\alpha\beta}^2, \tag{1}$$

where $M_{\alpha\beta}$ is the fictitious mass corresponding to the motion of $h_{\alpha\beta}$ that is the β-th component of the α-th cell vector. The first term on the RHS of (1) is different from the *physical* kinetic energy of the system, which is $\sum \frac{1}{2} m_i [d(hs_i)/dt]^2$. The consequence of this difference will be discussed in detail later. It worth noting that the Parrinello-Rahamn barostat [6, 7] also uses Eq. (1) as the definition of kinetic energy. The Lagrangian of the system is defined as

$$\mathcal{L}(\{s_i\}, \{\dot{s}_i\}, h, \dot{h}) = K - (U + P \det(h)), \tag{2}$$

where $U = U(hs_1, \cdots, hs_N)$ is the potential energy of the system, and P is the target pressure. The generalized momenta corresponding to r_i and h are

$$\pi_i = m_i h^\top h \dot{s}_i, \quad p_{\alpha\beta}^h = M_{\alpha\beta} \dot{h}_{\alpha\beta}, \tag{3}$$

respectively. Therefore, the Hamiltonian of the system yields

$$\mathcal{H}(\{s_i\}, \{\pi_i\}, h, p_h) = \sum_{\alpha\beta} \frac{(p_{\alpha\beta}^h)^2}{2M_{\alpha\beta}} + \sum_i \frac{(h^{-\top}\pi_i)^2}{2m_i} + U + P \det(h). \tag{4}$$

The first term on the RHS of (4) is the kinetic energy of cell vectors, and the summation of the last three terms gives the instantaneous enthalpy. The Langevin dynamics is defined by

$$\dot{s}_i = \frac{\partial \mathcal{H}}{\partial \pi_i} \tag{5a}$$

$$\dot{\pi}_i = -\frac{\partial \mathcal{H}}{\partial s_i} - \Gamma_i \frac{\partial \mathcal{H}}{\partial \pi_i} + \Sigma_i \dot{W}_i \tag{5b}$$

$$\dot{h}_{\alpha\beta} = \frac{\partial \mathcal{H}}{\partial p^h_{\alpha\beta}} \tag{5c}$$

$$\dot{p}^h_{\alpha\beta} = -\frac{\partial \mathcal{H}}{\partial h_{\alpha\beta}} - \hat{\gamma}_{\alpha\beta} \frac{\partial \mathcal{H}}{\partial p^h_{\alpha\beta}} + \hat{\sigma}_{\alpha\beta} \dot{W}_{\alpha\beta}, \tag{5d}$$

where W_i, $W_{\alpha\beta}$ denote the standard Wiener processes, which are independent for different particle and cell degrees of freedom. The friction Γ_i and noise magnitude Σ_i are h dependent matrices defined by $\Gamma_i = \gamma m_i h^\top h$, $\Sigma_i = \sigma \sqrt{m_i}\, h^\top$, where $\sigma^2 = 2\gamma/\beta$, then the fluctuation-dissipation theorem $\Sigma_i \Sigma_i^\top = 2\Gamma_i/\beta$ holds for (5a)–(5b). We define $\hat{\gamma}_{\alpha\beta} = M_{\alpha\beta}\gamma_{\alpha\beta}$ and $\hat{\sigma}_{\alpha\beta} = \sqrt{M_{\alpha\beta}}\sigma_{\alpha\beta}$, where $\sigma^2_{\alpha\beta} = 2\gamma_{\alpha\beta}/\beta$, then the fluctuation-dissipation theorem $\hat{\sigma}^2_{\alpha\beta} = 2\hat{\gamma}_{\alpha\beta}/\beta$ holds for (5c)–(5d). It is well known that the stationary probability density of the Langevin dynamics (5) is

$$\rho_{\text{equi}} \propto \exp\Big[-\beta \mathcal{H}(\{s_i\}, \{\pi_i\}, h, p^h) \Big]. \tag{6}$$

It is more convenient to represent the Langevin dynamics (5) in physical coordinates, so we introduce the following transformation

$$r_i = h s_i, \quad p_i = h^{-\top} \pi_i. \tag{7}$$

By using Ito's formula, and writing down all partial derivatives explicitly, we reach

$$\dot{r}_i = \frac{1}{m_i} p_i + \dot{h} h^{-1} r_i \tag{8a}$$

$$\dot{p}_i = -\partial_i U - h^{-\top} \dot{h}^\top p_i - \gamma\, p_i + \sqrt{m_i}\, \sigma\, \dot{W}_i \tag{8b}$$

$$\dot{h}_{\alpha\beta} = \frac{1}{M_{\alpha\beta}} p^h_{\alpha\beta} \tag{8c}$$

$$\dot{p}^h_{\alpha\beta} = \det(h)[(P_{\text{ins}} - P)h^{-\top}]_{\alpha\beta} - \gamma_{\alpha\beta}\, p^h_{\alpha\beta} + \sqrt{M_{\alpha\beta}}\, \sigma_{\alpha\beta}\, \dot{W}_{\alpha\beta}, \tag{8d}$$

where P_{ins} is the instantaneous pressure tensor defined by

$$P_{\text{ins}} = \frac{1}{\det(h)} \sum_i \Big(\frac{1}{m_i} p_i \otimes p_i + F_i \otimes r_i \Big). \tag{9}$$

It can be easily shown that the Jacobian determinant of the transform (7) is 1, therefore, by integrating out the cell vector momenta, the equilibrium probability density generated by dynamics (8a)–(8d) is

$$\rho_{\text{equi}} \propto \exp\Big[-\beta\Big(\sum_{i=1}^{N}\frac{p_i^2}{2m_i}+U(\{r_i\})+P\det(h)\Big)\Big]. \tag{10}$$

Remark 1: The stationary distribution (10) obtained here is different from that can be found in some literature, e.g. Ref. [8], by a prefactor of $[\det(h)]^{d-1}$ (d being the dimension of the physical space), regardless of the normalization constant. This prefactor stems from the conservation of the system momentum when using the deterministic extended phase-space methods. For Langevin dynamics, the system momentum is neither conserved nor restrained, so the prefactor $[\det(h)]^{d-1}$ does not appear in (10). The prefactor contributes the free energy at the order of $\mathcal{O}(d/N)$, therefore, when the system contains a large number of particles, this difference is expected to be negligible.

Remark 2: It should be noted here that the particle momentum p_i is not exactly the physical momentum p_i^{phys}. As a matter of fact, we have the relation:

$$p_i^{\text{phys}} = p_i + m_i\dot{h}h^{-1}r_i. \tag{11}$$

This inconsistency is the direct consequence of the definition of the kinetic energy (1), in which the particle contribution is not the physical kinetic energy. There is no substantial difficulty in using the physical definition, however, the derivation of the Langevin dynamics would become much more complicated. In most applications, only the correctness of system configuration is of importance. The current definition guarantees the configurational distribution to be consistent with that of the NPT ensemble, so we consider it to be acceptable.

Remark 3: The choice of the parameters in the Langevin dynamics (8) has been extensively discussed in literature, e.g. [19, 20]. We pick up the friction coefficients based on a rule of thumb [20]: $\gamma = \gamma_{\alpha\beta} = \omega_T/2\pi = 1/\tau_T$, where ω_T denotes the frequency of the thermostat and $\tau_T = 2\pi/\omega_T$ is the time-scale of the thermostat. In order to provide the way of choosing the fictitious mass $M_{\alpha\beta}$, we firstly assume that the cell matrix h is diagonal (a cuboid cell). With a first order expansion of the pressure with respect to the cell fluctuation in h_{11}, h_{22} and h_{33}, we have

$$P - P^0 = -\frac{1}{\kappa h_{11}^0 h_{22}^0 h_{33}^0}(h_{11}h_{22}h_{33} - h_{11}^0 h_{22}^0 h_{33}^0), \tag{12}$$

where $\kappa = -\frac{1}{V}\frac{\partial V}{\partial P}$ is the compressibility. The superscript "0" denotes the equilibrium value of the corresponding variable. We further assume that the fluctuation from the equilibrium value is small, i.e. $|h_{\alpha\alpha} - h^0_{\alpha\alpha}|$, $\alpha = 1, 2, 3$ are small. Arranging the equation (12) in the component-wise way and preserving only the first order fluctuations on the RHS, we have

$$\frac{1}{3}\Big[(P_{11} - P^0_{11}) + (P_{22} - P^0_{22}) + (P_{33} - P^0_{33})\Big] \\ = -\frac{1}{\kappa h^0_{11} h^0_{22} h^0_{33}}\Big[(h_{11} - h^0_{11})h^0_{22}h^0_{33} + (h_{22} - h^0_{22})h^0_{11}h^0_{33} + (h_{33} - h^0_{33})h^0_{11}h^0_{22}\Big].$$

One possible solution to the equation is

$$P_{\alpha\alpha} - P^0_{\alpha\alpha} = -\frac{3}{\kappa h^0_{\alpha\alpha}}(h_{\alpha\alpha} - h^0_{\alpha\alpha}), \quad \alpha = 1, 2, 3. \tag{13}$$

By inserting (8c) and (13) into (8d), and discarding the friction and noise, and higher order terms, we have

$$M_{\alpha\alpha}\ddot{h}_{\alpha\alpha} = -\frac{3\det(h^0)}{\kappa(h^0_{\alpha\alpha})^2}(h_{\alpha\alpha} - h^0_{\alpha\alpha}),$$

which is the equation of motion of harmonic oscillator $h_{\alpha\alpha}$ with spring constant $k_{\alpha\alpha} = \frac{3\det(h^0)}{\kappa(h^0_{\alpha\alpha})^2}$ and equilibrium position $h^0_{\alpha\alpha}$. By using the relation $\omega_{\alpha\alpha} = \sqrt{k_{\alpha\alpha}/M_{\alpha\alpha}}$, where $\omega_{\alpha\alpha}$ is the barostat frequency, we derive the expression for the fictitious mass

$$M_{\alpha\alpha} = \frac{3\det(h^0)}{\kappa(h^0_{\alpha\alpha})^2}\Big(\frac{1}{\omega_{\alpha\alpha}}\Big)^2 = \frac{3\det(h^0)}{\kappa(h^0_{\alpha\alpha})^2}\Big(\frac{\tau_{\alpha\alpha}}{2\pi}\Big)^2, \tag{14}$$

where $\tau_{\alpha\alpha}$ is the time-scale of the barostat. For simplicity, we choose the off-diagonal value of fictitious mass by

$$M_{\alpha\beta} = \frac{3\det(h^0)}{\kappa(h^0_{\alpha\alpha})^2}\Big(\frac{\tau_{\alpha\beta}}{2\pi}\Big)^2. \tag{15}$$

In practice, it is often impossible to predict the equilibrium cell matrix h^0 before performing the simulation, however, if the initial condition does not deviate very far from equilibrium, we take the initial value of h as a reasonable guess for h^0. The compressibility κ can either take an experimental value, or be estimated from short testing simulations using the formula: $\kappa = \beta(\langle V^2\rangle - \langle V\rangle^2)/\langle V\rangle$.

It worth noting that regardless of the choice of the parameters, the Langevin dynamics (8) samples the configurational distribution of the NPT ensemble at infinitely long time limit. The difference in using different parameters lies in the sampling efficiency. If the time-scales of the thermo- and barostat were chosen too large, then the temperature and pressure

of the system would not be adjusted in a responsive way, and the sampling of the NPT ensemble would be too slow. On the other hand, if they were chosen too small, then an unnecessarily small time-step would be needed to keep the MD simulation stable. Therefore, it was suggested that the inverse time-scales be chosen just below the typical molecular frequency [19].

III. DISCRETIZE THE LANGEVIN DYNAMICS

The evolution of a system governed by the Langevin dynamics (8a) – (8d) is equivalently described by the following Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} = \mathcal{F}\rho, \tag{16}$$

where $\rho(t, \{r_i\}, \{p_i\}, h, p^h)$ is the time dependent probability density defined on the phase space. $\mathcal{F}$ is the infinitesimal generator, which can be factorized as

$$\mathcal{F} = \mathcal{F}_K + \mathcal{F}_U + \mathcal{F}_O + \mathcal{F}_K^h + \mathcal{F}_U^h + \mathcal{F}_O^h, \tag{17}$$

with each term defined by

$$\begin{aligned}
\mathcal{F}_K &= \sum_i \Big[\frac{p_i}{m_i} + \dot{h}h^{-1}r_i\Big] \cdot \frac{\partial}{\partial r_i} \\
\mathcal{F}_U &= \sum_i \Big[- \partial_i U - h^{-\top}\dot{h}^{\top}p_i\Big] \cdot \frac{\partial}{\partial p_i} \\
\mathcal{F}_O &= \sum_{i=1}^{N} \Big[3\gamma + \gamma p_i \frac{\partial}{\partial p_i} + \frac{m_i\sigma^2}{2}\frac{\partial^2}{\partial p_i^2} \Big] \\
\mathcal{F}_K^h &= \sum_{\alpha\beta} \frac{p^h_{\alpha\beta}}{M_{\alpha\beta}} \frac{\partial}{\partial h_{\alpha\beta}} \\
\mathcal{F}_U^h &= \sum_{\alpha\beta} \det(h) \Big[(P_{\text{ins}} - P)h^{-\top}\Big]_{\alpha\beta} \frac{\partial}{\partial p^h_{\alpha\beta}} \\
\mathcal{F}_O^h &= \sum_{\alpha\beta} \Big[\gamma_{\alpha\beta} + \gamma_{\alpha\beta} p^h_{\alpha\beta} \frac{\partial}{\partial p^h_{\alpha\beta}} + \frac{M_{\alpha\beta}\sigma^2_{\alpha\beta}}{2} \frac{\partial^2}{\partial (p^h_{\alpha\beta})^2} \Big].
\end{aligned}$$

Given this factorization, the single-step propagator $e^{\Delta t \mathcal{F}}$ (Δt being the time-step) can be split by the Trotter theorem:

$$e^{\Delta t \mathcal{F}} = e^{\frac{\Delta t}{2}\mathcal{F}_U^h} e^{\frac{\Delta t}{2}\mathcal{F}_U} e^{\frac{\Delta t}{2}\mathcal{F}_K^h} e^{\frac{\Delta t}{2}\mathcal{F}_K} e^{\Delta t \mathcal{F}_O^h} e^{\Delta t \mathcal{F}_O} e^{\frac{\Delta t}{2}\mathcal{F}_K} e^{\frac{\Delta t}{2}\mathcal{F}_K^h} e^{\frac{\Delta t}{2}\mathcal{F}_U} e^{\frac{\Delta t}{2}\mathcal{F}_U^h} + \mathcal{O}(\Delta t^3). \tag{18}$$

This style of splitting is actually the "BAOAB" scheme proposed by Ref. [15]. The authors argued that the BAOAB splitting is more accurate than other schemes in the sense of configurational sampling. The action of propagator $e^{\Delta t \mathcal{F}_K}$ and $e^{\Delta t \mathcal{F}_U}$ corresponds to evolve r_i and p_i by Δt under the ordinary differential equation $\dot{r}_i = \frac{1}{m_i} p_i + \dot{h} h^{-1} r_i$ and $\dot{p}_i = -\partial_i U - h^{-\top} \dot{h}^\top p_i$, respectively. Here we adopt the convention that the cell matrix h is an upper triangular matrix, then one has to solve, in general, $\dot{x} = b + Ax$ with A being an upper or lower triangular matrix. In the current work, the analytic solution of $\dot{x} = b + Ax$ in the upper and lower triangular cases are denoted by $x(t) = S_u(x(0), t, b, A)$ and $x(t) = S_l(x(0), t, b, A)$, respectively. The explicit forms of function S_u and S_l are provided in Appendix A. The action of propagator $e^{\Delta t \mathcal{F}_O}$ and $e^{\Delta t \mathcal{F}_O^h}$ corresponds to evolve variables p_i and $p^h_{\alpha\beta}$ by Δt under the Ornstein-Uhlenbeck process. In general, an Ornstein-Uhlenbeck process $dp = -\gamma p dt + \sigma \sqrt{m} dw_t$ can be explicitly solved by $p(t) = e^{-\gamma t} p(0) + \frac{\sigma}{\sqrt{2\gamma}} \sqrt{1 - e^{-2\gamma t}} \sqrt{m} R$, where R is a random number subject to the normal distribution with vanishing mean and unit variance, i.e. $\mathcal{N}(0, 1)$.

Discarding the higher order terms and applying from left to right the propagators on the RHS of (18) yields the following numerical scheme:

1: **while** MD continues **do**
2: $\quad p^h_{\alpha\beta} \leftarrow p^h_{\alpha\beta} + \Delta t/2 \, \det(h) [(P_{\text{ins}} - P) h^{-\top}]_{\alpha\beta}$
3: $\quad p_i \leftarrow S_l(p_i, \Delta t/2, -\partial_i U, -h^{-\top} \dot{h}^\top)$
4: $\quad h_{\alpha\beta} \leftarrow h_{\alpha\beta} + \Delta t/2 \, (p^h_{\alpha\beta} / M_{\alpha\beta})$
5: $\quad r_i \leftarrow S_u(r_i, \Delta t/2, p_i/m_i, \dot{h} h^{-1})$
6: $\quad p^h_{\alpha\beta} \leftarrow e^{-\gamma_{\alpha\beta} \Delta t} p^h_{\alpha\beta} + \sqrt{1 - e^{-2\gamma_{\alpha\beta} \Delta t}} \sqrt{M_{\alpha\beta}/\beta} \, R$
7: $\quad p_i \leftarrow e^{-\gamma \Delta t} p_i + \sqrt{1 - e^{-2\gamma \Delta t}} \sqrt{m_i/\beta} R$
8: $\quad r_i \leftarrow S_u(r_i, \Delta t/2, p_i/m_i, \dot{h} h^{-1})$
9: $\quad h_{\alpha\beta} \leftarrow h_{\alpha\beta} + \Delta t/2 \, (p^h_{\alpha\beta} / M_{\alpha\beta})$
10: $\quad$ Compute the force for each particle
11: $\quad p_i \leftarrow S_l(p_i, \Delta t/2, -\partial_i U, -h^{-\top} \dot{h}^\top)$
12: $\quad$ Compute the instantaneous pressure tensor
13: $\quad p^h_{\alpha\beta} \leftarrow p^h_{\alpha\beta} + \Delta t/2 \, \det(h) [(P_{\text{ins}} - P) h^{-\top}]_{\alpha\beta}$
14: **end while**

The operators $\mathcal{F}_O$ and $\mathcal{F}_O^h$ are mutable, therefore, lines 6 and 7 in the algorithm can be

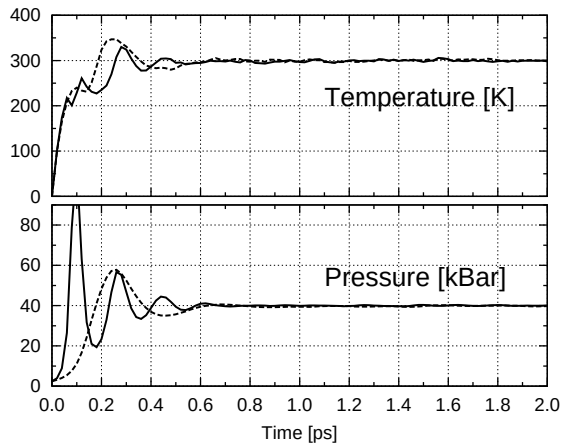

FIG. 1: The equilibriation of the Lennard-Jones argon system toward 300 K and 40 kBar. The simulation starts from a perfect FCC configuration at 0 K, and lasts for 100 ps. The plots present the equilibriation of temperature (upper plot) and pressure (lower plot) at the first 2 ps. The solid line uses a compressibility of 4.5×10^{-5} Bar^{-1}, and the dashed line uses a compressibility of 0.8×10^{-5} Bar^{-1}.

swapped.

IV. NUMERICAL RESULTS

The NPT Langevin method was implemented in the in-house molecular dynamics simulation package MOASP developed on infrastructure JASMIN [23]. In order to validate the theory and the implementation, we tested a solid argon system modeled by the Lennard-Jones interaction:

$$U(r) = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} + C, \tag{19}$$

where r is the distance between a pair of particles. $C_{12} = 2.71507 \times 10^{-7}$ kJ mol^{-1}nm^{-12} and $C_6 = 1.72685 \times 10^{-4}$ kJ mol^{-1}nm^{-6} are parameters taken from the CHARMM27 force field [24, 25]. C is a shifting constant that ensures the continuity of energy at the cut-off. The cut-off radius was chosen to be 0.9 nm in all simulations. The neighbor list was build for particles that are at most 1.1 nm apart, and was updated every 20 time-steps. The time-step of all simulations was chosen to be 1 fs.

An initial configuration of perfect FCC crystal was prepared by extending the Bravais

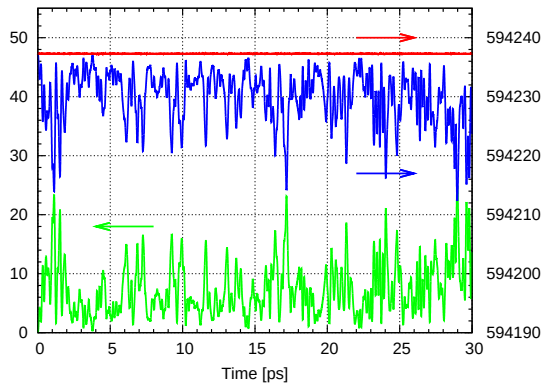

FIG. 2: The conservation of Hamiltonian when the friction and noise vanish. The time evolution of the cell kinetic energy (green), the instantaneous enthalpy (blue) and the Hamiltonian (red) are presented. The kinetic energy uses the left y-axis, while the instantaneous enthalpy and the Hamiltonian use the right y-axis, as the arrows in the Figure indicate. The unit of the energy is kJ/mol.

lattice cell of $|h_1| = |h_2| = |h_3| = 1.825$ nm and $\alpha = \beta = \gamma = 60°$ by $30 \times 20 \times 20$ times along three cell vectors, respectively. Therefore, the system contained 12,000 atoms in total. An 100 ps equilibriation simulation that used this configuration and zero initial velocities was conducted at 300 K and 40 kBar. The initial guess of the compressibility was 4.5×10^{-5} Bar^{-1} (which was actually a value taken from the liquid water under ambient condition). The time-scales of thermostat and barostat were set to 0.1 ps and 0.5 ps, respectively. The system was successfully equilibriated to the desired thermodynamic state within only 1 ps (the solid lines in Fig. 1), and the initial FCC solid structure was stable under this thermodynamic condition. The finial coordinates and velocities of atoms were recorded for productive simulations. The compressibility, 0.8×10^{-5} Bar^{-1}, was estimated from this simulation, and was used for all following simulations. Since the initial guess of the compressibility was much larger than 0.8×10^{-5} Bar^{-1}, the speed of equilibriation was actually faster than the user specified thermo- and barostat time-scales. We conducted the equilibriation again with the correct compressibility, and found that the speed of equilibriation was roughly the same as the specified thermo- and barostat time-scales (the dashed lines in Fig. 1).

An effective way to validate the correctness of the equations and software implementation is to check the conservation of Hamiltonian (4) when the friction and noise in (8) vanish.

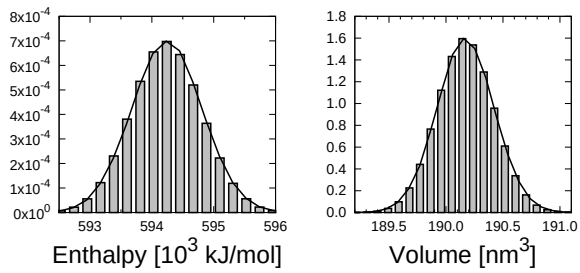

FIG. 3: The distribution of the instantaneous enthalpy (left) and cell volume (right) of the solid argon system. The gray bars present the probability densities calculated from the 10,000 ps Langevin NPT simulation (this work) at 300 K and 40 kBar. The solid lines present the same probability densities calculated from the reference simulation (see the text for more details).

We performed this simulation with the initial positions and velocities from the previous equilibriation, and plot the evolution of the cell kinetic energy, the instantaneous enthalpy and the Hamiltonian in Figure 2. A perfect conservation of the Hamiltonian is observed. It worth mentioning that the cell kinetic energy and the instantaneous enthalpy fluctuate at the magnitude of roughly ± 5 kJ/mol (see Fig. 2), and that the kinetic and potential energy of particles fluctuate at the magnitude of roughly ± 300 kJ/mol (not shown).

The productive NPT simulation lasted for 10,000 ps, and the data collection started from 100 ps. The instantaneous enthalpy and cell vectors were recorded every 0.02 ps. We plot the distribution of the instantaneous enthalpy and cell volume in Fig. 3. The distribution of the lengths of cell vectors and the angles between them are presented in Fig. 4. In the figures, the results of the NPT Langevin dynamics are compared and consistent with those of a reference simulation that uses velocity-rescaling thermostat [26] and Parrinello-Rahman barostat [6, 7]. The initial condition and the other parameters were set to be the same as the Langevin simulation. The reference simulation is considered to be reliable, because it was conducted by a well-tested MD simulation package Gromacs [27, 28] (version 4.6.5). The radial distribution functions (RDFs) calculated by Langevin and the reference simulation are shown to be overlapping in the upper plot of Fig. 5. From the lower plot of Fig. 5, we observe that the difference between them is dominated by the statistical uncertainty. This means that the solid structures sampled by Langevin dynamics reproduce those of the reference simulation.

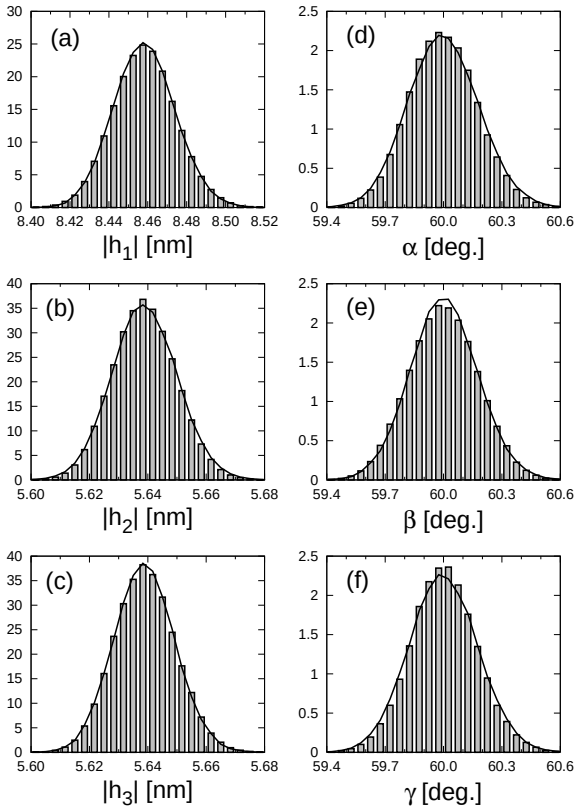

FIG. 4: The distribution of cell vectors. Plot (a) – (c) on the left column present the probability densities of lengths of three cell vectors, i.e. $|h_1|$, $|h_2|$ and $|h_3|$, respectively. Plot (d) – (f) on the right column present the probability densities of three angles α, β and γ between them, respectively. (α is the angle between h_2 and h_3. β is the angle between h_1 and h_3. γ is the angle between h_1 and h_2.) The simulated system is the same as Fig. 3.

V. CONCLUSIONS AND DISCUSSIONS

In this work, we proposed a new fully-flexible-cell Langevin dynamics in NPT ensemble. Our approach couples stochastic terms to both the particle and cell degrees of freedom, and is proved to correctly sample the configurational stationary distribution of the NPT ensemble. We noted that the choice of the working parameters (friction coefficients and the fictitious masses of cell vectors) does not affect the sampling of the NPT ensemble in the infinitely long time limit, however it determines the sampling efficiency and the size of the MD time-step. Therefore, we suggested a practical guide that automatically computes the fictitious mass and the friction coefficients by using the user specified thermo-/barostat time-scales, and the compressibility of the simulated material. In order to solve the Langevin equations

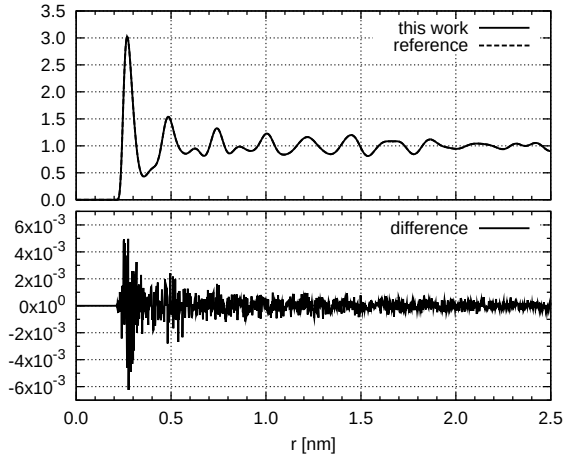

FIG. 5: The radial distribution functions calculated by the NPT Langevin (this work) and reference simulations. The two RDFs are plotted together in the upper plot; The difference between them is presented in the lower plot.

by computers, a discretization scheme was developed by using the Trotter splitting of the single-step propagator of the Fokker-Planck equation. This scheme is, by construction, of second order accuracy. A solid argon system modeled by the Lennard-Jones interaction was simulated to validate the proposed Langevin dynamics and the numerical scheme. The conservation of the Hamiltonian in the case of vanishing friction and noise was firstly check, and the correctness of the equations and the software implementation was confirmed. Then the equilibrium distributions of enthalpy, cell volume and vectors were calculated, and were found to be in good consistency with those calculated by a reference simulation (conducted by package Gromacs 4.6.5). The accuracy of the equilibrium structure of the system was verified by comparing the radial distribution function with the reference simulation.

Although the Langevin dynamics is an efficient sampling tool of the equilibrium ensembles, the computed dynamical properties like time correlation functions are usually wrong, because the stochastic terms in the momentum equations break the physical dynamics of the system that is described by the Newton's equations of motion. Therefore, in the cases where the dynamical properties are intended to be computed precisely, more carefully designed methods, e.g. local Langevin thermostat, should be used [29].

VI. ACKNOWLEDGMENT

The authors acknowledge the valuable discussions with Aiqing Zhang, Xu Liu and Jianjun Liu, and the technical supports from them. The authors gratefully acknowledge the financial support from National High Technology Research and Development Program of China under Grant 2015AA01A304. X.G. is supported by the National Science Foundation of China under Grants 91430218 and 61300012. H.W. is supported by the National Science Foundation of China under Grant 11501039.

Appendix A: Solve the equation $\dot{x} = b + Ax$

The ordinary differential equation $\dot{x} = b + Ax$ can be solved analytically. We start with the case that A is an upper triangular matrix:

$$\begin{bmatrix} \dot{x}_0 \\ \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ 0 & a_{11} & a_{12} \\ 0 & 0 & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}. \tag{A1}$$

The solution is given by:

$$\begin{aligned} x_0(t) = & x_0(0)e^{a_{00}t} + t\, b_0 F_1(0, a_{00}t) \\ & + t\, a_{01}x_1(0)F_1(a_{00}t, a_{11}t) + t^2\, a_{01}b_1F_2(a_{00}t, 0, a_{11}t) \\ & + t\, a_{02}x_2(0)F_1(a_{00}t, a_{22}t) + t^2\, a_{02}b_2F_2(a_{00}t, 0, a_{22}t) \\ & + t^2 a_{01}a_{12}x_2(0)F_2(a_{00}t, a_{11}t, a_{22}t) + t^3 a_{01}a_{12}b_2F_3(a_{00}t, a_{11}t, 0, a_{22}t) \end{aligned} \tag{A2}$$

$$\begin{aligned} x_1(t) = & x_1(0)e^{a_{11}t} + t\, b_1 F_1(0, a_{11}t) \\ & + t\, a_{12}x_2(0)F_1(a_{11}t, a_{22}t) + t^2\, a_{12}b_2F_2(a_{11}t, 0, a_{22}t) \end{aligned} \tag{A3}$$

$$x_2(t) = x_2(0)e^{a_{22}t} + t\, b_2 F_1(0, a_{22}t), \tag{A4}$$

where the function F_1, F_2 and F_3 are defined to be

$$F_1(A, B) = \frac{e^A - e^B}{A - B} \tag{A5}$$

$$F_2(A, B, C) = \frac{1}{B - C}(F_1(A, B) - F_1(A, C)) \tag{A6}$$

$$F_3(A, B, C, D) = \frac{1}{C - D}(F_2(A, B, C) - F_2(A, B, D)). \tag{A7}$$

It should be noticed that the definitions have singularity when any two of the A, B, C and D are equal, and we do not exclude these cases in real simulations. Noticing that we only need the solution of (A1) at small t, the singularity can be avoided by Taylor expansion of the exponential functions around 0. Following this idea, we reach the serial expansions of F_1, F_2 and F_3:

$$F_1(A,B) = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} \sum_{\substack{0\le\alpha,\beta\le k\\ \alpha+\beta=k}} A^\alpha B^\beta \tag{A8}$$

$$F_2(A,B,C) = \sum_{k=0}^{\infty} \frac{1}{(k+2)!} \sum_{\substack{0\le\alpha,\beta,\gamma\le k\\ \alpha+\beta+\gamma=k}} A^\alpha B^\beta \tag{A9}$$

$$F_3(A,B,C,D) = \sum_{k=0}^{\infty} \frac{1}{(k+3)!} \sum_{\substack{0\le\alpha,\beta,\gamma,\delta\le k\\ \alpha+\beta+\gamma+\delta=k}} A^\alpha B^\beta C^\gamma D^\delta. \tag{A10}$$

For the case that A is a lower triangular matrix:

$$\begin{bmatrix} \dot{x}_0 \\ \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} a_{00} & 0 & 0 \\ a_{10} & a_{11} & 0 \\ a_{20} & a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}, \tag{A11}$$

the solution can be written down in a similar way:

$$x_0(t) = x_0(0)e^{a_{00}t} + t\, b_0 F_1(0, a_{00}t) \tag{A12}$$

$$\begin{aligned} x_1(t) = {} & x_1(0)e^{a_{11}t} + t\, b_1 F_1(0, a_{11}t) \\ & + t\, a_{10}x_0(0)F_1(a_{11}t, a_{00}t) + t^2\, a_{10}b_0 F_2(a_{11}t, 0, a_{00}t) \end{aligned} \tag{A13}$$

$$\begin{aligned} x_2(t) = {} & x_2(0)e^{a_{22}t} + t\, b_2 F_1(0, a_{22}t) \\ & + t\, a_{21}x_1(0)F_1(a_{22}t, a_{11}t) + t^2\, a_{21}b_1 F_2(a_{22}t, 0, a_{11}t) \\ & + t\, a_{20}x_0(0)F_1(a_{22}t, a_{00}t) + t^2\, a_{20}b_0 F_2(a_{22}t, 0, a_{00}t) \\ & + t^2 a_{21}a_{10}x_0(0)F_2(a_{22}t, a_{11}t, a_{00}t) + t^3 a_{21}a_{10}b_0 F_3(a_{22}t, a_{11}t, 0, a_{00}t). \end{aligned} \tag{A14}$$

---

[1] Han Wang, Christof Schuütte, Giovanni Ciccotti, and Luigi Delle Site. Exploring the conformational dynamics of alanine dipeptide in solution subjected to an external electric field: A nonequilibrium molecular dynamics simulation. *Journal of Chemical Theory and Computation*, 10(4):1376–1386, 2014.

[2] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[3] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[4] Glenn J Martyna, Michael L Klein, and Mark Tuckerman. Nosé–hoover chains: the canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, 1992.

[5] H.C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72:2384, 1980.

[6] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52:7182, 1981.

[7] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, 45(14):1196–1199, 1980.

[8] G.J. Martyna, D.J. Tobias, and M.L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994.

[9] G.J. Martyna, M.E. Tuckerman, D.J. Tobias, and M.L. Klein. Explicit reversible integrators for extended systems dynamics. *Molecular Physics*, 87(5):1117–1157, 1996.

[10] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic Processes and Their Applications*, 101(2):185–232, 2002.

[11] Jonathan C Mattingly and Andrew M Stuart. Geometric ergodicity of some hypo-elliptic diffusions for particle motions. *Markov Process. Related Fields*, 8(2):199–214, 2002.

[12] G. Bussi and M. Parrinello. Accurate sampling using langevin dynamics. *Physical Review E*, 75(5):056707, 2007.

[13] S. Melchionna. Design of quasisymplectic propagators for langevin dynamics. *The Journal of Chemical Physics*, 127(4):44108–44108, 2007.

[14] N. Bou-Rabee and H. Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM Journal on Numerical Analysis*, 48(1):278–297, 2010.

[15] B. Leimkuhler and C. Matthews. Robust and efficient configurational molecular sampling via langevin dynamics. *The Journal of Chemical Physics*, 138:174102, 2013.

[16] Benedict Leimkuhler and Charles Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research eXpress*, 2013(1):34–56, 2013.

[17] Elias Alphonsus Jozef Franciscus Peters, Nicolae Goga, and Herman JC Berendsen. Stochastic dynamics with correct sampling for constrained systems. *Journal of Chemical Theory and Computation*, 10(10):4208–4220, 2014.

[18] S.E. Feller, Zhang Y., R.W. Pastor, and B.R. Brooks. Constant pressure molecular dynamics simulation: the langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995.

[19] A Kolb and B Dünweg. Optimized constant pressure stochastic dynamics. *The Journal of Chemical Physics*, 111(10):4453–4459, 1999.

[20] D Quigley and MIJ Probert. Langevin dynamics in constant pressure extended systems. *The Journal of Chemical Physics*, 120(24):11432–11441, 2004.

[21] D Quigley and MIJ Probert. Constant pressure langevin dynamics: theory and application. *Computer Physics Communications*, 169(1):322–325, 2005.

[22] Evan Kelly, Michael Seth, and Tom Ziegler. Calculation of free energy profiles for elementary bimolecular reactions by ab initio molecular dynamics: sampling methods and thermostat considerations. *The Journal of Physical Chemistry A*, 108(12):2167–2180, 2004.

[23] Zeyao Mo, Aiqing Zhang, Xiaolin Cao, Qingkai Liu, Xiaowen Xu, Hengbin An, Wenbing Pei, and Shaoping Zhu. Jasmin: a parallel software infrastructure for scientific computing. *Frontiers of Computer Science in China*, 4(4):480–488, 2010.

[24] Nicolas Foloppe and Alexander D MacKerell Jr. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*, 21(2):86–104, 2000.

[25] Alexander D Mackerell and Nilesh K Banavali. All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution. *Journal of Computational Chemistry*, 21(2):105–120, 2000.

[26] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126:014101, 2007.

[27] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput*, 4(3):435–447, 2008.

[28] 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, B. Hess, and E. Lindahl. Gromacs 4.5: a high-throughput and

highly parallel open source molecular simulation toolkit. *Bioinformatics*, page btt055, 2013.

[29] Han Wang and A. Agarwal. Adaptive resolution simulation in equilibrium and beyond. *The European Physical Journal Special Topics*, 224(12):2269–2287, 2015.