

# Further cautionary tales on thermostatting in Molecular Dynamics: energy equipartitioning and non-equilibrium processes in gas-phase simulations

Roope Halonen,<sup>\*,†</sup> Ivo Neefjes,<sup>‡</sup> and Bernhard Reischl<sup>‡</sup>

<sup>†</sup>*Center for Joint Quantum Studies and Department of Physics, School of Science, Tianjin  
University, 92 Weijin Road, Tianjin 300072, China*

<sup>‡</sup>*Institute for Atmospheric and Earth System Research / Physics, Faculty of Science,  
University of Helsinki, P.O. Box 64, FI-00014, Finland*

E-mail: roope@tju.edu.cn

## Abstract

Molecular dynamics (MD) simulations of gas-phase chemical reactions are typically carried out on a small number of molecules near thermal equilibrium by means of various thermostatting algorithms. Correct equipartitioning of kinetic energy among translations, rotations and vibrations of the simulated reactants is critical for many processes occurring in the gas phase. As thermalizing collisions are infrequent in gas-phase simulations, the thermostat has to efficiently reach equipartitioning in the system during equilibration and maintain it throughout the actual simulation. Furthermore, in non-equilibrium simulations where heat is released locally, the action of the thermostat should not lead to unphysical changes in the overall dynamics of the system. In this study, we explore issues related to both obtaining and maintaining thermal

equilibrium in MD simulations of an exemplary ion–molecule dimerization reaction. We first compare the efficiency of Nosé-Hoover, Canonical Sampling through Velocity Rescaling, and Langevin thermostats for equilibrating the system and find that of these three only the Langevin thermostat achieves equipartition in a reasonable simulation time. We also study the effect of unphysical removal of latent heat released during simulations involving multiple dimerization events, when global thermostating schemes are applied, which effectively cools down the reactants and leads to an overestimation of the dimerization rate. Our findings underscore the importance of thermostating for the proper thermal initialization of gas-phase systems and the consequences of global thermostating in non-equilibrium MD simulations.

## 1 Introduction

Atomistic simulations of processes like gas-phase chemical reactions are typically performed on a small number of reactants near thermal equilibrium. Molecular dynamics (MD) simulations in particular are often limited to small system sizes and short trajectories due to the computational cost, especially if the interactions are described quantum mechanically. In MD simulations, thermal equilibrium can be realized by employing various thermostating algorithms which are somewhat artificial but computationally highly effective. Some standard thermostats are, however, problematic as they do not produce a proper canonical ensemble and/or are prone to introduce spurious unphysical dynamics to the simulated system.<sup>1–15</sup> Perhaps the most famous of such thermostating problems is the so-called “flying ice cube” effect,<sup>1,2,15</sup> which arises from unphysical transfer of kinetic energy from fast to slow modes of motion, causing the system’s vibrational modes to freeze, while simultaneously resulting in enormous translational and rotational energies. These kinds of violations of the equipartition theorem, which states that the thermal energy should be equally distributed over the system’s degrees of freedom (dof), ultimately affect the structural, dynamic, and thermodynamic properties of the simulated system.<sup>2,12,15</sup> Since most atomistic simulations

rely on the use of available, well-tested simulation packages, understanding the implementation details of different thermostats in these codes is crucial, especially for systems with a small number of dof.<sup>14</sup> Despite a large body of studies and cautionary publications, pitfalls of thermostatting are still recurring issues in many fields of computational physics, chemistry, and molecular biology.<sup>12,14,15</sup>

In condensed matter systems, continuous energy exchange between molecules usually ensures equipartition of intramolecular dof. However, in the gas phase, this form of thermal equilibration or energy relaxation is not efficient due to infrequent collisions at low densities. At the same time, it is especially important to achieve equipartition when modeling reactions in gas-phase systems, as the different partitions of the kinetic energy (translational, rotational, and vibrational; hereafter referred to by the subscripts tr, rot and vib, respectively) each have their own specific contribution to the reaction rate. Unimolecular reactions (decay, evaporation, dissociation, decomposition, etc.) mostly depend on vibrational energy and the rate of energy redistribution within a single molecular compound or cluster.<sup>16,17</sup> For canonical bimolecular reactions (related to e.g., association, charge exchange, or chemical ionization), the reaction rate can be divided into a kinetic prefactor and an exponential thermodynamic part,<sup>18</sup>

$$k_{\text{bi-rxn}} = \beta(T) \exp\left(-\frac{E_{\text{a}}}{k_{\text{B}}T}\right), \quad (1)$$

where  $\beta(T)$  is the collision rate coefficient,  $E_{\text{a}}$  ( $\geq 0$ ) the reaction-specific activation energy barrier,  $k_{\text{B}}$  the Boltzmann constant and  $T$  the overall temperature. Ideally, the equipartition of energy over different modes of motion translates to  $T_{\text{tr}} = T_{\text{rot}} = T_{\text{vib}} = T$  at equilibrium. The kinematics of freely moving gas particles are largely dominated by the translational motion (for non-interacting hard spheres,  $\beta \propto T_{\text{tr}}^{1/2}$ ), whereas the post-collision activation depends mainly on the vibrational motion (leading to  $\ln k_{\text{bi-rxn}} \propto T_{\text{vib}}^{-1}$ ). The rotational motion usually contributes little to the activation (for large compounds,  $\text{dof}_{\text{vib}} \gg \text{dof}_{\text{rot}}$ ), although for systems with strong long-range interactions, rotational motion can affect the collision trajectories, e.g., through dipole orientations.<sup>19,20</sup> Given these distinct functional

dependencies of the bimolecular reaction rate on the kinetic energies of the different modes of motion, it is paramount to obtain the statistically correct energy distribution over these modes when setting up MD simulations. Moreover, it is important to maintain this distribution throughout the simulation if multiple reactions take place.

Direct reactive MD simulations have become common in many fields of physical chemistry. For example, a recent advance in computational chemistry are so-called *ab initio* nanoreactors<sup>21-23</sup> in which the reaction pathways and mechanisms are explored without any a priori heuristic rules by simulating a group of reactants. Also, in theoretical studies of atmospheric new particle formation, where formation rates have mainly been determined based on cluster free energies obtained from static configurational sampling,<sup>24</sup> *ab initio* studies probing *real time* cluster formation might become the new focus of interest,<sup>25</sup> as little is known about the dynamics of new-born clusters.<sup>26,27</sup> Such simulations require efficient relaxation of the system to the desired temperature and equipartition of energy prior to the simulation phase involving reactions, and proper temperature control throughout the production simulation.

Chemical reactions and molecular association and dissociation, as a rule, involve some excess energy or latent heat. Following the principle of energy conservation, the cohesive energy of a bond is absorbed in/released from the system when a bond is broken/formed. This excess energy is initially absorbed/released very locally, affecting the atomistic motion of a small subset of the system only. Most thermostats, however, act *globally* by coupling the entire system to a heat bath and are as such unable to distinguish between reactants and products. Consequently, the heat release is balanced by uniformly scaling the kinetic energies of all compounds subjected to the thermostat. While this issue is rarely considered in the field of MD simulations of chemical reactions, the problem related to global thermostatting has been previously acknowledged in studies of fullerene formation from carbon vapor condensation,<sup>7</sup> and silicon or metal nanoparticle synthesis.<sup>9,28</sup> The effect of excess energy release is analyzed in detail in the special case of homogeneous nucleation in the gas phase.<sup>3,11,29-31</sup> In these nucleation simulations the number of reacting monomers is often very

large and the amount of excess energy relatively low, and hence the problem is likely even more pronounced in small-scale reactive MD simulations where a more substantial excess energy is shared between a smaller number of reactants.

In this study, we explore issues related to both obtaining and maintaining thermal equilibrium when performing MD simulations of bimolecular reactions in the gas phase, close to the free molecular regime.<sup>32,33</sup> We focus on simulations of flexible molecular compounds with explicit internal dof, i.e., vibrational modes. First, to study issues related to establishing thermal equilibrium and equipartitioning, we tested the often-used Nosé-Hoover, Canonical Sampling through Velocity Rescaling, and Langevin thermostats, using a system of 30 dimethylammonium ions,  $(\text{CH}_3)_2\text{NH}_2^+$ . Then, to study the effect of thermostating during the non-equilibrium production stage, we compared results obtained from two distinct dimerization simulation setups. The first setup is a stochastic dimerization simulation of 15 dimethylamine,  $(\text{CH}_3)_2\text{NH}$ , molecules and 15  $(\text{CH}_3)_2\text{NH}_2^+$  ions in a constant volume and periodic boundary conditions. The second setup consists of trajectory simulations of a large number of independent  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization events over a range of relevant collision geometries and relative velocities.

As the aim of this study is primarily methodological, the potential problems related to thermostatted simulations are investigated using a computationally inexpensive classical force field approach, rather than a chemically more realistic reactive force field or a quantum mechanical description of the compounds and their interactions. This approach allows us to minimize the statistical uncertainties of our results, while the observations and conclusions remain equally valid for more complex reactive or ab initio simulations of gas-phase systems.

The remainder of this paper is organized as follows: In Section 2, we briefly discuss the force field model, implementation details, thermostat algorithms, and simulation schemes. In Section 3, we report and discuss the efficiency of the different thermostats for thermalizing and realizing equipartition of energy in small gas-phase systems. In Section 4, we study the effect of thermostating on a non-equilibrium process, by analyzing the results of the

dimerization simulations. In Section 5, we summarize the main issues and best practices related to thermostating in gas-phase MD simulations and conclude the paper.

## 2 Methods and Simulation Details

### 2.1 Simulated Compounds and Force Field

As test compounds, we studied the neutral dimethylamine molecule,  $(\text{CH}_3)_2\text{NH}$ , and the positively charged dimethylammonium ion,  $(\text{CH}_3)_2\text{NH}_2^+$ . To describe these compounds and their interactions, we employed a force field fitted according to the OPLS all-atom procedure.<sup>34</sup> In the OPLS force field, the intramolecular interactions consist of harmonic bond potentials between covalently bonded atoms, harmonic angle potentials between atoms separated by two covalent bonds, and dihedral angle potentials between atoms separated by three covalent bonds,

$$U_{\text{intra}}^{\text{OPLS}} = \sum_{i=1}^{N_{\text{bonds}}} \frac{k_i^{\text{b}}}{2} (r_i - r_i^0)^2 + \sum_{j=1}^{N_{\text{angles}}} \frac{k_j^{\theta}}{2} (\theta_j - \theta_j^0)^2 + \sum_{k=1}^{N_{\text{dihedrals}}} \sum_{n=1}^4 \frac{V_n}{2} [1 + \cos(n\phi^k - \phi_n^k)], \quad (2)$$

where  $k_i^{\text{b}}$ ,  $r_i$ , and  $r_i^0$  are the force constant, instantaneous, and equilibrium length of bond  $i$ ,  $k_j^{\theta}$ ,  $\theta_j$ , and  $\theta_j^0$  the force constant, instantaneous, and equilibrium value of angle  $j$ , and  $V_n$ ,  $\phi_n^k$ , and  $\phi^k$  the Fourier coefficients, phase angles, and instantaneous value of the dihedral angle  $k$ .

The intermolecular interactions, as well as intramolecular interactions between atoms separated by more than three covalent bonds, are described by Lennard-Jones potentials between atoms  $i$  and  $j$  separated by a distance  $r_{ij}$ , with distance and energy parameters  $\sigma_{ij}$  and  $\epsilon_{ij}$ , and Coulomb interactions between the atoms' partial charges  $q_i$  and  $q_j$ ,

$$U_{\text{inter}} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (3)$$

where  $\epsilon_0$  is the vacuum permittivity.

The OPLS force field parameters used in this study were obtained from Loukonen et al.<sup>35</sup>. We note that in the original OPLS force field, Lennard-Jones and Coulomb interactions between atoms separated by three covalent bonds (“1-4 interactions”) are scaled by a factor 0.5. Loukonen et al.<sup>35</sup> set this scaling factor to zero when fitting the force field parameters. For consistency, we have also set these interactions to zero in our simulations.

## 2.2 Implementation Details

All MD simulations were carried out with the LAMMPS code<sup>36</sup>, using a Velocity-Verlet integrator with a time step  $\delta t = 1$  fs. Coulomb and Lennard-Jones potentials in the OPLS force field were cut off at 50 Å, or 150 Å, for gas-phase simulations, and collision trajectory simulations, respectively. We tested that, for the gas-phase simulations with periodic boundary conditions, truncating the Coulomb potential does not affect system properties compared to computing the long-range electrostatics using a particle-particle-particle-mesh (PPPM) solver.

In MD simulations, the temperature of the system is defined by the kinetic energy of the atoms. To determine the different partitions of the kinetic energy and temperature of specific compounds, the built-in functionalities of the LAMMPS simulation package<sup>37</sup> were used. In LAMMPS, up to 30 individual groups of atoms can be defined. Partly due to this restriction, the simulations here are limited to systems with 30 molecules or ions. For each group, the command `compute ID group-ID temp` can be used to calculate the group’s total instantaneous temperature ( $T_{\text{temp}}$ ), and commands `compute ID group-ID temp/com` and `compute ID group-ID temp/rotate` to obtain the temperature after subtracting the center-of-mass (com) motion ( $T_{\text{com}}$ ) and after subtracting both the com and rotational motion ( $T_{\text{rotate}}$ ), respectively. Please consult the online documentation for LAMMPS<sup>37</sup> for more details about the commands and the particular use of temperature (and dof) in the code.

Each of these “temperatures” are related to the respective kinetic energy through the relation

$$K_x = \frac{3n}{2}k_B T_x, \tag{4}$$

where  $n$  is the number of atoms in the group. Thus, the actual instantaneous temperatures corresponding to the translational, rotational and vibrational motion of a non-linear compound can be calculated according to the number of respective dof as

$$T_{\text{tr}} = n(T_{\text{temp}} - T_{\text{com}}), \tag{5}$$

$$T_{\text{rot}} = n(T_{\text{com}} - T_{\text{rotate}}), \tag{6}$$

$$T_{\text{vib}} = \frac{n}{n-2}T_{\text{rotate}}. \tag{7}$$

Note that in LAMMPS three dof are subtracted by default when computing temperatures, to account for the com motion of the entire system. When applying eqs 5–7, this default should be disregarded for each `compute` command (by using command `compute_modify compute-ID extra/dof 0`). Otherwise, for small groups in particular, the values of the different energy components will be severely underestimated.

### 2.3 Thermostat Algorithms

In our simulations, we focus on the widely used Nosé-Hoover (NH), Canonical Sampling through Velocity Rescaling (CSV), and Langevin thermostats, which are all able to generate correct canonical kinetic energy distributions unlike some other thermostats (e.g., basic Velocity Rescaling or the Berendsen thermostat<sup>38,39</sup>). The strength of the coupling of the system temperature to the heat bath temperature is conveniently defined by a coupling time constant  $\tau$ , which can be translated to intrinsic parameter values for different thermostating schemes. As a very general rule,  $\tau > 20\delta t$ .<sup>40</sup> In the following, we briefly introduce the different thermostating algorithms, for more details we refer the reader to the original

articles<sup>41–44</sup> and other useful sources.<sup>12,37,39,40,45</sup>

### 2.3.1 Nosé-Hoover

The Nosé-Hoover thermostat<sup>41,42</sup> (NH) is one of the most widely used thermostats (e.g., it is the default thermostat in LAMMPS). In this extended ensemble approach, the heat bath is reduced to a fictitious dof (or “virtual particle”) which interacts with the actual system to achieve the correct kinetic energy distribution. The time constant determines the temperature-dependent friction parameter of the extra dof by changing the mass parameter of the the virtual particle. The Nosé-Hoover thermostat is deterministic and global. By chaining NH thermostats, some ergodicity issues related to small systems can be avoided.<sup>46</sup> Here, we have used a NH thermostat with three chains.

### 2.3.2 Canonical Sampling Through Velocity Rescaling

The Canonical Sampling through Velocity Rescaling (CSVR) or Bussi-Donadio-Parrinello thermostat<sup>43</sup> *stochastically* alters the momenta of the atoms globally according to the canonical energy distribution at the end of each simulation time step. Thus, unlike other rescaling thermostats (namely the Berendsen or the Velocity Rescaling), CSVR adequately samples the canonical ensemble as natural temperature fluctuations are allowed. The time constant scales the temperature relaxation, analogously to the Berendsen thermostat.<sup>38</sup>

### 2.3.3 Langevin

Another widely used stochastic thermostat is the Langevin thermostat.<sup>44,47</sup> The simulated system is made to obey Langevin dynamics by introducing a virtual solvent around the atoms causing friction through *localized* random collisions. The system undergoes semi-Brownian motion to a degree set by the coupling constant  $\tau$ . This shift towards Brownian dynamics by the Langevin thermostat results in notable dampening of dynamical processes which may not reflect the desired environment and transport properties of the studied system. The

systems using global NH and CSVN thermostats produce more realistic gas-phase dynamics regardless of applied  $\tau$ . A detailed study on the subject is provided by Basconi and Shirts<sup>8</sup>.

## 2.4 Simulation Schemes

### 2.4.1 Equilibration of a Gas in the Canonical Ensemble

We have carried out a set of equilibration MD simulations to test the dynamics of kinetic energy redistribution with the aforementioned thermostats. In these simulations, a system consisting of 30  $(\text{CH}_3)_2\text{NH}_2^+$  ions is equilibrated to a target temperature of 300 K. The reason for simulating a system consisting only of cations is purely practical: the repulsive forces prevent clustering which would affect the analysis of kinetic energy partitioning as rotational and translational dof would be substituted for vibrational modes in the cluster. Since no bond formation takes place, the simulated system will stay in a gaseous state and should achieve equilibrium in a finite time.

The major difference to the setup used in the recent study on thermostats by Braun et al.<sup>12</sup> is that instead of first equilibrating each system with a Langevin thermostat, here the systems are thermostatted with one particular thermostat, with initial random velocities drawn from a Boltzmann distribution corresponding to the target temperature. As these atomic velocities are arbitrarily assigned to the entire system without any consideration for molecular structure, or equipartition within the different modes of motion, the initial states of individual compounds can be far from equilibrium. In this standard initialization procedure, the energy re-partitioning via subsequent thermostating is considered to be swift,<sup>39</sup> but as will be seen, this assumption may not hold for flexible compounds in the gas phase, where energy transfer via collisions is relatively rare. More details of the equilibration simulation setup is given in Section 3.

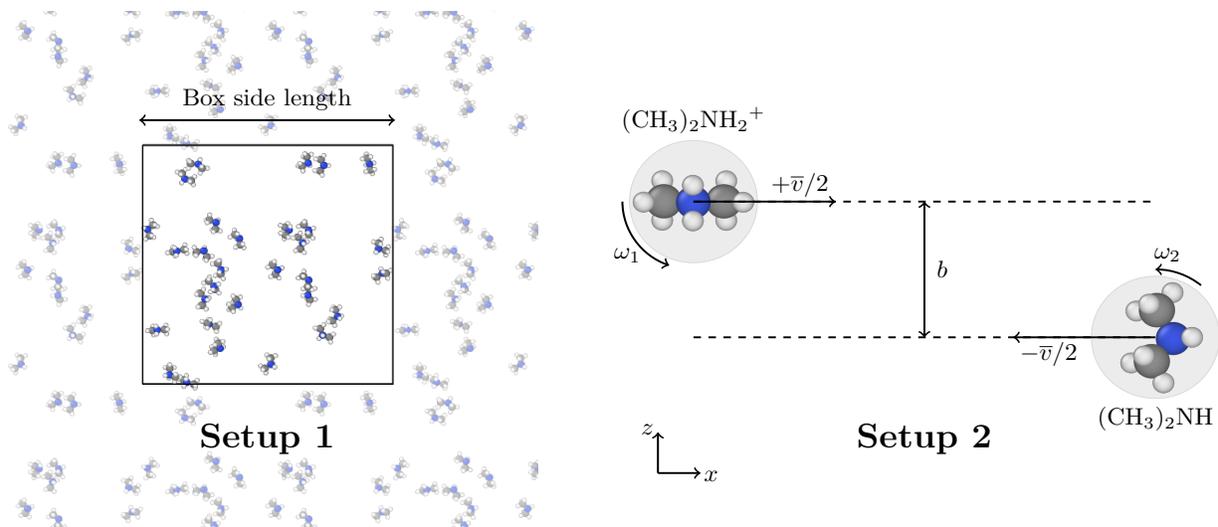


Figure 1: Illustration of the two simulation setups for the dimerization of  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$  (not to scale). Left: In Setup 1, we study the time evolution of 15  $(\text{CH}_3)_2\text{NH}$  and 15  $(\text{CH}_3)_2\text{NH}_2^+$  originally in thermal equilibrium, in a cubic simulation box with 3D periodic boundary conditions (periodic images of the system are shown as translucent atoms). Right: In Setup 2, we sample the collision statistics of two individual reactants as a function of their impact parameter  $b$  along the  $z$ -axis and their relative velocity  $v$  along the  $x$ -axis, for different randomized starting orientations and angular velocities  $\omega_1$  and  $\omega_2$ . Reactant atoms are color coded as follows: carbon is gray, nitrogen is blue, hydrogen is white.

## 2.4.2 Dimerization Simulations

To investigate the problems of maintaining thermal equilibrium when simulating the aggregation of molecules and ions, we compare two distinct simulation setups for the dimerization of  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$ . The key features of the setups are illustrated in Figure 1.

**Setup 1:** A straightforward approach to study reaction pathways, rate parameters and other mechanistic details of physico-chemical processes is to simulate a freely evolving continuous system consisting of multiple reactants near thermal equilibrium. In these simulations, the observed reactions occur stochastically and thus, to acquire statistically significant results, one can either perform a substantial number of individual simulations of small systems (where only a few events occur), or one simulation of a very large system containing numerous reactants (leading to many events). To emulate a typical simulation setup for systems with complex interactions, the simulations presented here are carried out for rather

small systems containing 15 molecules and 15 ions in a simulation box with a constant cubic volume of  $V = (300 \text{ \AA})^3$ , using periodic boundary conditions. The system is first shortly equilibrated using the best performing thermostat according to the equipartition test. After equilibration, the system is allowed to evolve in either the NVE or NVT ensemble, using a NH thermostat. The total linear momentum is conserved and *three* dof are hence disregarded during thermostatting (i.e., the default setting in LAMMPS is used). This setup allows us to study both equilibrium properties and non-equilibrium processes.

**Setup 2:** In addition, we carried out more controlled *trajectory* simulations for isolated systems of  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$  moving towards each other with a certain initial relative velocity  $v$  and impact parameter  $b$  (the right-hand side of Figure 1). In these simulations, the two compounds are equilibrated together beyond the force field cut-off of  $150 \text{ \AA}$  using the best performing thermostat in our equipartition test. After each time step, the respective com motion of each compound is removed keeping them both at rest during the equilibration and the loss of translational energy is *not* compensated by rescaling other energy components. Thereby the correct equipartition of rotational and vibrational modes was reached over the sampled configurations. After equilibration, the two compounds are set on a trajectory towards each other in the NVE ensemble. Numerous trajectories are simulated for various  $(v,b)$  pairs to obtain sufficient statistics allowing accurate calculations of collision cross sections and dimerization rate coefficients, as well as detailed geometric and dynamic analyses of the collision trajectories.

Further details of both dimerization simulation setups are provided in Sections 4.2 and 4.3, respectively.

### 3 Equipartition in Thermostatted Flexible Molecular Compounds

Systems consisting of  $N = 30$   $(\text{CH}_3)_2\text{NH}_2^+$  ions were simulated for several nanoseconds coupled to Langevin, Canonical Sampling through Velocity Rescaling (CSVR), or Nosé-Hoover (NH) thermostats, set to a target temperature  $T_0 = 300$  K, using three different time constants,  $\tau = 0.1, 1.0,$  or  $10.0$  ps. Each simulation was started from an identical initial configuration of atomic positions and velocities.

In each system, we observed that the overall kinetic temperature  $T$  fluctuates around 300 K with a standard deviation (std) of  $\sim 13.5$  K, which matches the expected value,<sup>48</sup>  $\text{std} = \sqrt{2/\text{dof}} T_0 = \sqrt{2/(3nN - 3)} T_0$ . However, maintaining the correct distribution of total kinetic energy alone does not guarantee correct equipartition into the different modes of motion through the action of the thermostats. The respective temperatures of different modes computed according to eqs 5–7 are shown in Figure 2. For these simulated systems of classical, flexible compounds to satisfy the equipartition theorem, it should hold that  $T_{\text{tr}} = T_{\text{rot}} = T_{\text{vib}} = T_0$ . As shown in Figure 2, the Langevin thermostat clearly performs best at bringing the system to equilibrium and the correct equipartitioning of energies. Even for the weakest coupling,  $\tau = 10.0$  ps, the vibrational temperature reaches 300 K in less than 100 ps. The temperature fluctuations are again close to the theoretical values for the canonical ensemble: for both translational and rotational temperature,  $\text{std} = 44.7 \pm 0.1$  K  $\equiv \sqrt{(2/3N)} T_0$ , and for vibrational temperature,  $\text{std} = 14.9 \pm 0.1$  K  $\equiv \sqrt{2/(N(3n - 6))} T_0$ . Note that apparent increase of the fluctuations with time in Figure 2 is only due to the logarithmic time axis.

When using either the CSVR or NH thermostat, the system remains out of thermal equilibrium for a significant amount of time. First the kinetic energies are redistributed and a brief quasi-steady state is achieved where the translational and rotational temperatures are above and the vibrational temperature is below the target temperature  $T_0$ . Even for the

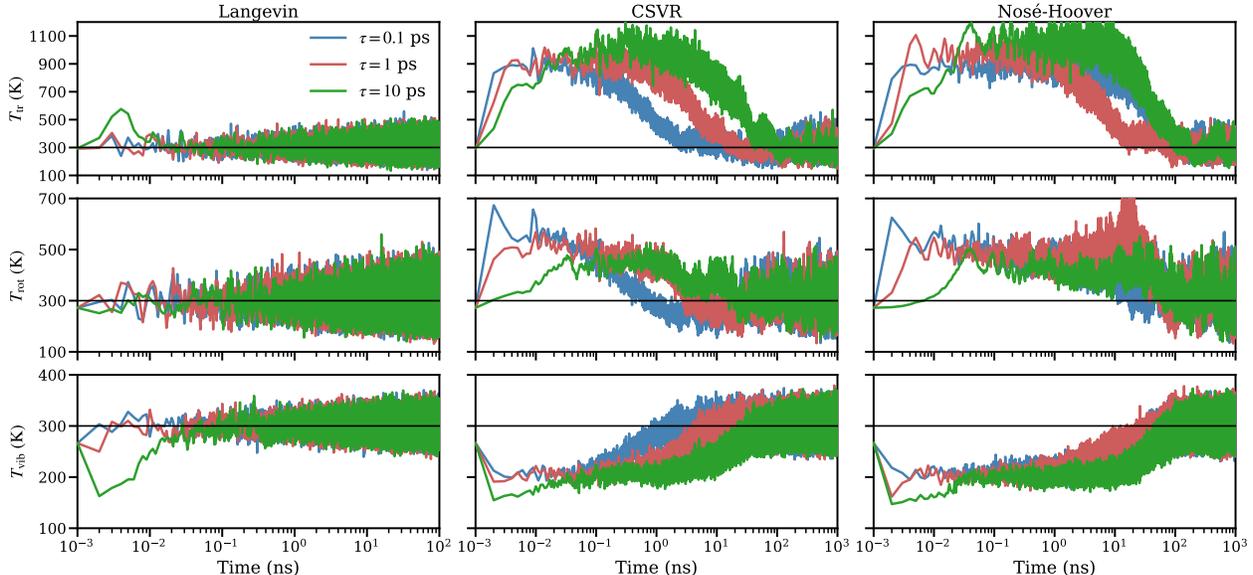


Figure 2: Time evolution of different temperature partitions of  $(\text{CH}_3)_2\text{NH}_2^+$  ions during equilibration using different thermostats set to a target temperature  $T_0 = 300$  K. From left to right: Langevin, Canonical Sampling through Velocity Rescaling (CSVr) and Nosé-Hoover thermostats. From top to bottom: translational, rotational and vibrational temperatures. Thermostat time constants  $\tau = 0.1, 1,$  and  $10$  ps are represented by blue, red, and green lines, respectively.

smallest value of  $\tau$  (i.e., strongest coupling), achieving equipartitioning with CSVr takes over one nanosecond. For the most commonly used canonical thermostat, i.e., NH, the equilibration is a very slow process and does not significantly depend on  $\tau$ . Almost one microsecond of thermostating with NH is required to achieve equipartition of the different modes of kinetic motion for the studied system. For rather trivial gas-phase MD simulations, these are very long time scales to properly equilibrate the systems.

We further verified the general independence of these results from the starting configuration by studying 100 individual simulations for each thermostat with  $\tau$  set to 1.0 ps, starting from different initial atomic positions and velocities. The obtained average temperatures, sampled during the quasi-steady state reached between 100 and 200 ps are presented in Figure 3. Similarly as for the long single runs presented in Figure 2, the CSVr and NH thermostats systematically assign too much energy to the translational and rotational modes at the expense of the vibrational modes.

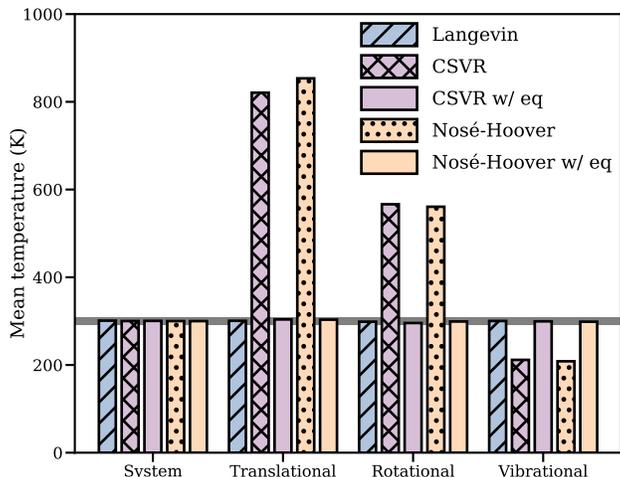


Figure 3: Overall system temperature and the different temperature partitions calculated averaged between 100 and 200 ps for 100 individual simulations of 30  $(\text{CH}_3)_2\text{NH}_2^+$  ions, using different thermostats. The unhatched bars correspond to simulations carried out with the Canonical Sampling through Velocity Rescaling (CSV) and Nosé-Hoover thermostats, after an initial equilibration of 100 ps using a Langevin thermostat. The target temperature,  $T_0 = 300$  K, is indicated by the horizontal gray line.

The failure of CSV and NH originates in their global working principle: after every time step the system’s momenta undergo a uniform transformation and thus no net energy exchange between translational/rotational and vibrational degrees of freedom occurs. This exchange of energy requires physical interactions, i.e., collisions, between the compounds or an inert carrier gas. Indeed, we verified that for a condensed matter system of 216  $(\text{CH}_3)_2\text{NH}$  molecules at  $T = 250$  K and liquid density  $\rho = 676 \text{ kg m}^{-3}$  ( $V = (28.8 \text{ \AA})^3$ ), equilibration and proper equipartition within a few picoseconds could be observed using Langevin, CSV, or NH thermostats, for all coupling time constants  $\tau$  (here, long-range electrostatics were evaluated using the PPPM method, otherwise the simulation protocol was identical to the gas-phase simulations). However, achieving equilibration through a large number of consecutive collisions is very inefficient when studying gas-phase systems near the free molecular regime. Assuming that the ions considered here act as hard spheres with cross sections of approximately  $100 \text{ \AA}^2$ , two ions are expected to interact roughly every nanosecond at the simulated conditions. This is comparable to the time scales of equilibration seen in Figure 2 for CSV and NH thermostats. On the other hand, using a

Langevin thermostat, the momenta are not conserved and the energy exchange is facilitated by the fictitious solvent.

In practice, the issues of NH and CSVN can easily be overcome by first driving the system towards equilibrium with a short run using the efficient Langevin thermostat, before starting the production run with a global thermostat, as required by the dynamics of the system. The unhatched bars in Figure 3 show averages of the respective temperature partitions for the CSVN and NH thermostat after an initial equilibration of 100 ps was performed with a Langevin thermostat. This indeed shows that both CSVN and NH are able to preserve equipartition, once it has been achieved. A more general demonstration of consecutive use of different thermostats is given by Braun et al.<sup>12</sup>

## 4 Effect of Thermostatting and Equilibration on Dimerization Rates

In Sections 4.2 and 4.3, we concentrate on determining the rate of dimerization of  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$  with two standard molecular dynamics approaches: stochastic dimerization simulations in a finite volume of gas (Setup 1) and individual collision trajectory simulations (Setup 2). The presented results and observations will not only highlight the importance of proper equilibration of the simulated systems but also the underlying issues related to thermostatting in simulations of non-equilibrium processes.

### 4.1 Definition of a Dimerization Event

First, the analysis of dimerization simulations requires a rigorous definition for the dimerization event. Due to the large amount of data generated by the presented calculations, the criterion for dimerization should be based on a minimum amount of information while being unambiguous so that the monitoring of the system can be automatized. The following rules for an actual dimerization event are used in both simulation approaches.

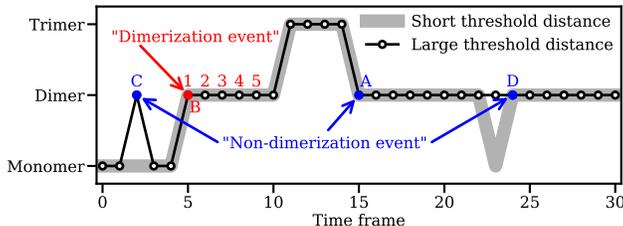


Figure 4: Schematic illustration of the dimerization detection procedure used to analyze the simulated  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  systems. The evolution of an individual “cluster” is monitored in every time frame (recorded each picosecond), based on the com distances. Time frames at which the cluster becomes a dimer (marked by full circles) correspond to dimerization events only if specific conditions are fulfilled: the cluster has to emerge from monomers (to prevent false detections from dissociation processes, shown as Non-event A) and the cluster has to be identified as a dimer for at least 5 subsequent time frames (Event B and Non-event C). Having a too short threshold distance might also result in misidentified evaporation events (Non-event D).

Generally, in our analysis a dimer is identified when the center-of-mass (com) distance between  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$  is shorter than some threshold value, and no other molecules or ions are within this range. More specifically, a dimer has to originate from two free monomers (molecule and ion), i.e., possible fragmentations from larger clusters into dimers are disregarded. As the fluctuations of a formed dimer around the equilibrium com distance of the two monomers (about  $3.6 \text{ \AA}$ ) can be rather large, especially immediately after collision, the threshold distance has to be large enough to prevent overcounting of dimerization events due to recombination. Here, we have used a threshold distance of  $10 \text{ \AA}$ , which is roughly equal to the observed maximum impact parameter in the trajectory simulations at the high velocity limit (see Section 4.3). Such a large threshold value, on the other hand, can lead to mistaking a “fly-by” for a dimerization event. To prevent this, we have introduced an additional criterion that the com distance between the monomers has to be below the threshold for at least 5 subsequent time frames recorded every 1 ps. The complete dimerization detection procedure is illustrated in Figure 4 for an individual “cluster”, fluctuating between monomer, dimer, and trimer, during a fictitious simulation.

## 4.2 Stochastic Dimerization Simulations

We simulated multiple systems of 15  $(\text{CH}_3)_2\text{NH}$  and 15  $(\text{CH}_3)_2\text{NH}_2^+$  in a constant volume of  $V = (300 \text{ \AA})^3$  at target temperatures  $T_0 = 200$  and 300 K. In these freely evolving simulations, dimerization can occur stochastically. Based on the discussion in Section 3, we have guaranteed the correct initial kinetic energy distribution in the system by applying a Langevin thermostat ( $\tau = 0.1$  ps) for 100 ps before starting the actual production simulation. Note that here we have refrained from studying or demonstrating the direct dynamical effect of improper equilibration while using CSVN or NH thermostats mainly because the effect depends highly on the selected length of the equilibration period and the studied system in general. The overall effect of improper initialization/equilibration, however, will be revealed in Section 4.4 by further analyzing the role of different modes of motion in the dimerization of  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$ .

After the short equilibration with the Langevin thermostat, the system is simulated for another 5 ns either without a thermostat (NVE) or with a thermostat applied to all atoms (NVT). Despite the demonstrated efficiency of the Langevin thermostat to bring a system to thermal equilibrium, the applied dissipative forces are incompatible with simulations of the free molecular regime. In fact, due to the Brownian motion under the Langevin thermostat, no clustering can occur during the equilibration period. Thus, for the NVT production simulations we have used the NH thermostat with a coupling constant  $\tau = 0.1$  ps.

The additional NVE dynamics runs serve as a reference point for the NVT simulations since the issues related to latent heat and its subsequent removal via the thermostat are not present. Every occurrence of dimerization will release some quanta of energy,  $q > 0$ , equal to the binding energy, which is eventually dissipated through the thermostat. However, due to the global working principle of thermostats like NH and CSVN, latent heat release results in the temperature of the whole system being scaled down. The specific atoms of a cluster carrying the excess energy  $q$  are not distinguished by the algorithm and these atoms remain relatively “hot” while the rest of the system is cooled below the target temperature. The

magnitude of this divergence in molecular temperatures during clustering can be estimated in a simplified manner: On average, a thermostat (connected to a heat bath at  $T_0$ ) conserves the total kinetic energy of a system consisting of  $N$  identical monomers with  $n$  atoms each,  $K = (3nN - 3)k_B T_0/2$ . In addition to the kinetic energy carried by individual monomers (corresponding to temperature  $T_1$ ), each dimerization event introduces roughly  $q/2$  kinetic energy (for large compounds  $q$  is almost equally divided between kinetic and potential energy). Thus, the system’s kinetic energy after  $n_{\text{dim}}$  dimerization events can be expressed as

$$K \approx \frac{3nN}{2}k_B T_0 = \frac{3nN}{2}k_B T_1 + \frac{n_{\text{dim}}q}{2}. \quad (8)$$

The effective monomer temperature  $T^{\text{mon}}$  due to dimerizations in a globally thermostatted system is thus expected to be

$$T^{\text{mon}} = T_1 = T_0 - \frac{n_{\text{dim}}q}{3nNk_B}. \quad (9)$$

Clearly, the decrease in  $T^{\text{mon}}$  depends highly on the system size,  $N$  and  $n$ , and the excess energy,  $q$ . The latter can be readily calculated as the difference between the potential energies of thermally equilibrated monomers and the dimer. Carrying out such MD calculations for isolated monomers and dimers at  $T_0 = 200$  and 300 K, we obtained  $q = 0.80 \pm 0.01$  eV for  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimers. Thus, for example, if half of the monomers have transformed into dimers ( $n_{\text{dim}} = N/4$ ,  $n = 10.5$ ), the observed temperature of the rest of the monomers would be reduced by about 37% and 25% at  $T_0 = 200$  and 300 K, respectively. Since the monomer temperature decrease is not directly seen in the standard simulation output, such an effect can cause a notable hidden dynamical bias in the NVT simulations.<sup>11</sup> Thus, events observed at different moments during a simulation are not necessarily comparable with each other.

Possible temperature evolution of the system aside, the canonical rate of dimerization can be obtained from the first-order reaction equation for freely varying concentrations of

the reactants (“mol” and “ion”) at thermal equilibrium. Accordingly, the time-evolution of dimerization events in a constant volume  $V$  can be described as<sup>49</sup>

$$\frac{1}{V} \frac{dn_{\text{dim}}(t)}{dt} = \beta(T) \frac{N_{\text{mol}}(t)}{V} \frac{N_{\text{ion}}(t)}{V}. \quad (10)$$

Since all dimerization events are considered to be independent of each other, data obtained from  $N_{\text{sim}}$  individual simulations (carried out at equal temperature  $T_0$ ) can be analyzed together, and the average dimerization rate coefficient can be calculated as

$$\beta_{\text{stoc}}(T) = \frac{V \sum_{j=1}^{N_{\text{sim}}} n_{\text{dim}}(j)}{\sum_{j=1}^{N_{\text{sim}}} \sum_{i=1}^{n_{\text{dim}}(j)} N_{\text{mol}}(i, j) N_{\text{ion}}(i, j) \Delta t(i, j)}, \quad (11)$$

where  $N_{\text{mol}}(i, j)$  and  $N_{\text{ion}}(i, j)$  are the average numbers of monomers between two dimerization events  $i$  and  $i - 1$  in the  $j$ th simulation, and the time elapsed between these events is  $\Delta t(i, j)$ . Similarly, the average monomer temperatures ( $T^{\text{mon}}$ ,  $T_{\text{tr}}^{\text{mon}}$ ,  $T_{\text{rot}}^{\text{mon}}$  and  $T_{\text{vib}}^{\text{mon}}$ ) are monitored between the events (for the first dimerization event of a simulation, the averaging starts after the equilibration period has ended).

A large amount of data and thus large  $N_{\text{sim}}$  is required to achieve sufficient accuracy for  $\beta_{\text{stoc}}$ . For each studied system (NVE or NVT,  $T_0 = 200$  or  $300$  K), we carried out at least 1,000 individual simulations, resulting in  $N_{\text{events}} = \sum_{j=1}^{N_{\text{sim}}} n_{\text{dim}}(j) > 10,000$ . The detailed numbers are presented in Table 1. These statistics allow for accurate determination of the rate coefficients using eq 11. While the formation of  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimers is most likely, other clusters occur as well and this is taken into account when calculating the time-dependent  $N_{\text{mol}}$  and  $N_{\text{ion}}$ . Here, we only concentrate on the rate of the ion–molecule dimer formation mainly due to the lack of good statistics for the other events.

The dimerization rates given by eq 11 for systems at  $T_0 = 200$  and  $300$  K are shown in Figures 5a and b, respectively. The simulations are run until the averages of  $\beta_{\text{stoc}}$  are properly converged, which requires about 1,000 individual dimerization events for the studied systems. Based on the results shown in Figure 5, we have roughly estimated that the uncertainty of

Table 1: Stochastic MD simulation results for  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization, carried out for NVT or NVE systems initialized at temperatures  $T_0 = 200$  and 300 K: Simulation box volume ( $V$  in  $10^7 \text{ \AA}^3$ ), number of individual simulations ( $N_{\text{sim}}$ ), total number of observed dimerization events ( $N_{\text{events}}$ ), dimerization rate coefficient given by eq 11 ( $\beta_{\text{stoc}}$  in  $10^{-15} \text{ m}^3\text{s}^{-1}$ ), and average monomer temperatures of different partitions (in K)

	$T_0$	$V$	$N_{\text{sim}}$	$N_{\text{events}}$	$\beta_{\text{stoc}}$	$T^{\text{mon}}$	$T_{\text{tr}}^{\text{mon}}$	$T_{\text{rot}}^{\text{mon}}$	$T_{\text{vib}}^{\text{mon}}$
NVT	200	2.7	2,180	22,608	$3.02 \pm 0.10$	$155.5 \pm 2.5$	$138.9 \pm 2.0$	$137.8 \pm 2.9$	$159.5 \pm 1.9$
NVE	200	2.7	1,000	11,296	$2.41 \pm 0.10$	$204.9 \pm 1.7$	$217.8 \pm 1.6$	$216.0 \pm 1.1$	$202.0 \pm 1.7$
NVT	300	2.7	1,000	10,891	$2.26 \pm 0.10$	$254.7 \pm 5.2$	$237.1 \pm 3.1$	$239.4 \pm 2.6$	$258.5 \pm 4.0$
NVE	300	2.7	1,000	13,074	$1.94 \pm 0.10$	$312.2 \pm 4.8$	$330.9 \pm 9.7$	$322.6 \pm 7.3$	$308.8 \pm 4.3$

$\beta_{\text{stoc}}$  is  $\delta\beta_{\text{stoc}} \approx 10^{-16} \text{ m}^3\text{s}^{-1}$ . For both initial temperatures considered, the simulations carried out with a thermostat (NVT) result in statistically significantly higher values of  $\beta_{\text{stoc}}$  compared to the NVE simulations. This difference is especially notable for  $T_0 = 200$  K.

As anticipated based on prior studies<sup>3,11</sup> and eq 9, thermostating during dimer formation affects the monomers by decreasing their kinetic energies. The average monomer temperatures ( $T^{\text{mon}}$ ,  $T_{\text{tr}}^{\text{mon}}$ ,  $T_{\text{rot}}^{\text{mon}}$ ,  $T_{\text{vib}}^{\text{mon}}$ ) over all simulations are presented in Table 1. Average  $T^{\text{mon}}$  are further shown as a function of formed dimers for the studied systems in Figures 5c and d. While the monomer temperatures remain near  $T_0$  in the NVE simulations, the NVT simulations exhibit a significant drop of the average value of  $T^{\text{mon}}$ . The monomer cooling in the NVT simulations closely follows the simplified model given by eq 9 until about 10 dimers have appeared in the system. Based on the general trend of decreasing  $\beta_{\text{stoc}}$  with increasing  $T^{\text{mon}}$  (a characteristic feature for ion–dipole capture as will be discussed later), the observed discrepancy between  $\beta_{\text{stoc}}$  obtained from NVT and NVE simulations can be attributed to the altered monomer temperature in the NVT simulations. At the late stages of the simulations, the average time between consecutive dimerization events is long. During this time, existing dimers (or other clusters) are able to dissociate and thus release “hot” monomers (a reverse process to association). This is the main reason for the discrepancy between eq 9 and the observed  $T^{\text{mon}}$  as  $n_{\text{dim}} > 9$ . Unlike the NVT simulations, the monomer temperatures stay near the target temperature when no thermostat is applied during the period of dimerization. In these NVE simulations, the excess energy does not affect the

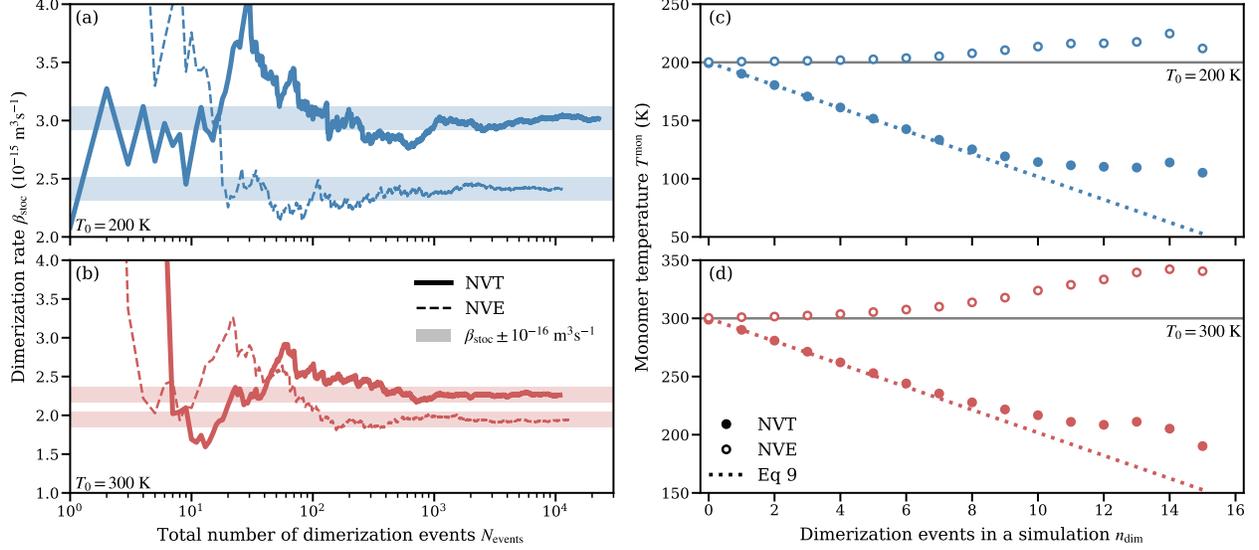


Figure 5: Left: Running average of the dimerization rate coefficients obtained from the stochastic MD simulations at (a)  $T_0 = 200$  and (b) 300 K as a function of observed total events  $N_{\text{events}}$  calculated using eq 11. The results from the NVT and NVE simulations are shown as thick solid lines, and thin dashed lines, respectively. The shaded areas in panels (a) and (b) indicate the respective final averages  $\beta_{\text{stoc}} \pm 10^{-16} \text{ m}^3 \text{ s}^{-1}$ . Right: Average monomer temperatures after  $n_{\text{dim}}$  dimerization events observed in the simulations at (c)  $T_0 = 200$  and (d) 300 K. Results from NVT and NVE simulations are shown as full, and open circles, respectively. The dashed lines show the predicted monomer temperature evolution due to latent heat release during dimerization, given by eq 9.

monomers significantly, but again at the late stages the monomer temperatures increase. It should also be noted that the statistics are best for small  $n_{\text{dim}}$ : about 90% of the simulations result in at least 9 dimerization events, whereas 13 or more events were observed in only 10% of the simulations.

Due to the small number of degrees of freedom and natural fluctuation of temperature, the observed standard deviations cannot be taken as the uncertainties of the monomer temperatures. Instead, we have crudely estimated the error of  $T^{\text{mon}}$  to be the difference of the observed and expected standard deviation. In the error analysis, the effect of dimerization on monomer count is considered. The estimated errors are presented in Table 1; the uncertainties are minor, only about 1–3%, for most systems and temperatures.

### 4.3 Dimerization Trajectory Sampling Simulations

The issues related to excess energy release during dimerization and its appropriate removal from the simulated system can be avoided by concentrating on the dynamics of an isolated event. A straightforward approach is to set two compounds on a collision course with some well-defined initial relative velocity  $v$  and collision geometry, i.e., impact parameter  $b$  (depicted in Figure 1). In these classical trajectory simulations,<sup>20,32,33,50,51</sup> a number of individual trajectories are studied for each pair of  $v$  and  $b$  by varying the initial state of the system. Adequate sampling over different rotational and vibrational states (and molecular orientations) at the desired temperature was achieved by equilibrating both  $(\text{CH}_3)_2\text{NH}$  and  $(\text{CH}_3)_2\text{NH}_2^+$  with a Langevin thermostat at a separation distance larger than the force field cut-off (150 Å). After equilibration, the compounds were moved to 50 Å apart along the  $x$ -axis and  $b$  along the perpendicular  $z$ -axis, and given a velocity along the  $x$ -direction of  $v_x = \pm v/2$  towards each other. The range of sampled relative velocities was selected so that at least 99% of the Maxwell-Boltzmann distribution was covered in intervals of 20  $\text{ms}^{-1}$ , e.g., for the simulation temperature of 300 K, the range was 100 to 1100  $\text{ms}^{-1}$ . The system was then simulated without any thermostating until either the criterion for dimerization (given in Section 4.1) was satisfied, or the com distance exceeded 150 Å (no dimerization). By running  $n_{\text{samp}} = 300$  to 600 individual simulations for each combination of  $v$ ,  $b$ , and  $T$ , the dimerization probability function  $P(v, b, T)$  was determined. For each value of  $v$ , we sampled impact parameters in steps of 1 Å, starting from 0 Å up to the first impact parameter for which the collision probability was zero. The statistical uncertainty for  $P(v, b, T)$  is taken from a binomial distribution given by

$$\delta P(v, b, T) = \sqrt{\frac{P(v, b, T) - P(v, b, T)^2}{n_{\text{samp}}(v, b, T)}}. \quad (12)$$

An example of the obtained  $P(v, b, T)$  distributions for the  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  system at  $T = 300$  K is shown in Figure 6a. For slowly approaching reactants, there is a

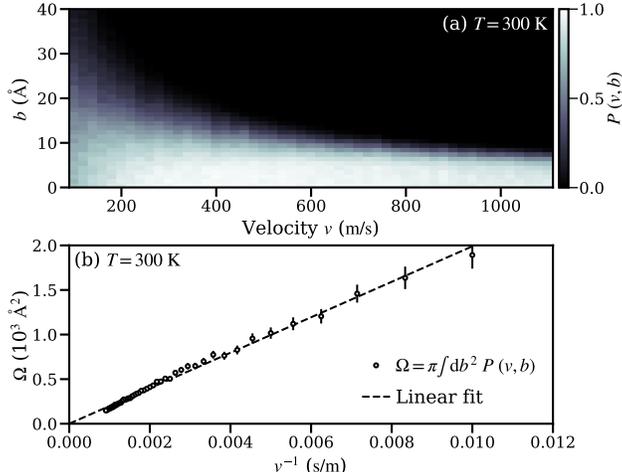


Figure 6: Example results from a trajectory simulation carried out for the  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  system at  $T = 300 \text{ K}$ : (a) Heat map of the dimerization probability as a function of relative velocity  $v$  and impact parameter  $b$ . (b) Reaction cross section as a function of reciprocal velocity  $v^{-1}$ . The computed cross sections (open circles) show linear dependency on reciprocal velocity,  $v^{-1}$ , within the uncertainty of the simulations.

significant probability of a dimerization even at relatively large values of  $b$ , as translational energy is required for to escape the field of force. For increasing values of  $v$ , the maximum value of  $b$  at which dimerization will occur decreases, and the transition in the dimerization probability as a function of  $v$  becomes more step-like. Based on the obtained dimerization probabilities, the reaction cross sections for each  $v$  can be determined readily as

$$\Omega(v, T) = \pi \int_0^\infty db^2 P(v, b, T). \quad (13)$$

For systems with notable long-range interaction, the cross section depends strongly on the impact velocity; as shown in Figure 6b,  $\Omega \propto v^{-1}$ .

The canonical reaction rate coefficient can be further calculated from the cross sections by integrating over the Maxwell-Boltzmann distribution function  $f(v, T)$ :

$$\beta_{\text{traj}}(T) = \int_0^\infty dv v f(v, T) \Omega(v, T). \quad (14)$$

The results given by eq 14 with the respective values from the stochastic MD simulations are

Table 2: Rate coefficients of  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization (in  $10^{-15} \text{ m}^3\text{s}^{-1}$ ) at various temperatures obtained from both trajectory and stochastic MD simulations

	Temperature (K)						
	100	150	200	250	300	350	400
$\beta_{\text{traj}}$	$3.23 \pm 0.10$	$2.72 \pm 0.08$	$2.38 \pm 0.08$	$2.16 \pm 0.07$	$1.98 \pm 0.07$	$1.86 \pm 0.06$	$1.74 \pm 0.06$
$\beta_{\text{stoc}} (\text{NVT})$			$3.02 \pm 0.10$		$2.26 \pm 0.10$		
$\beta_{\text{stoc}} (\text{NVE})$			$2.41 \pm 0.10$		$1.94 \pm 0.10$		

gathered in Table 2 and further presented in Figure 7. The rate coefficients determined from the trajectory simulations (open circles in Figure 7) are very close to the respective values of  $\beta_{\text{stoc}}$  obtained from the stochastic NVE simulations (full squares), whereas the results from the stochastic NVT simulations (full diamonds) significantly overestimate the dimerization rate. As pointed out in Section 4.2, the monomer temperatures (especially  $T_{\text{tr}}^{\text{mon}}$  and  $T_{\text{rot}}^{\text{mon}}$ ) are substantially lower than the target temperature  $T_0$  in the stochastic NVT simulations. Hence, for better comparison, the value of  $\beta_{\text{stoc}}$  obtained from NVT simulations should be assigned to either  $T^{\text{mon}}$  or some partitioning of it ( $T_{\text{tr}}^{\text{mon}}$ ,  $T_{\text{rot}}^{\text{mon}}$  or  $T_{\text{vib}}^{\text{mon}}$ ). Indeed, as shown in Figure 7 with the open diamonds, the results from the stochastic NVT simulations agree overall much better with the highly controlled trajectory simulations when the results are plotted against the rotational temperature  $T_{\text{rot}}^{\text{mon}}$  instead of  $T_0$ . For the stochastic NVE simulations, the temperature discrepancy is much smaller, as shown with the open and full squares. The magnitude and general temperature dependence of  $\beta_{\text{traj}}(T)$  and  $\beta_{\text{stoc}}(T_{\text{rot}}^{\text{mon}})$  are almost identical despite the fundamental difference of the simulation approaches. The agreement is slightly worse at low temperatures. The observed discrepancies can have many sources, but one distinctive contribution in the stochastic simulations is the presence of multiple monomers in a rather small volume that interact with each other at long range through electrostatic forces. On the other hand, in the trajectory simulations, the molecule and the ion are interacting beyond the cutoff distance used in the stochastic simulations (50 Å). While the longer range might result in higher  $\beta_{\text{traj}}$ , it has also been shown that even small periodic oscillations in the interaction energy can cause enough repulsive forces to effectively repel the molecules already at intermediate distances.<sup>32</sup>

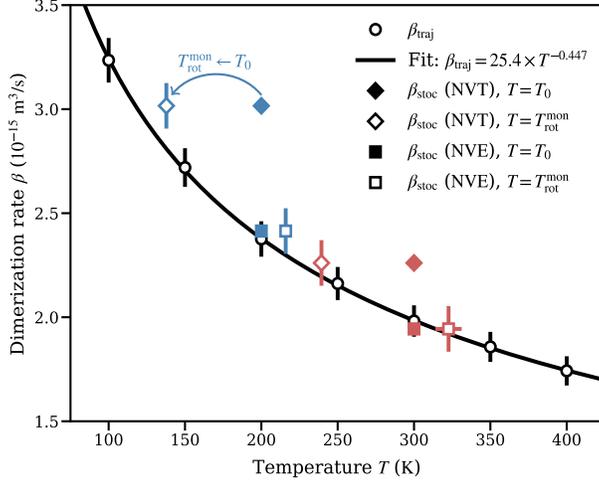


Figure 7: Simulated rate coefficients for  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization from Table 2 plotted as a function of temperature  $T$ . For the trajectory simulations (open circles),  $T$  corresponds to the well-defined temperature during equilibration. The results obtained from the stochastic simulations are presented for both the target temperature ( $T = T_0$ , full diamonds and squares) and the averaged rotational temperature of the monomers ( $T = T_{\text{rot}}^{\text{mon}}$ , open diamonds and squares). The blue and red markers correspond to stochastic simulations carried out for  $T_0 = 200$  and  $300$  K, respectively. The solid line shows the parametrization to the trajectory simulation data.

The rotational temperature  $T_{\text{rot}}$  is more important than  $T_{\text{tr}}$  or  $T_{\text{vib}}$  to characterize the temperature dependence of dimerization for the following reasons: Overall, the vibrational motion around the com of the monomers has a very minor effect on the intermolecular forces at distances relevant for dimerization ( $\gtrsim 10 \text{ \AA}$ ) as it will be averaged out quickly due to the high vibrational frequencies ( $\sim 50 \text{ ps}^{-1}$ ). Furthermore, the key reason for not presenting the results with respect to  $T_{\text{tr}}^{\text{mon}}$  can be seen in Figure 6b: the cross section of  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization is proportional to  $v^{-1}$  and according to eq 14 the reaction rate does not depend on the velocity distribution i.e., the translational temperature, because the integral  $\int dv f(v, T) = 1$ .

## 4.4 Importance of Proper Equilibration in Dimerization Simulations

The inverse relationship found between  $\beta$  and  $T_{\text{rot}}$  is characteristic of gas-phase ion–dipole collisions. This can be shown analytically by assuming that the interaction between a point-like ion and a rod-like dipole is isotropic (as the strong orientational dependencies are averaged out) and follows the relation<sup>52</sup>

$$U(r) = -\frac{A}{T_{\text{rot}}}r^{-4}, \quad (15)$$

where  $A$  is a system-specific constant and  $r$  is the distance between the ion and the dipole. Using a central field approach,<sup>33,53</sup> we consider that the ion approaches a stationary dipole from infinitely far away with some initial velocity  $v$  and impact parameter  $b$ . Both the system’s energy (initially only kinetic energy  $K = \mu v^2/2$ , where  $\mu$  is the reduced mass of the system) and angular momentum ( $\mu v b$ ) are conserved during the process. Thus, the following condition must hold:<sup>32,33</sup>

$$U(r) + K b^2 r^{-2} - K \leq 0. \quad (16)$$

The maximum of this equation is located at  $r_{\text{max}} = \sqrt{2A/(KT_{\text{rot}}b^2)}$ , and for a successful collision, this critical distance has to be crossed. Inserting  $r_{\text{max}}$  into eq 16, we obtain the maximum value of  $b^2$  leading to a “capture” (i.e., the ion collapses into the center of the field):

$$b_{\text{max}}^2 = \sqrt{\frac{8A}{\mu T_{\text{rot}}}}v^{-1}. \quad (17)$$

Ref 33 provides more detailed modeling and analyses of the dynamics of different atmospheric ion–dipole complexes.

As the cross section  $\Omega = \pi b_{\text{max}}^2$ , the behavior  $\Omega \propto v^{-1}$  seen in the trajectory simulations (Figure 6b) is indeed expected. Furthermore, according to eq 14,  $\beta \propto T_{\text{rot}}^{-1/2}$ , whereas the power law estimated from the trajectory simulations is  $\beta_{\text{traj}} \propto T^{-0.447}$ . Thus, we conclude

that the process of  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization behaves like a classical ion-dipole system, for which the rotational temperature is crucial.

The analysis of the specific process of gas-phase  $(\text{CH}_3)_2\text{NH}-(\text{CH}_3)_2\text{NH}_2^+$  dimerization not only demonstrates the general issue of thermostating in non-equilibrium systems but also emphasizes the role of proper initialization of the system’s molecular kinetic energy partitions. Here, it has been ensured that before entering the production stage the presented dimerization simulations are perfectly equilibrated and in accordance with the equipartition theorem. Our test simulations presented in Section 3, specifically Figure 3, show that the rotational temperature due to insufficient thermostating (by either NH or CSV) can be about two times higher than the target temperature. The effect of imperfect relaxation of rotational modes of motion is thus much greater than due to the improper rescaling of released excess energy during the stochastic NVT dimerization simulations, and the temperature discrepancy of  $T_{\text{rot}} = 2T_0$  would manifest itself in about 30% lower dimerization rate coefficients. Both biases are, however, highly setup- and system-specific: The efficiency of the equilibration process depends not only on the chosen thermostat, but also on the system density and the duration of the equilibration period. However, for a setup like in the trajectory simulations where natural collisions during equilibration do not occur, an improper equipartition will persist. When modeling non-equilibrium processes involving multiple energy-releasing events, the drift of reactant temperature away from  $T_0$  depends on the number of atoms and events, and the amount of excess energy released (see eq 9).

## 5 Conclusions

Temperature-controlled molecular dynamics (MD) simulations are routinely applied to study both equilibrium and non-equilibrium properties of various systems. Although the failures of naïve temperature control through simple velocity scaling algorithms are well documented<sup>2,12</sup> (e.g., the flying ice cube effect), even standard “best-practice” thermostating can still lead to

major alterations of the underlying dynamics both during and after preparation of a system in thermal equilibrium. The consequences are particularly severe for (but not limited to) simulations of gas-phase processes due to the limited interactions between compounds, the specific roles of their translational, rotational and vibrational motion, and the considerable heat of vaporization. Ideally, the thermostat should produce the correct kinetic energy distributions in a short simulation time, to minimize the computational cost before the start of the actual production run. In this study, we compared the Nosé-Hoover (NH; global and deterministic), the Canonical Sampling through Velocity Rescaling (CSVR; global and stochastic) and the Langevin (local and stochastic) thermostats, which all produce and maintain the canonical ensemble in ideal circumstances. The performance of these thermostats was tested specifically for relatively small, gas-phase systems with nanosecond-scale dynamics by studying their ability to obtain equilibrium and equipartitioning, and produce correct dynamics during simulations of the non-equilibrium process of  $(\text{CH}_3)_2\text{NH}-\text{(CH}_3)_2\text{NH}_2^+$  dimerization.

Despite the theoretical adequacy of a thermostat algorithm (rigorous temperature control, ability to produce canonical ensemble and ergodicity, etc.), the intrinsic design of the algorithm can lead to deviations in kinetic energy distributions and introduce dynamical artifacts. Crucially, undesired deviations in kinetic energies can easily go unnoticed as the system appears to be in equilibrium, with the correct overall temperature, and temperature fluctuations. Within the standard scheme for preparing a gas-phase system of flexible many-atom monomers, the Langevin thermostat is found to be very efficient, whereas the NH and CSVR thermostats are unable to produce the correct kinetic energy partitioning in an acceptable amount of equilibration time. Our test simulations for NH and CSVR thermostats show that without actual collisions, monomers are overexcited with respect to their translational and rotational modes at the expense of vibrational energy. In the context of barrierless ion-dipole dimerization, where rotational temperature plays a key role, this hidden bias significantly reduces the dimerization rate coefficient. Effects arising from erroneous

equipartition of reactants can be even more drastic for systems with stronger temperature dependencies, e.g., bimolecular reaction complexes involving finite activation barriers.

In addition to the failure to produce correct equipartition during the initial preparation of the system, the kinetic energies can also drift away from the desired distribution during non-equilibrium NVT simulations using global thermostating schemes, where locally released excess energy is uniformly removed from the entire system. On the other hand, locally operating thermostats, such as Langevin, can alter the dynamics of the simulated system.<sup>8</sup> Here, the detailed monitoring of different modes of motion revealed the incorrect state of monomers in stochastic NVT simulations of the dimerization process using the NH thermostat. Redistribution of the released latent heat ( $q = 0.8$  eV per dimerization event) effectively leads to a reduction of monomer temperature by several tens of Kelvins, unphysically accelerating the dimerization process. This effect would be even more pronounced if covalent or ionic bonds were formed, where  $q \gg 1$  eV. If the post-reaction energy dissipation is not explicitly studied, the proper monomer temperature can be maintained rather well by running the simulations under NVE dynamics i.e., without thermostating. The issues related to global thermostating have been previously acknowledged in MD simulation setups for gas-phase nucleation,<sup>3,29–31</sup> where the latent heat released by a chain-reaction-like cluster formation and its redistribution by the thermostat can reduce the nucleation rates by several orders of magnitude.<sup>11</sup> Page et al.<sup>7</sup> reported similar issues regarding quantum mechanical MD simulations of carbon vapor condensations, and based on their analysis concluded that “reliable reaction dynamics and rigorous temperature control are apparently mutually exclusive”. As demonstrated in refs 3,7,9,29–31, to overcome the issues of using global thermostating in non-equilibrium simulations, an explicit chemically inert carrier gas (connected to a thermostat) can be introduced to the simulated system. Addition of carrier gas inevitably increases the computational cost notably and affects the transport properties, but unlike a thermostat, carrier gas introduces an actual atmosphere around the monomers and its effect on the nucleation rate can be accurately assessed theoretically.<sup>31</sup>

The issues pointed out in this paper can be avoided, but no single solution or perfect thermostat exists. We suggest the following, especially when simulating small systems involving bond formation: First and foremost, we advocate modelers to scrutinize the basic properties of their systems as exhaustively as possible: test energy partitioning and local temperature distributions (e.g., using the built-in commands given in Section 2.2), compare the observed temperature fluctuations against statistical prediction, and study the general dynamic and transport properties. When carrying out small-scale NVT simulations, special attention should also be paid to the default settings (thermostat implementation and effective number of dof) in the specific MD code. Very recently Xu et al.<sup>14</sup> demonstrated the importance of correct dof setting by calculating diffusion in nanoporous materials; incorrect settings are particularly disruptive when only a subset of atoms is coupled to a thermostat, or isolated non-continuous systems are simulated. In addition to the fact that the Langevin thermostat realizes correct kinetic energy distributions irrespective of the dof settings,<sup>14</sup> we have demonstrated that in gas-phase simulations it leads to perfect thermalization and equipartition with a minimal computational effort, making it an ideal choice for system initialization, where dynamics are secondary. Finally, during the production phase, one should carefully decide between NVT and NVE dynamics. In case of non-equilibrium processes involving excess energy, global thermostating is not a suitable approach to maintain isothermal conditions of the reactants, while local thermostating might not produce the desired transport properties. Overall, despite its limitations, only the NVE ensemble strictly represents physically well-defined Hamiltonian dynamics.

## Acknowledgement

We acknowledge funding from the European Research Council (Project No. 692891 DAMOCLES), the Academy of Finland flagship program (Grant No. 337549), the Centers of Excellence Program (Project No. 1346368 VILMA), and the University of Helsinki, Faculty

of Science ATMATH project. We also acknowledge the CSC-IT Center for Science in Espoo, Finland, for computational resources.

## References

- (1) Lemak, A.; Balabaev, N. On the Berendsen thermostat. *Mol. Simul.* **1994**, *13*, 177–187.
- (2) Harvey, S. C.; Tan, R. K.-Z.; Cheatham III, T. E. The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.* **1998**, *19*, 726–740.
- (3) Wedekind, J.; Reguera, D.; Strey, R. Influence of thermostats and carrier gas on simulations of nucleation. *J. Chem. Phys.* **2007**, *127*, 064501.
- (4) Lingenheil, M.; Denschlag, R.; Reichold, R.; Tavan, P. The “hot-solvent/cold-solute” problem revisited. *J. Chem. Theory Comput.* **2008**, *4*, 1293–1306.
- (5) Rosta, E.; Buchete, N.-V.; Hummer, G. Thermostat artifacts in replica exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **2009**, *5*, 1393–1399.
- (6) Toton, D.; Lorenz, C. D.; Rompotis, N.; Martsinovich, N.; Kantorovich, L. Temperature control in molecular dynamic simulations of non-equilibrium processes. *J. Phys. Condens. Matter.* **2010**, *22*, 074205.
- (7) Page, A. J.; Isomoto, T.; Knaup, J. M.; Irle, S.; Morokuma, K. Effects of molecular dynamics thermostats on descriptions of chemical nonequilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 4019–4028.
- (8) Basconi, J. E.; Shirts, M. R. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2887–2899.

- (9) Zhao, J.; Singh, V.; Grammatikopoulos, P.; Cassidy, C.; Aranishi, K.; Sowwan, M.; Nordlund, K.; Djurabekova, F. Crystallization of silicon nanoclusters with inert gas temperature control. *Phys. Rev. B* **2015**, *91*, 035419.
- (10) Ruiz-Franco, J.; Rovigatti, L.; Zaccarelli, E. On the effect of the thermostat in non-equilibrium molecular dynamics simulations. *Eur. Phys. J. E* **2018**, *41*, 1–13.
- (11) Halonen, R.; Zapadinsky, E.; Vehkamäki, H. Deviation from equilibrium conditions in molecular dynamic simulations of homogeneous nucleation. *J. Chem. Phys.* **2018**, *148*, 164508.
- (12) Braun, E.; Moosavi, S. M.; Smit, B. Anomalous effects of velocity rescaling algorithms: the flying ice cube effect revisited. *J. Chem. Theory Comput.* **2018**, *14*, 5262–5272.
- (13) Li, Z.; Xiong, S.; Sievers, C.; Hu, Y.; Fan, Z.; Wei, N.; Bao, H.; Chen, S.; Donadio, D.; Ala-Nissilä, T. Influence of thermostatting on nonequilibrium molecular dynamics simulations of heat conduction in solids. *J. Chem. Phys.* **2019**, *151*, 234105.
- (14) Xu, H.; Cabriolu, R.; Smit, B. Effects of degrees of freedom on calculating diffusion properties in nanoporous materials. *J. Chem. Theory Comput.* **2022**, *18*, 2826–2835.
- (15) Korpelin, V.; Kiljunen, T.; Melander, M. M.; Caro, M. A.; Kristoffersen, H. H.; Mammen, N.; Apaja, V.; Honkala, K. Addressing dynamics at catalytic heterogeneous interfaces with DFT-MD: anomalous temperature distributions from commonly used thermostats. *J. Phys. Chem. Lett.* **2022**, *13*, 2644–2652.
- (16) Baer, T.; Hase, W. L. *Unimolecular reaction dynamics: theory and experiments*; Oxford University Press, 1996; Vol. 31.
- (17) Hansen, K. *Statistical Physics of Nanoparticles in the Gas Phase*; Springer, 2018; Vol. 73.
- (18) Laidler, K. J. *Chemical kinetics*; McGraw-Hill New York, 1965; Vol. 2.

- (19) Clary, D. C. Rate constants for the reactions of ions with dipolar polyatomic molecules. *J. Chem. Soc., Faraday Trans. 2* **1987**, *83*, 139–148.
- (20) Chesnavich, W. J.; Su, T.; Bowers, M. T. Collisions in a noncentral field: a variational and trajectory investigation of ion–dipole capture. *J. Chem. Phys.* **1980**, *72*, 2641–2655.
- (21) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044–1048.
- (22) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated discovery and refinement of reactive molecular dynamics pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638–649.
- (23) Ford, J.; Seritan, S.; Zhu, X.; Sakano, M. N.; Islam, M. M.; Strachan, A.; Martínez, T. J. Nitromethane decomposition via automated reaction discovery and an ab initio corrected kinetic model. *J. Phys. Chem. A* **2021**, *125*, 1447–1460.
- (24) Elm, J.; Kubečka, J.; Besel, V.; Jääskeläinen, M. J.; Halonen, R.; Kurtén, T.; Vehkamäki, H. Modeling the formation and growth of atmospheric molecular clusters: A review. *J. Aerosol Sci.* **2020**, *149*, 105621.
- (25) Jiang, S.; Liu, Y.-R.; Huang, T.; Feng, Y.-J.; Wang, C.-Y.; Wang, Z.-Q.; Ge, B.-J.; Liu, Q.-S.; Guang, W.-R.; Huang, W. Towards fully ab initio simulation of atmospheric aerosol nucleation. *Nat. Commun.* **2022**, *13*, 6067.
- (26) Loukonen, V.; Kuo, I. W.; McGrath, M.; Vehkamäki, H. On the stability and dynamics of (sulfuric acid)(ammonia) and (sulfuric acid)(dimethylamine) clusters: A first-principles molecular dynamics investigation. *Chem. Phys.* **2014**, *428*, 164–174.
- (27) Loukonen, V.; Bork, N.; Vehkamäki, H. From collisions to clusters: first steps of sulphuric acid nanocluster formation dynamics. *Mol. Phys.* **2014**, *112*, 1979–1986.

- (28) Zhao, J.; Baibuz, E.; Vernieres, J.; Grammatikopoulos, P.; Jansson, V.; Nagel, M.; Steinhauer, S.; Sowwan, M.; Kuronen, A.; Nordlund, K.; Djurabekova, F. Formation mechanism of Fe nanocubes by magnetron sputtering inert gas condensation. *ACS Nano* **2016**, *10*, 4684–4694.
- (29) Ayuba, S.; Suh, D.; Nomura, K.; Ebisuzaki, T.; Yasuoka, K. Kinetic analysis of homogeneous droplet nucleation using large-scale molecular dynamics simulations. *J. Chem. Phys.* **2018**, *149*, 044504.
- (30) Halonen, R.; Tikkanen, V.; Reischl, B.; Dingilian, K. K.; Wyslouzil, B. E.; Vehkamäki, H. Homogeneous nucleation of carbon dioxide in supersonic nozzles II: molecular dynamics simulations and properties of nucleating clusters. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4517–4529.
- (31) Tikkanen, V.; Reischl, B.; Vehkamäki, H.; Halonen, R. Nonisothermal nucleation in the gas phase is driven by cool subcritical clusters. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2201955119.
- (32) Halonen, R.; Zapadinsky, E.; Kurtén, T.; Vehkamäki, H.; Reischl, B. Rate enhancement in collisions of sulfuric acid molecules due to long-range intermolecular forces. *Atmos. Chem. Phys.* **2019**, *19*, 13355–13366.
- (33) Neefjes, I.; Halonen, R.; Vehkamäki, H.; Reischl, B. Modeling approaches for atmospheric ion–dipole collisions: all-atom trajectory simulations and central field methods. *Atmos. Chem. Phys.* **2022**, *22*, 11155–11172.
- (34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (35) Loukonen, V.; Kurtén, T.; Ortega, I.; Vehkamäki, H.; Padua, A. A.; Sellegri, K.; Kul-

- mala, M. Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water—a computational study. *Atmos. Chem. Phys.* **2010**, *10*, 4961–4974.
- (36) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (37) Our simulations were performed using the 29 Sep 2021 release of LAMMPS. For more detail about the simulation package and the used commands, please consult LAMMPS Documentation, <https://docs.lammps.org/Manual.html>.
- (38) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (39) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2018**, *1*, 5957.
- (40) Tuckerman, M. *Statistical mechanics: theory and molecular simulation*; Oxford University Press, 2010.
- (41) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (42) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.
- (43) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (44) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **1978**, *17*, 1302.

- (45) Allen, M. P.; Tildesley, D. J. *Molecular simulation of liquids*; Oxford University Press, 1987.
- (46) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (47) Bussi, G.; Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **2007**, *75*, 056707.
- (48) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics: Volume 5: Statistical Physics*; Pergamon Press, 1969.
- (49) Kröger, L. C.; Kopp, W. A.; Döntgen, M.; Leonhard, K. Assessing statistical uncertainties of rare events in reactive molecular dynamics simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3955–3960 .
- (50) Dugan Jr, J. V.; Magee, J. L. Capture collisions between ions and polar molecules. *J. Chem. Phys.* **1967**, *47*, 3103–3112.
- (51) Goudeli, E.; Lee, J.; Hogan Jr, C. J. Silica nanocluster binding rate coefficients from molecular dynamics trajectory calculations. *J. Aerosol Sci.* **2020**, *146*, 105558.
- (52) Israelachvili, J. N. *Intermolecular and surface forces*; Academic Press, 2011.
- (53) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics: Volume 1: Mechanics*; Elsevier, 1976.

# TOC Graphic

