

# On the effects of the degrees of freedom on calculating diffusion properties in nanoporous materials

Henglu Xu,<sup>†</sup> Raffaella Cabriolu,<sup>\*,†,‡</sup> and Berend Smit<sup>\*,†</sup>

<sup>†</sup>*Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Switzerland*

<sup>‡</sup>*Department of Physics, Norwegian University of Science and Technology (NTNU), Høgskoleringen 5, Realfagbygget D5-149, 7491 Trondheim, Norway*

E-mail: raffaela.cabriolu@ntnu.no; berend.smit@epfl.ch

## Abstract

If one carries out a molecular simulation of  $N$  particles using periodic boundary conditions, linear momentum is conserved and hence the number of degrees of freedom is set to  $3N - 3$ . In most programs this number of degrees of freedom is the default settings. However, if one carries out a molecular simulation in an external field, one needs to ensure that degrees of freedom are changed from this default setting to  $3N$ , as in an external field the velocity of the center of mass can change. Using the correct degrees of freedom is important in calculating the temperature and in some algorithms to simulate at constant temperature.

For sufficiently large systems, the difference between  $3N$  and  $3N - 3$  is negligible in the way. However, there are systems in which the comparison with experimental data requires molecular dynamics simulations of a small number of particles. In this work, we illustrate the effect of an incorrect setting of degrees of freedom in molecular

dynamic simulations studying the diffusion properties of guest molecules in nanoporous materials. We show that previously published results have reported a surprising diffusion dependence on the loading, which could be traced back to an incorrect setting of the degrees of freedom. As the correct settings are convoluted and counter-intuitive in some of the most commonly used molecular dynamics programs, we carried out a systematic study on the consequences of the various commonly used (incorrect) settings. Our conclusion, is that for systems smaller than 50 particles the results are most likely unreliable as these are either performed at an incorrect temperature or the temperature is incorrectly used in some of the results.

Furthermore, a novel and efficient method to calculate diffusion coefficients of guest molecules into nanoporous materials at zero loading conditions is introduced.

## 1 Introduction

Since the early days of molecular simulation, molecular dynamics has been used to compute the diffusion coefficients in gases and liquids.<sup>1</sup> Later these studies were extended to molecules in confined media. Understanding the diffusion properties of guest molecules adsorbed in the pores of nanoporous materials is of importance for almost all their applications. For example, the performances of molecular separations using nanoporous membranes are based on differences in diffusion coefficients of the components. And, in catalytic applications, diffusion of the reactants to the active sites can be a rate-limiting step. Ideally, by tuning basic building modules and topology of networks, a variety of pore sizes and shapes of nanoporous materials can be designed to enhance the separation efficiency and selectivity.<sup>2</sup> Furthermore, from a fundamental point of view, diffusion in confined geometries is an intriguing topic, as the diffusion behavior of molecules in these geometries can be very different from the gas or liquid phase.<sup>2-4</sup>

In experimental and computational studies of diffusion of guest molecules, different diffusion coefficients can be measured. For example, incoherent quasi-elastic neutron scattering

(QENS) and pulse-field gradient NMR (PFG-NMR) experiments report the self-diffusion coefficient ( $D_s$ ), while uptake experiments obtain the Fickian or transport diffusion coefficient ( $D_t$ ).<sup>5</sup> In a molecular simulation one computes either the self-diffusion coefficient ( $D_s$ ) or the collective diffusion coefficient ( $D_c$ ). The diffusion of particles due to a concentration gradient is quantified by the transport diffusion coefficient,  $D_t$ , while the collective diffusion coefficient,  $D_c$ , refers to the motion driven by a chemical potential gradient. If we know the adsorption isotherm, we can compute a correction factor to convert the concentration gradient into a gradient of the chemical potential, and hence relate the  $D_t$  to  $D_c$ .<sup>2,6,7</sup>

The self-, collective, and transport diffusion coefficients, coincide in the limit of zero loading condition, that is an idealized situation usually obtained by extrapolating the diffusion-loading dependence curve to the state with no guest molecules. Therefore zero-loading diffusion coefficient,  $D_0$ , allows an easy comparison with different experimental data, and hence most molecular simulation studies report an estimate of the zero loading diffusion coefficient. In the section 1 of Supporting Information (SI), a general overview on the mathematical formulation of the diffusion coefficients is given.

Diffusion coefficients in different nanoporous material can vary orders of magnitude, which makes this a challenging problem for molecular simulation. For systems with a very low diffusion coefficient, special simulation methods have been developed to estimate  $D_0$ .<sup>4,8-12</sup> For those systems in which the diffusion coefficient is high, one typically uses molecular dynamics. To save computational time, most of these simulations assume that the porous material is rigid, which sometimes can give unrealistic results.<sup>3,13</sup>

In this work, we focus on these molecular dynamics simulations. To approach the zero loading limit, one typically tries to reduce the number of molecules and/or increase the size of the simulation cell. This work was triggered by some recent work of Witherspoon et al.,<sup>14</sup> who have used a combination of NMR and MD simulations to study the  $\text{CH}_4$  diffusion and their binding at the coordinatively unsaturated metal sites in  $\text{M}_2(\text{dobdc})$  ( $\text{M} = \text{Mg}, \text{Ni}, \text{and Zn}$ ) frameworks at low and high guests' loading. At low loading, Witherspoon et al. observed

in the molecular dynamics simulation that the self-diffusion coefficient slowly increases until it reaches a maximum. This trend however, could not be confirmed experimentally. A maximum in the diffusion coefficient as a function of loading, has been observed in other simulation studies (e.g., Skoulidas et al.,<sup>15</sup> Dubbeldam et al.,<sup>16</sup> and Beerdsen et al.<sup>10</sup>), but unlike these studies a possible mechanism was lacking in the work of Witherspoon et al.

We show that the maximum found in the simulation of Witherspoon et al. is an artifact caused by the implementation of the algorithm used in these simulations to keep the temperature constant. In particular, the convoluted way in which the degrees of freedom appear in the different algorithms, made it difficult to detect errors in the way the temperature is set or computed in these MD programs.

In most of the simulations of porous materials, it is assumed that the material is rigid. From a simulation point of view, the crystal is then equivalent to an external field in a simulation. In an external field the number of degrees of freedom is  $3N$ , while in the gas phase or in a flexible material the number of degrees of freedom is  $3N - 3$  as the momentum of center of mass of the system is set to zero to avoid that the system moves with a constant velocity. In the default settings of several popular MD engines (e.g., LAMMPS<sup>17</sup>), one can set the degrees of freedom, but as we show changing this default setting is not sufficient to correctly simulate at the desired temperature, as these changes in default setting are not automatically applied in the algorithms for computing the temperature or simulation at constant temperature. For most simulations the number of particles is sufficiently large that the difference between  $3N$  and  $3N - 3$  is sufficiently small that it won't impact the conclusions. However, for diffusion in porous media this difference turns out to result in qualitatively different behaviours. For large systems the difference are expected to be very small, yet one may run into difficulties reproducing the results if someone simulates the system correctly.

Furthermore, this paper introduces a simple method to obtain zero loading diffusion coefficient that avoids having to simulate a very small number of particles. This method

makes feasible the direct comparison with experiments for which the low loading in those kind of materials is extremely relevant for industrial applications.

## 2 Computational Methods

### 2.1 Temperature and degrees of freedom

For our diffusion study, we have performed constant temperature simulations of  $N$  guest molecules diffusing in a crystalline nanoporous material. In most simulations it is assumed that the atoms of the crystal can be frozen. In such a rigid framework approximation, the crystal acts as an external field on the guest molecules. For such a system, the total linear momentum is **not conserved** and the total number of degrees of freedom ( $N_{\text{DoF}}$ ) coincides with  $3N$ , being  $N$  the number of guest molecules to which the thermostat is applied. In contrast, for a simulation in the gas or liquid phase, without the crystal, the number of degrees of freedom is  $3N - 3$ , as in this case the total linear momentum is conserved, and the center of mass movement is set to zero in a typical simulation. Also, if one would allow the atoms of the crystal to move, the total degrees of freedom would be  $3N + 3N_{\text{Crystal}} - 3$ .

In classical statistical mechanics, according to the equipartition theorem, the energy carried by each degree of freedom (DoF) is equal to  $\frac{1}{2}k_{\text{B}}T$ , and, for a system made of  $N$  particles having a velocity  $|v|$  each, the total kinetic energy at temperature  $T$  is equal to:

$$E_{k,\text{tot}} = \frac{1}{2}k_{\text{B}}T \times N_{\text{DoF}} = \sum_{i=1}^N \frac{1}{2}mv_i^2, \quad (1)$$

which gives the well-known results that one needs to ensure to have the correct number of degrees of freedom  $N_{\text{DoF}}$  to compute the temperature of the system.

## 2.2 Thermostats

There are different algorithms that can be used to simulate the  $NVT$  canonical ensemble. In these algorithms a thermal reservoir is coupled to the system, and, in their implementation it is important to take into account the degrees of freedom (DoF) correctly.

Below we will briefly review how the DoF appears in some commonly used algorithms (e.g., the Nosé-Hoover thermostat,<sup>1,18</sup> canonical sampling velocity rescaling thermostat,<sup>19</sup> and, Langevin thermostat<sup>20</sup>).

### 2.2.1 Nosé-Hoover thermostat

In the Nosé-Hoover approach an additional “virtual particle”, with coordinate  $s$ , effective mass  $Q$ , and momenta  $p_s$ , is inserted in the classical system of  $N$  particles. The equations of motion of this system follow from the extended Lagrangian, or corresponding Hamiltonian.<sup>1</sup>

The conserved quantity,  $H_{\text{Nosé}}$ , can be expressed in terms of the real variables,  $\mathbf{p}, \mathbf{r}$ :

$$H_{\text{Nosé}} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) + \frac{s^2 p_s^2}{2Q} + \frac{g}{\beta} \ln(s) \quad (2)$$

$$= \mathcal{H}(p, r) + \frac{\xi^2 Q}{2} + \frac{g}{\beta} \ln(s) \quad (3)$$

The  $H_{\text{Nosé}}$  is not a Hamiltonian, but it contains the usual Hamiltonian,  $\mathcal{H}(p, r)$ , of the classical system of the  $N$  particles. Furthermore, the above equation contains the so called thermodynamics friction  $\xi = sp/Q$ , the coefficient,  $g$ , the inverse temperature,  $\beta$ , and, the logarithm of the virtual coordinate, i.e.,  $\ln(s)$ , which are essential to reproduce the statistic of the canonical ensemble.<sup>1</sup>

From the conserved quantity  $H_{\text{Nosé}}$  the partition function  $Z_{\text{Nosé}}$  can be derived.<sup>1</sup> In this derivation one carries out the integration over  $s$  and  $p_s$ , and what remains is the term that is important for the present discussion:

$$Z_{\text{Nosé}} = C \int d\mathbf{p}^N \int d\mathbf{r}^N e^{-\frac{3N}{g}\beta\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)} \quad (4)$$

In Equation 4 one can see that only if  $g$  is set equal to the degrees of freedom, i.e.,  $3N$ , we recover the canonical ensemble. In this case, we integrate over all positions and momenta, and hence  $g = 3N$ , but if we constrain the center of mass,  $g = 3N - 3$ . Hence, one needs to be careful to set correctly the  $g$ , as this has an important effect on the equation of motion in the Nosé-Hoover scheme. This effect is more pronounced for a relatively small number of particles  $N$ .

### 2.2.2 Canonical Sampling Velocity Rescaling thermostat

The Canonical Sampling Velocity Rescaling (CSVR) thermostat<sup>19</sup> is scaling the total kinetic energy to maintain the desired temperature. The CSVR method consists in multiplying all the particles' velocities by a rescaling factor  $\alpha$ :

$$\alpha = \sqrt{\frac{E_{k,t}}{E_k}} \quad (5)$$

with the target  $E_{k,t}$  drawn from the canonical distribution for the kinetic energy:

$$P(E_{k,t}) dE_{k,t} \propto E_{k,t}^{(N_{\text{DoF}}/2)-1} e^{-\beta E_{k,t}} dE_{k,t} \quad (6)$$

The idea of the algorithm consists in enforcing the system's total kinetic energy,  $E_k$ , to a target kinetic energy value,  $E_{k,t}$ , using a stochastic procedure, that obtain the desired canonical distribution of  $E_{k,N}$  (Equation 9). This approach leads to the dynamics where the driving force is the difference between the target and the actual kinetic energy:

$$dE_k = (E_{k,t} - E_k) \frac{dt}{\tau} + 2 \sqrt{\frac{E_k E_{k,t}}{N_{\text{DoF}}}} \frac{dW}{\sqrt{\tau}} \quad (7)$$

where  $dW$  is a “Wiener” noise, and  $\tau$  is a relaxation time or damping constant that defines the strength of the coupling with the heat bath. From the formula above, it is clear that one needs to input the correct  $N_{\text{DoF}}$  in the CSVR thermostat.

### 2.2.3 Langevin thermostat

Another stochastic dynamics that leads to the canonical distribution is the Langevin dynamics on which the Langevin thermostat is built.<sup>20</sup> The Langevin equation of motion can be written as:

$$m_i \ddot{\mathbf{r}}_i = -\nabla U(\mathbf{r}_i) - \gamma_i \dot{\mathbf{r}}_i + \sqrt{\frac{2m\gamma}{\beta}} \mathbf{W}(t), \quad (8)$$

$-\nabla U(\mathbf{r}_i)$  is the deterministic force, while  $\gamma_i \dot{\mathbf{r}}_i$  and  $\mathbf{W}(t)$  represent the dissipative force and the stochastic noise, respectively. Due to the existence of dissipative forces, the system doesn't conserve linear momentum, and the center of mass undergoes a random walk. In the simulation of a system without external forces, one typically subtract the force applied on the center of mass, setting the constrain on the center of mass velocity. Because of this constrain, one need to compute the instantaneous temperature (Equation 1) with the correct degrees of freedom to have a consistent result with  $\beta$  in Equation 8.

## 3 Computation details

### 3.1 System and computational details

To calculate diffusion coefficients of guest molecules in different nanoporous structures, we perform MD simulations in the canonical ( $NVT$ ) ensemble using LAMMPS.<sup>17</sup> Periodic boundary conditions are applied to all MD simulations. Each simulated system is composed by the crystal which is treated with the rigid approximation, and mobile guest molecules. According to the rigid approximation, the atoms in the framework do not move during MD simulations. All unit cells are sufficiently replicated in order to avoid self interactions between periodic replica.

Our protocol consisted of two steps: insertion of guest molecules and MD simulation in the  $NVT$  ensemble. The corresponding guest molecules are randomly inserted into pores of the framework cell, except for MOF-74 and LTL, where  $\text{CH}_4$  molecules are always inserted in



the same independent channel. After the insertion, constant temperature simulations were carried out for 1 ns to equilibrate the system using Nosé–Hoover thermostat at the target temperature. The unit time step of all MD simulations is 1 fs.

After the equilibration of the samples, *NVT* simulations were carried out for a sufficiently long time to ensure the guest molecules motion reached the diffusion regime. This time varies depending on the structures and on the number of guest molecules per unit cell; In general, this time was around 100 ns for the simulations involving zeolites structures and 10 ns for the MOF samples. In all the simulations that have been performed, we ensured that the velocities of the guest molecules motion respected the canonical distribution at the imposed temperature. For data analysis purposes, the trajectories and velocities of guest molecules were stored every 1000 fs during the MD production simulations. The detailed process of calculating the diffusion coefficients and the relative errors from MD trajectories is provided in the SI section 1.

### 3.2 Simulation at zero loading

Experimentally one can measure different diffusion coefficients (e.g., self-diffusion coefficients, transport diffusion coefficients), which in the limit of zero loading become identical (see the SI section 1). It is important to simulate at the limit of zero loading. In this work we used a simple trick to simulate at these conditions.

For example if we simulate methane molecules in a porous material, the  $\text{CH}_4\text{--CH}_4$  interactions are neglected here in order to study the diffusion of  $\text{CH}_4$  at zero loading. The zero loading condition is achieved by excluding  $\text{CH}_4$  molecules from the neighbor lists of all the other  $\text{CH}_4$  during MD simulations, which not only turns off  $\text{CH}_4\text{--CH}_4$  interactions, but also allows different  $\text{CH}_4$  molecules to occupy the same space. This specific situation is equivalent to the ideal gas condition, where  $N$  far apart particles are inside an infinitely large supercell. In these conditions, 160  $\text{CH}_4$  molecules are inserted in each framework cell, and the zero-loading diffusion coefficients,  $D_0$ , from one single MD simulation is calculated.

This avoids computing the  $D_0$  by extrapolating the  $D$ -loading curve to zero guest molecules, while this simple trick provides accurate results with good statistics at a much lower computational cost. As discussed,  $D_s$ ,  $D_c$ , and  $D_t$  are equal in the limit of zero loading, i.e.,  $D_0$ , since diffusion is not affected by correlation effects at zero loading.

### 3.3 Default setting of thermostats

In several molecular simulation engines, the default implementation of the dynamics is tailored to systems with zero net force applied, i.e., no external field. For example, GRO-MACS<sup>21</sup> explicitly states that the center of mass velocity is set to zero at every steps. In other programs, e.g., LAMMPS, one can carry out a simulation in an external field and one can adjust the number of degrees of freedom. However, the commands to set correctly the number of degrees of freedom are counter intuitive, and this has been the cause of confusion in the literature for those simulations that use a relatively small number of particles. In the SI section 2, we outline in detail the settings we used to correctly and incorrectly simulate our systems. It is instructive to give here a short summary of the causes of these incorrect simulations.

The main cause of confusion is that in LAMMPS one can set the degrees of freedom using the command: `compute_modify compute-ID extra/DOF 0` to modify  $N_{\text{DoF}}$ , i.e. number of degrees of freedom. This is a so-called “command-line” change, and, this is given at the input at the same place where one sets for example the temperature. However, the issue is that this is not a system-wide change but only affect the specific property of `compute-ID`, which is used to compute the temperature but is not changing the degrees of freedom used in the thermostats and many other properties that rely on the instantaneous temperature (e.g., viscosity, pressure,  $\dots$ ).

To properly change the degrees of freedom one needs to carry out a “pseudo” temperature-gradient non-equilibrium simulation in which one defines a part of the system that is coupled to a separate thermostat. If the simulation is set-up like this, one can, both, set the tem-

perature and instruct LAMMPS for this special region to over write the default degrees of freedom. In this way one can change all the variables that depend on DoF correctly.

In the next section, we will refer to three types of settings:

1. Default settings: for which the degrees of freedom,  $N_{\text{DoF}}$ , is set to  $3N - 3$ .
2. Naive DoF settings: for which the degrees of freedom is set to  $3N$  at the “command line,” and,
3. Pseudo sub-system settings (or correct settings): here we ensure that the degrees of freedom are  $3N$  at all routines. With these settings in all subroutines the correct degrees of freedom are used.

### 3.4 Importance of DoF setting

In this section, we illustrate the importance of the correct setting by comparing the results of a simulations in which we use the three different LAMMPS settings. For this, we study a system of  $N$  methane molecules adsorbed in the pores of a metal-organic framework (MOF-74).

As the framework is rigid, the  $N$  methane molecules move in a external field, hence, there is no conservation of linear momentum and the correct degrees of freedom are  $N_{\text{DoF}} = 3N$ . It is instructive to monitor distribution of the kinetic energy ( $P(E_{k,N})$ ) in these simulations, and compare this distribution with the theoretical distribution, which, in the canonical ensemble of a system of  $N$  particles is given by:

$$P(E_{k,N}) = \frac{\beta^{(N_{\text{DoF}}/2)}}{\Gamma\left(\frac{N_{\text{DoF}}}{2}\right)} E_{k,N}^{(N_{\text{DoF}}/2)-1} \exp(-\beta E_{k,N}), \quad (9)$$

$$P(E_{k,i}) = \frac{\beta^{\frac{3}{2}}}{\Gamma\left(\frac{3}{2}\right)} E_{k,i}^{\frac{1}{2}} \exp(-\beta E_{k,i}), \quad (10)$$

where  $\beta = 1/k_B T$  is the inverse temperature, with  $k_B$  being the Boltzmann constant,  $T$  the temperature, and  $\Gamma$  represents the so called Gamma function.<sup>1</sup>

Figure 1 compares the  $E_k$  distribution for six  $\text{CH}_4$  molecules at 313 K for the Nosé-Hoover (a), CSVr (b), and Langevin (c) thermostat, using the default and the correct (pseudo sub-system settings) with the expected theoretical canonical results. These graphs clearly show

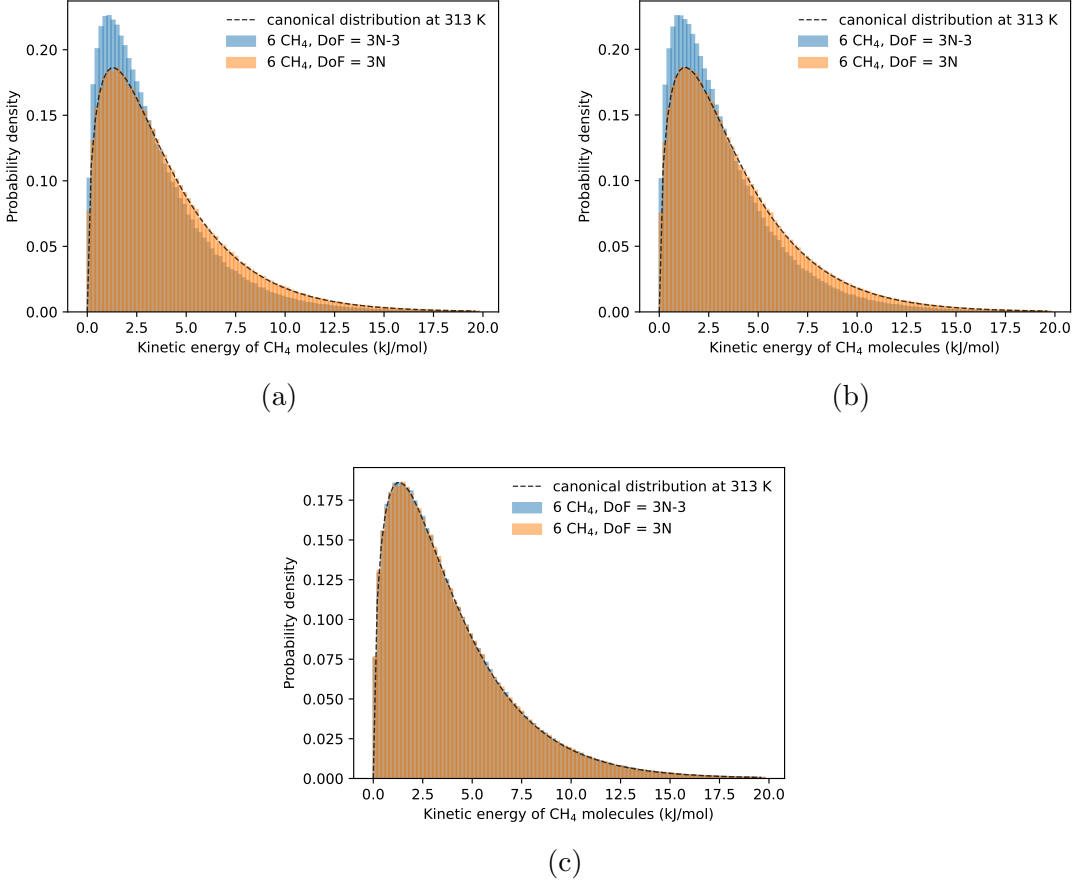


Figure 1:  $E_k$  distribution of six guest molecules of the default (blue histogram) and correct (pseudo-sub system) (orange histogram) DoF setting using (a) Nosé-Hoover, (b) CSVr, and (c) Langevin thermostat at 313 K. The black dashed line represents the expected theoretical canonical distribution.

that the default setting does not give the correct distribution. In fact, Equation 4 shows that an incorrect value of  $g$  in the Nosé-Hoover thermostat will give a Boltzmann distribution, but at a lower temperature (see Table 1). For the Langevin thermostat Figure 1 (c) shows that we do sample the correct distribution.

In Table 1, we show how the average kinetic energy per particle and the average instantaneous temperature (as reported by LAMMPS) depend on the number of particles for the

same case discussed above. As expected, if the number of particles is increased, the system approaches the correct kinetic energy per particle, but surprisingly for 100 particles one can still see a small but significant deviation. Additional figures of kinetic energy distribution of small  $N$  are presented in the SI section 3, while for larger  $N$  the distributions are almost identical to the expected canonical distribution.

This table also shows that using the default setting for Nosé-Hoover and CSVR thermostats, the incorrectly computed average temperature is very close to the desired target temperature, while for the Langevin simulation the temperature reported by LAMMPS deviates significantly from the desired temperature. This is unfortunate, as the incorrectly computed temperature in Nosé-Hoover and CVSR using incorrect  $N_{\text{DoF}}$  exactly “compensates” the incorrect kinetic energy, and hence the LAMMPS output gives users the suggestion that the simulations have been carried out at the desired temperature while from the kinetic energy it is clear that we are simulating a wrong dynamic. If one uses the naive DoF setting, one can see from the output that something is wrong.

For the Langevin dynamics case, irrespective of the settings, one does simulate a correct dynamics, i.e., kinetic energy per particle, and hence at the desired temperature (see Table 1). However, in the default setting the temperature is computed incorrectly. In the naive DoF setting, the temperature is computed correctly, but unfortunately the correct settings are not used globally. If one computes, for example, the viscosity or pressure the instantaneous temperature is computed with the incorrect default settings, and one also needs to define a pseudo sub-system to have the correct DoF for all calculations if a Langevin thermostat is used.

## 4 Results and discussions

We studied the diffusion properties for selected MOFs and zeolites at low loading, and, furthermore, we also applied a new approach to investigate the diffusion in the limit of zero

Table 1: Average kinetic energy per particle  $\langle E_k \rangle / N$  and instantaneous temperature  $\langle T_{\text{ins}} \rangle$  of  $N$  guest molecules using the default settings (a), the naive DoF settings (b), and correct settings (c) and three thermostats. The bold numbers didn’t reproduce the expected canonical distribution at 313 K.

(a) Default settings						
$N$	$\langle E_k \rangle / N$ (kJ/mol)			$\langle T_{\text{ins}} \rangle$ (K)		
	NH	CSVR	Langevin	NH	CSVR	Langevin
6	<b>3.26</b>	<b>3.26</b>	3.90	314.0	314.0	<b>375.6</b>
10	<b>3.52</b>	<b>3.51</b>	3.91	313.9	313.0	<b>348.7</b>
14	<b>3.62</b>	<b>3.62</b>	3.90	312.9	312.9	<b>337.1</b>
40	<b>3.80</b>	<b>3.81</b>	3.90	312.8	313.6	<b>321.0</b>
100	<b>3.86</b>	<b>3.86</b>	3.90	312.9	312.9	<b>316.2</b>
Canonical	3.90	3.90	3.90	313.0	313.0	313.0
(b) Naive DoF settings						
$N$	$\langle E_k \rangle / N$ (kJ/mol)			$\langle T_{\text{ins}} \rangle$ (K)		
	NH	CSVR	Langevin	NH	CSVR	Langevin
6	<b>3.26</b>	<b>3.26</b>	3.90	<b>261.7</b>	<b>261.7</b>	312.9
10	<b>3.52</b>	<b>3.51</b>	3.91	<b>282.5</b>	<b>281.7</b>	313.4
14	<b>3.62</b>	<b>3.62</b>	3.90	<b>290.6</b>	<b>290.6</b>	313.1
40	<b>3.80</b>	<b>3.81</b>	3.90	<b>305.0</b>	<b>305.8</b>	313.2
100	<b>3.86</b>	<b>3.86</b>	3.90	<b>310.0</b>	<b>310.0</b>	313.0
Canonical	3.90	3.90	3.90	313.0	313.0	313.0
(c) Correct settings						
$N$	$\langle E_k \rangle / N$ (kJ/mol)			$\langle T_{\text{ins}} \rangle$ (K)		
	NH	CSVR	Langevin	NH	CSVR	Langevin
6	3.91	3.90	3.90	313.5	313.1	312.9
10	3.90	3.90	3.91	313.3	313.4	313.4
14	3.90	3.91	3.90	313.0	313.1	313.1
40	3.91	3.90	3.90	312.9	313.2	313.2
100	3.90	3.90	3.90	313.1	313.0	313.0
Canonical	3.90	3.90	3.90	313.0	313.0	313.0

loading. The results from our work are compared to previous results, in which, we think for some works the conclusions were seriously affected by the misleading results generated by the default DoF settings in the LAMMPS MD simulations. The MD settings and interactions of systems for different case studies are given below. For more details on the LAMMPS input files used to generate the data in this paper, please find the files on the Materials Cloud

platform (<https://doi.org/10.24435/materialscloud:9w-gz>).

#### 4.1 Case study 1: CH<sub>4</sub> diffusion in three M<sub>2</sub>(dobdc) structures (MOF-74)

For this case study, we used the same structures and the same LAMMPS files, adopting the same simulation conditions as Witherspoon et al.,<sup>14</sup> and repeated the simulations of the diffusion of CH<sub>4</sub> molecules at low loading inside three M<sub>2</sub>(dobdc) structures (M: Mg, Ni, and Zn) at 313 K.

In these simulations, the interactions between CH<sub>4</sub> and MOF-74 structures are described using a parameterized density functional theory-derived force field.<sup>22</sup> CH<sub>4</sub>–CH<sub>4</sub> interactions are described using the TraPPE-UA force field from the Transferable Potentials for Phase Equilibria family, which adopts united-atom (UA) representations for alkyl groups.<sup>23</sup> The cutoff of non-bonded interactions is set at 12.8 Å. For zero-loading diffusion coefficients, we take the settings described in the Method section 3.2 where the CH<sub>4</sub>–CH<sub>4</sub> interactions are not included.

We computed the different diffusion coefficients (self-, collective, and transport) of CH<sub>4</sub> molecules at low loading (i.e., less than 0.6 CH<sub>4</sub> per open metal site M<sup>2+</sup>), inside the three M<sub>2</sub>(dobdc) structures. The results of these simulations are shown in Figure 2 (see SI section 1 for details on these different diffusion coefficients). It is interesting to compare our results for the self-diffusion coefficient with those obtained by Witherspoon et al.<sup>14</sup> Figure 2a shows that for a large number of particles, we obtain excellent agreement. If we decrease the concentration of particles, we observe an increase of the diffusion coefficient, while Witherspoon et al. observed a decrease. Such a decrease can be reproduced if we use the default setting of LAMMPS. Indeed, inspection of the input files of Witherspoon et al. confirms that these simulations have been carried out with the default settings (i.e.,  $N_{\text{DoF}} = 3N - 3$ ). Hence, the observed decrease of the diffusion coefficient with decreasing loading is caused by a decrease of the temperature.

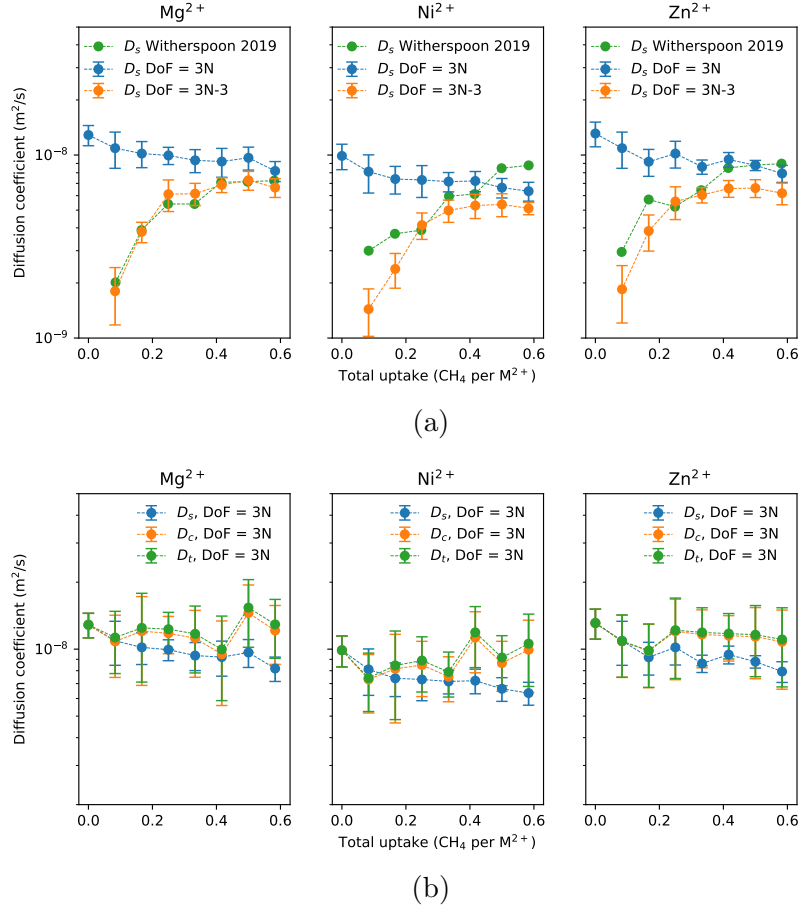


Figure 2: Self-, collective, and transport diffusion coefficients of CH<sub>4</sub> in M<sub>2</sub>(dobdc) (M: Mg, Ni, and Zn) (a) comparison of our results for the self-diffusion coefficient with the data of Witherspoon et al. (b) comparison of the different diffusion coefficients.



In Figure 2b we also report the collective and the transport diffusion coefficients. As expected, the three diffusion coefficients are gradually coinciding when the loading is approaching zero loading of guest molecules. In such range of low loading,  $\Gamma$  approaches to 1 and  $D_c$  is close to  $D_t$ , both appearing loading-independent, while  $D_s$  is slightly decreasing due to the steric effect.

## 4.2 Case study 2: Noble gases diffusion in the HKUST-1 structure

Parkes et al.<sup>24</sup> and Skoulidas et al.<sup>15</sup> studied the diffusion coefficients of Noble gases (Argon, Krypton, and Xenon) at low loading in a MOF structure, i.e. HKUST-1 at 298 K. Interestingly these studies gave a different behaviour of the loading dependence of the diffusion coefficient. The early result of Skoulidas et al.<sup>15</sup> results showed a small increase of the diffusion coefficient as a function of loading at low loading, while the later results of Parkes et al.<sup>24</sup> showed a much larger increase from low to high loading. In the context of this study it is important to note that Skoulidas et al. used their own code while Parkes et al. used LAMMPS.

To resolve these differences, we repeated these simulation and investigated the diffusion of these Noble gases for a wide range of loading (8 to 80 molecules/unitcell) in HKUST-1. The interactions between HKUST-1 and Noble gases are described using the Universal Force Field (UFF),<sup>25</sup> with the Lorentz-Berthelot rules applied. The cutoff of non-bonded interactions is set at 12.5 Å.

In Figure 3 we compared our results with the data of Parkes et al.<sup>24</sup> and Skoulidas et al.<sup>15</sup> The  $D_s$  results of Argon using correct DoF setting are consistent with that of Skoulidas et al.,<sup>15</sup> but don't agree with Parkes et al.<sup>24</sup> These simulations of Parkes et al., computed in the NVE ensembles using LAMMPS, are most likely using the default settings, and hence a temperature was computed in LAMMPS that is higher than the actual temperature of the simulations. In conclusion, the temperature of these simulations was most likely artificially lowered.

In our results, as expected, the self-diffusion coefficients of Argon and Krypton show that at low loading, there is a competition between the “small-cage” effect and the steric effect. The small cage effect exists in the structures with narrower windows and larger cages, for which, the presence of other molecules in the cages lowers the energy barriers of crossing-window motions. The window size of the small octahedral cages in HKUST-1, i.e. 4.1 Å, is close to the size of Argon and Krypton, 3.5 Å and 3.7 Å respectively, thus the small cage effect is observed here. As for Xenon with the size of 4.1 Å, the small octahedral cages cannot be explored by such a big guest, hence the steric effect dominates the diffusion coefficients, determining the  $D_s$  decreasing with the loading increases. However, the results of Xenon using default DoF setting underestimate the diffusion coefficients at low loading, indicating that the small cage effect also exists in Xenon diffusion in HKUST-1, which will result in incorrect conclusions.

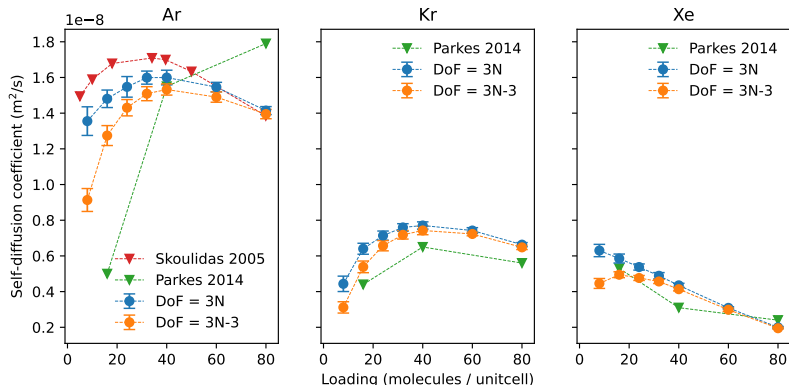


Figure 3: Diffusion coefficients of three Noble gases (left to right: Ar, Kr and Xe) in HKUST-1 in comparison with that of Parkes et al. and Skoulidas et al.

### 4.3 Case study 3: CH<sub>4</sub> diffusion in different zeolites structures

Beerdsen et al.<sup>9</sup> studied the diffusion of CH<sub>4</sub> in various zeolites structures and here we took four of them for representative comparison, i.e. MFI, FAU, ISV, and LTL. These studies also showed an increase of the diffusion coefficient as a function of loading for some of the structures. To reproduce these data we used the same crystalline structures and simulation

conditions described in the paper.<sup>9</sup>

In these simulations, the interactions between zeolites and  $\text{CH}_4$  are dominated by the oxygen atoms and the  $\text{CH}_4$  molecule is considered as a single reaction center. The interaction parameters are taken from the reported model which is optimized to reproduce adsorption properties.<sup>26,27</sup> The cutoff of non-bonded interactions is set at 12.5 Å.

In Figure 4 our results of self- and collective diffusion coefficients are compared with those obtained by Beerdsen et al.,<sup>9</sup> using a self-written code. The results using correct DoF setting are in excellent agreement with the values obtained by Beerdsen et al. In the same figure, the results using the default DoF setting, as expected, underestimate the diffusion coefficients at low loading.

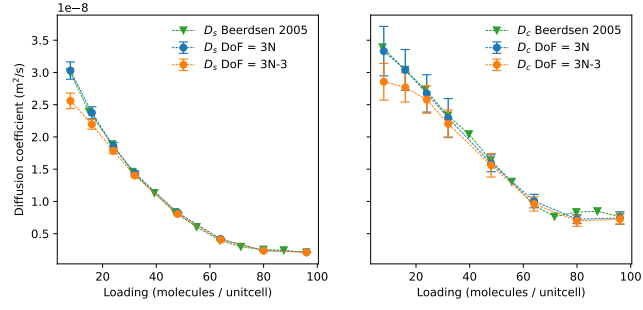
It is important to note that for these systems we could reproduce the increase of the diffusion coefficient as a function of loading that was observed for some of the structures. Hence, in this particular case this is not the result of an artifact in the simulation, but caused by the underlying diffusion mechanism as explained by Beerdsen et al.

#### **4.4 Case study 4: $\text{CH}_4$ diffusion at zero loading in 115 zeolites structures**

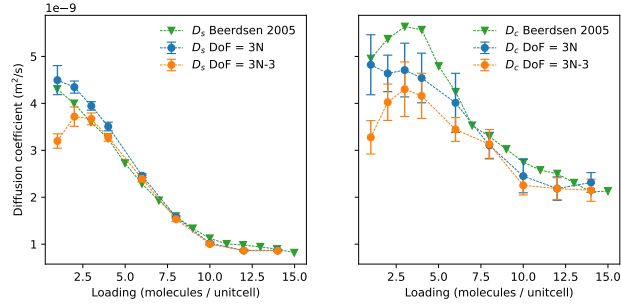
In case study 4, we investigated the  $\text{CH}_4$  diffusion at zero loading inside 115 zeolites at 500 K and we compared our results with the work by Mace et al.<sup>11</sup> In this work, the diffusion coefficients were obtained using standard MD and Transition State Theory methods.

All zeolites are described using the Universal Force Field<sup>25</sup> and  $\text{CH}_4$  molecules using TraPPE force field, respectively. The Lorentz-Berthelot mixing rules are applied for zeolites- $\text{CH}_4$  interactions, and, the cutoff of non-bonded interactions is set at 12.5 Å.  $\text{CH}_4$ - $\text{CH}_4$  interactions are not included here, following the setting described in Method section 3.2.

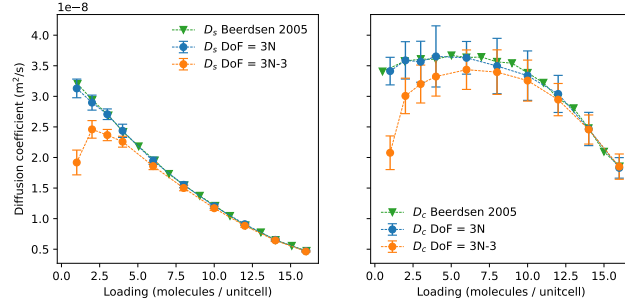
In Figure 5a, we compare our  $D_0$  for the 115 zeolites structures at 500 K with those obtained by Mace et al.<sup>11</sup> As expected, we observe that zero loading diffusion coefficients  $D_0$



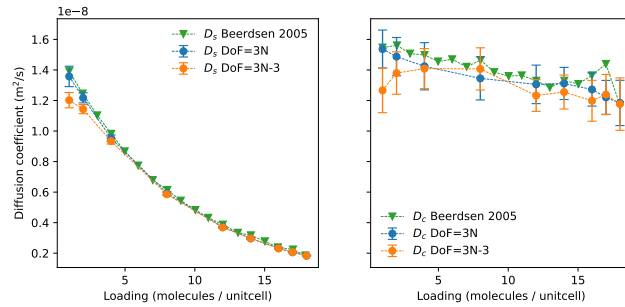
(a) FAU



(b) LTL



(c) ISV



(d) MFI

Figure 4: Self- (left) and collective (right) diffusion coefficients of CH<sub>4</sub> molecules in four zeolites with cage-type pores, FAU (a); channel-type pores LTL (b); and intersecting-channel-type pores ISV (c) and MFI (d). Note that for each of the structure the lowest loading has different number of methane molecules: MFI 16; LTL, FAU 8; and ISV 4.

(x-axis) using the correct DoF settings are systematically larger than those simulated using the default settings (y-axis).

In Figure 5b, we compare the  $D_0$  results (x-axis) with the results using TuTraSt method (y-axis) from Mace et al.<sup>11</sup> Whereas Mace et al. showed that the TuTraSt method slightly overestimates the diffusion coefficients, our results present, in general, lower diffusion coefficients. As this difference is about one order of magnitude, it does not change the main conclusion of the work of Mace et al. related to the practical applicability of the TuTraSt method.

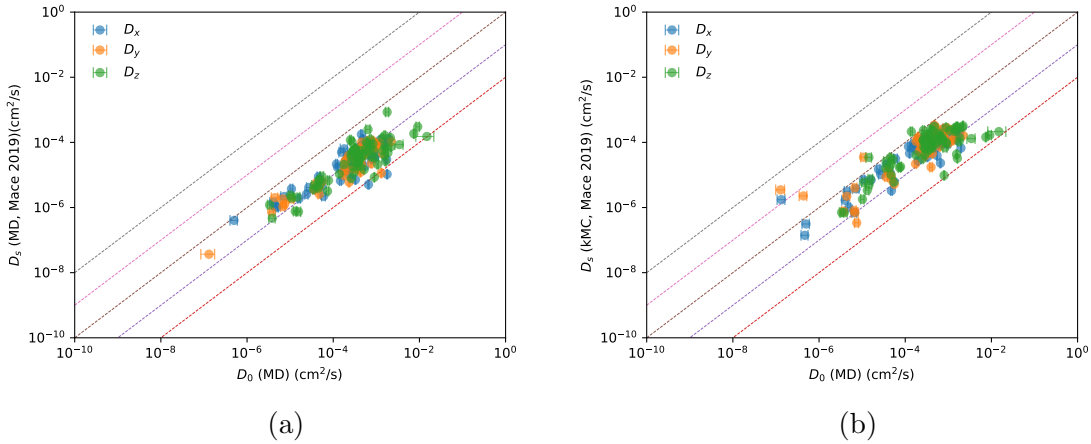


Figure 5: Diffusion coefficients at zero loading for 115 zeolites structures (a) compared to MD results, (b) compared to kMC results.

## 5 Concluding remarks

The main conclusion of this work is that molecular dynamics simulations in an external field with some of the most popular molecular dynamics programs are prone to error. Fortunately, in most simulations the number of particle is sufficiently large that these errors have little effect. However, if someone would like to reproduce these simulations, these errors may give small but inexplicable differences. This difference could easily be avoided, if one would use the correct way to simulate particles in an external field.

There are some systems in which simulations of a small number of particles are essential to compare with experimental data. As an example, we have studied the diffusion of gas molecules in porous media, where comparison with experiments requires to compute the diffusion coefficient in the limit of zero loading. As we have shown in this work, incorrect setting of the degrees of freedom can lead to qualitatively different results. In particular, the incorrect settings show a decrease of the diffusion coefficient if one decreases the number of particles, which is caused by the fact that the simulations were inadvertently done at a lower temperature. On the other hand, there are systems in which such a decrease is a real effect. In this article, we have repeated some of these simulations for which such a decrease was observed, including some of our own, and showed that for some, but not all, of these studies the effect was an artifact of the simulation.

From a computational point of view, our conclusions are not new and, interestingly, in the limited sample of simulations we reproduced, those simulations that relied on their own codes correctly implemented the degrees of freedom. However, as programs such as LAMMPS have become the default program for many MD studies, we felt it is important to emphasize the issue on the degrees of freedom. In particular, as, because of an “unfortunate” cancellation, in the default setting one simulates at the wrong temperature but as the temperature is also calculated incorrectly, the user gets the wrong impression that the simulations have been performed at exactly the desired temperature. We show that one can do the simulations correctly, but as we have illustrated, the settings are far from obvious.

Our conclusion is that the incorrect setting for systems of less than 50 particles may significantly impact the results. Either the simulations in an external field are carried out at a too low temperature (Nosé-Hoover or CSVN thermostats) or subsequent results that depend on the temperature are incorrectly computed (Langevin dynamics). For systems over 100 particles the differences are most likely too small to impact the results. Yet one should be aware of these issues when trying to reproduce literature results.

## Associated contents

### Supporting Information:

The supporting information is available free of charge on the ACS Publications website at DOI: DOI link.

The contents of SI organize as follows: theory and mathematical formalism of different diffusion coefficients and calculation details; A brief illustration of three DoF setting in LAMMPS command lines; Additional figures of kinetic energy distribution of different amounts of guest molecules.

LAMMPS input files for all case studies are available on the Materials Cloud platform (<https://doi.org/10.24435/materialscloud:9w-gz>).

## Author information

### Corresponding Authors:

Berend Smit: [berend.smit@epfl.ch](mailto:berend.smit@epfl.ch)

Raffaella Cabriolu: [raffaella.cabriolu@ntnu.no](mailto:raffaella.cabriolu@ntnu.no)

### ORCID

Henglu Xu (0000-0001-6457-8715)

Raffaella Cabriolu (0000-0002-9859-7444)

Berend Smit (0000-0003-4653-8562)

### Notes:

The authors declare no competing financial interest.

## Acknowledgments

This research is funded by the Swiss National Science Foundation (SNSF). H. X. acknowledges support from SNSF under grant number 200021\_172759. This work is supported by

the MARVEL National Centre for Competence in Research funded by the Swiss National Science Foundation (grant agreement ID 51NF40-182892). H. X. and R. C. acknowledge PRACE for awarding us access to Piz Daint, at the Swiss National Supercomputing Centre (CSCS), Switzerland with projects “pr128” and “s1019”. In Addition, R. C. would also like to acknowledge HPC-Europa3 visiting project (INFRAIA-2016-1-730897) for granting access to the computational facilities of CINECA. H. X. and R. C. thank Dr. Axel Kohlmeyer from LAMMPS developers group for the useful discussion and help regarding the input files.

## References

- (1) Frenkel, D.; Smit, B. Understanding Molecular Simulation: From Algorithms to Applications; Academic Press, 2002; Vol. 1.
- (2) Smit, B.; Maesen, T. L. M. Molecular Simulations of Zeolites: Adsorption, Diffusion, and Shape Selectivity. Chemical Reviews **2008**, 108, 4125–4184.
- (3) Jakobtorweihen, S.; Verbeek, M. G.; Lowe, C. P.; Keil, F. J.; Smit, B. Understanding the Loading Dependence of Self-Diffusion in Carbon Nanotubes. Physical Review Letters **2005**, 95, 1–4.
- (4) Beerdsen, E.; Smit, B.; Dubbeldam, D. Molecular Simulation of Loading Dependent Slow Diffusion in Confined Systems. Physical Review Letters **2004**, 93, 1–4.
- (5) Kärger, J.; Ruthven, D. Diffusion in Zeolites and other Microporous Solids; Wiley: New York, 1992.
- (6) Springer, T. Quasielastic Neutron Scattering for the investigation of Diffusive Motion in Solids and Liquids; Springer Tracts in Modern Physics; Springer, Berlin, 1972; Vol. 64.
- (7) Karger, J.; Ruthven, D. Diffusion in Zeolites and other Microporous Solids; Wiley: New York, 1992; Vol. 1.



- (8) Dubbeldam, D.; Smit, B. Computer Simulation of Incommensurate Diffusion in Zeolites: Understanding Window Effects. The Journal of Physical Chemistry B **2003**, 107, 12138–12152.
- (9) Beerdsen, E.; Dubbeldam, D.; Smit, B. Loading Dependence of the Diffusion Coefficient of Methane in Nanoporous Materials. The Journal of Physical Chemistry B **2006**, 110, 22754–22772.
- (10) Beerdsen, E.; Dubbeldam, D.; Smit, B. Understanding Diffusion in Nanoporous Materials. Physical Review Letters **2006**, 96, 1–4.
- (11) Mace, A.; Barthel, S.; Smit, B. Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field. Journal of Chemical Theory and Computation **2019**, 15, 2127–2141.
- (12) Tian, Y.; Xu, X.; Wu, J. Thermodynamic Route to Efficient Prediction of Gas Diffusivity in Nanoporous Materials. Langmuir **2017**, 33, 11797–11803.
- (13) Zimmermann, N. E. R.; Jakobtorweihen, S.; Beerdsen, E.; Smit, B.; Keil, F. J. In-depth study of the influence of host-framework flexibility on the diffusion of small gas molecules in one-dimensional zeolitic pore systems. J. Phys. Chem. C **2007**, 111, 17370–17381.
- (14) Witherspoon, V. J.; Mercado, R.; Braun, E.; Mace, A.; Bachman, J.; Long, J.; Blümich, B.; Smit, B.; Reimer, J. A. Combined Nuclear Magnetic Resonance and Molecular Dynamics Study of Methane Adsorption in M2(dobdc) Metal–Organic Frameworks. The Journal of Physical Chemistry C **2019**, 123, 12286–12295.
- (15) Skoulidas, A. I.; Sholl, D. S. Self-Diffusion and Transport Diffusion of Light Gases in Metal–Organic Framework Materials Assessed Using Molecular Dynamics Simulations. The Journal of Physical Chemistry B **2005**, 109, 15760–15768.

- (16) Dubbeldam, D.; Beerdsen, E.; Vlugt, T. J. H.; Smit, B. Molecular simulation of loading-dependent diffusion in nanoporous materials using extended dynamically corrected transition state theory. The Journal of Chemical Physics **2005**, 122, 224712.
- (17) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. Journal of computational physics **1995**, 117, 1–19.
- (18) Nosé, Canonical dynamics: Equilibrium phase-space distribution. Mol. Phys. **1986**, 57, 187–191.
- (19) Bussi, G.; Donadio, D.; Parinello, M. Canonical sampling through velocity rescaling. J. Chem. Phys. **2007**, 126.
- (20) Bussi, G.; Parinello, M. Accurate sampling using Langevin dynamics. Phys. Rev. E **2007**, 75.
- (21) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX **2015**, 1–2, 19–25.
- (22) Mercado, R.; Vlaisavljevich, B.; Lin, L.-C.; Lee, K.; Lee, Y.; Mason, J. A.; Xiao, D. J.; Gonzalez, M. I.; Kapelewski, M. T.; Neaton, J. B.; Smit, B. Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal-Organic Frameworks. The Journal of Physical Chemistry C **2016**, 120, 12590–12604.
- (23) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. The Journal of Physical Chemistry B **1998**, 102, 2569–2577.
- (24) Parkes, M. V.; Demir, H.; Teich-McGoldrick, S. L.; Sholl, D. S.; Greathouse, J. A.; Allendorf, M. D. Molecular dynamics simulation of framework flexibility effects on

- noble gas diffusion in HKUST-1 and ZIF-8. Microporous and Mesoporous Materials **2014**, 194, 190–199.
- (25) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. Journal of the American Chemical Society **1992**, 114, 10024–10035.
- (26) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H.; Krishna, R.; Maesen, T. L. M.; Smit, B. United Atom Force Field for Alkanes in Nanoporous Materials. The Journal of Physical Chemistry B **2004**, 108, 12301–12313.
- (27) Dubbeldam, D.; Calero, S.; Vlugt, T. J. H.; Krishna, R.; Maesen, T. L. M.; Beerdsen, E.; Smit, B. Force Field Parametrization through Fitting on Inflection Points in Isotherms. Physical Review Letters **2004**, 93, 088302.