Molecular Simulation towards Efficient and Representative Subsurface Reservoirs Modeling

Dissertation by

Ahmad Kadoura

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology, Thuwal,

Kingdom of Saudi Arabia

September 2016

The examination committee for the PhD dissertation defense by Ahmad Kadoura

Committee Chairperson: Shuyu Sun

Associate Professor, Earth Sciences and Engineering

Associate Professor, Applied Mathematics and Computational Science

First Internal Examiner: Georgiy Stenchikov

Professor, Earth Sciences and Engineering

Second Internal Examiner: Zhiping Lai

Associate Professor, Chemical and Biological Engineering

External Examiner: Abbas Firoozabadi

Professor, Chemical and Environmental Engineering

Yale University, New Haven, United States

Founder and Director, Reservoir Engineering Research Institute

Palo Alto, California, United States

King Abdullah University of Science and Technology

2016

ABSTRACT

Molecular Simulation towards Efficient and Representative Subsurface Reservoirs Modeling

Ahmad Kadoura

This dissertation focuses on the application of Monte Carlo (MC) molecular simulation and Molecular Dynamics (MD) in modeling thermodynamics and flow of subsurface reservoir fluids. At first, MC molecular simulation is proposed as a promising method to replace correlations and equations of state in subsurface flow simulators. In order to accelerate MC simulations, a set of early rejection schemes (conservative, hybrid, and non-conservative) in addition to extrapolation methods through reweighting and reconstruction of pre-generated MC Markov chains were developed. Furthermore, an extensive study was conducted to investigate sorption and transport processes of methane, carbon dioxide, water, and their mixtures in the inorganic part of shale using both MC and MD simulations. These simulations covered a wide range of thermodynamic conditions, pore sizes, and fluid compositions shedding light on several interesting findings. For example, the possibility to have more carbon dioxide adsorbed with more preadsorbed water concentrations at relatively large basal spaces. The dissertation is divided into four chapters. The first chapter corresponds to the introductory part where a brief background about molecular simulation and motivations are given. The second chapter is devoted to discuss the theoretical aspects and methodology of the proposed MC speeding up techniques in addition to the corre-

sponding results leading to the successful multi-scale simulation of the compressible single-phase flow scenario. In chapter 3, the results regarding our extensive study on shale gas at laboratory conditions are reported. At the fourth and last chapter, we end the dissertation with few concluding remarks highlighting the key findings and summarizing the future directions.

ACKNOWLEDGEMENTS

At first, I would like to thank my supervisor and dissertation defense committee chair, Dr. Shuyu Sun and the committee members, Dr. Georgiy Stenchikov, Dr. Zhiping Lai and Dr. Abbas Firoozabadi for their valuable feedback to help making this dissertation decent.

My deep appreciation for the help and support I received throughout my PhD years from Dr. Shuyu Sun, Dr. Amgad Salama and Dr. Arun Nair during the whole process of coding, writing and scientific discussions.

Finally, my heartfelt gratitude is extended to my parents and my wife for their encouragement, care and prayers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Motivation

It is crucial for reservoir flow simulators to have robust thermodynamic models that can accurately estimate the various thermophysical properties of the simulated subsurface fluid. These quantities may include both equilibrium properties (vapor pressure, heat capacity, compressibility, etc.) and transport properties (viscosity, diffusion coefficient, etc.). So far, the continuum-based models (e.g. equations of state and correlations) have played the major role in fulfilling this task [1,2]. Nonetheless, with the increase in exploiting the non-conventional reservoirs, the continuum-based models have faced two main challenges [3, 4]: (i) Lack of experimental data availability due to operating under extreme pressure and/or temperature conditions, in addition to presence of hazardous materials (e.g. corrosive, toxic and explosive). (ii) Working at extremely small scale (e.g. tight formations and shale gas reservoirs), where the continuum assumption might be violated.

On the other hand, and based on the concepts of statistical thermodynamics and quantum mechanics, these thermophysical properties of interest are predictable if one can monitor the evolution of the positions and momenta of the molecules forming the fluid system beside the adequate potential function to evaluate the inter- and intra-molecular interactions [5–8]. Lately, and due to advancements in computational

capabilities and algorithms, researchers were successful in developing and applying the Monte Carlo (MC) and Molecular Dynamics (MD) simulation methods. However, in comparison to classical thermodynamic models, the main disadvantage of molecular simulation is the much higher computational cost needed. Nonetheless, the advantages of molecular simulation are many, and evident [3, 4].

Therefore, we believe that molecular simulation can provide a suitable set to model subsurface reservoirs over the classical models where extreme thermodynamic conditions, often at small scales, are expected. This reflects the main motivation behind proposing the usage of molecular simulation techniques in order to simulate subsurface reservoirs. The conducted research has mainly focused on two cases. In the first case, MC molecular simulation was used to replace the classical thermodynamic models in estimating fluids thermophysical properties needed for flow simulators. For this purpose, several methods have been developed in order to accelerate MC simulations and make such coupling possible. In the second, both MC and MD simulations were used to investigate different aspects of gas sorption and transport phenomena in the inorganic part of shale at various thermodynamic conditions.

1.2 General Background

Molecular simulation considers systems at extremely small scale. It can predict a wide range of properties of these systems by monitoring the interactions of their components. Mainly two molecular simulation methods exist (see Fig. 1.1), Monte Carlo (MC) and Molecular Dynamics (MD). MC molecular simulation is a statistical approach that works by generating a large number of molecular configurations. Various elemental changes, known as MC trials (Fig. 1.2), are introduced in order to generate these configurations. Whenever a trial is performed it is subjected to an acceptance test. The process of accepting and rejecting these MC trials creates different con-

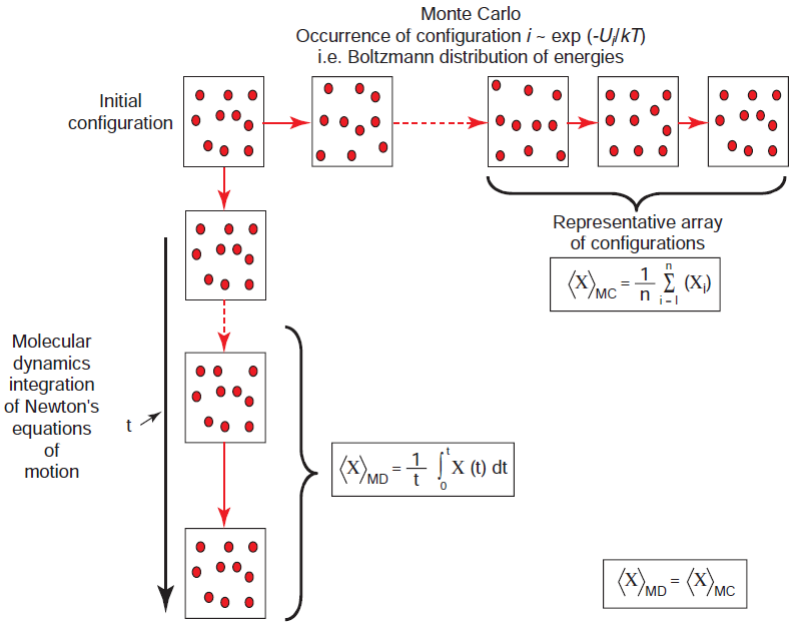

Figure 1.1: Average properties are determined from ensemble averages in MC while time averages are used in MD. Nonetheless, these averages are equivalent as a consequence of ergodicity (Ungerer et al., 2006).

figurations. The collective sequence of these created configurations is called the MC Markov chain. Averages and statistical fluctuations of the MC Markov chains can help evaluate equilibrium properties of fluids at single- and two-phase states.

On the other hand, MD simulation involves solving the Newton's equations of motion. In contrast to MC, MD method tracks the time evolution of the molecular system. Thus, MD simulation is mainly used in predicting time dependent properties while it can be also used in studying equilibrium ones. Another important aspect of molecular simulation is the statistical ensemble. Statistical ensembles define the environment at which the simulation is conducted. The experimental environment is created by subjecting the simulated molecules to a certain set of constraints, such as imposed temperature (T) or imposed pressure (p). Therefore, depending on the required application different ensembles could be chosen (Table B.1 in Appendix B). Fortunately various ensembles exist, each serving a different purpose. Each of these ensembles is characterized by its constrained variables and probability density.

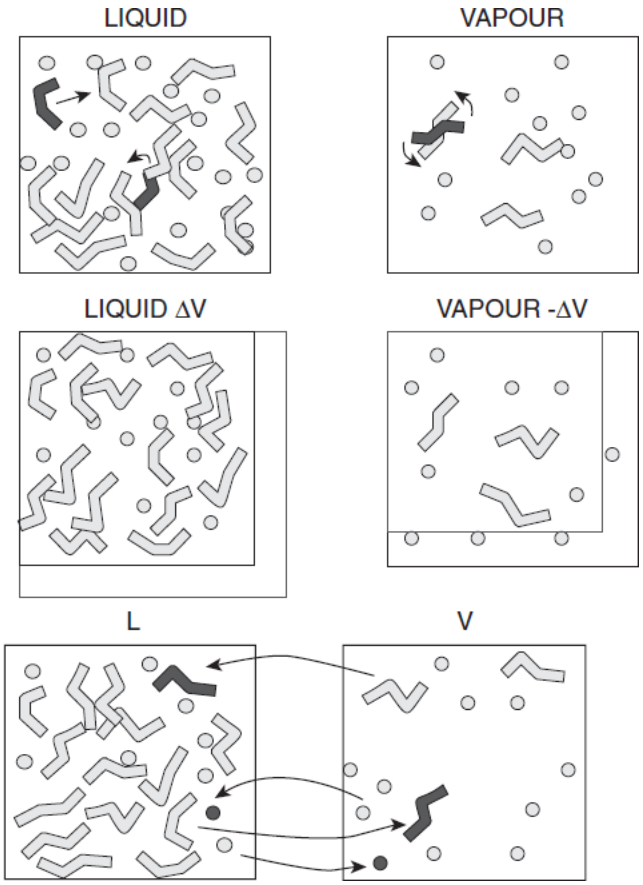

Figure 1.2: Group of possible MC trials in Gibbs ensemble, starting from top, configurational trials (molecule displacement, rotation, bending, etc.), volume changes and molecules transfers (Ungerer et al., 2006).

Table B.1 in Appendix B summarizes the main ensembles that have been used in the dissertation work. Noting that N, V, T, p and μ refer to the number of simulated molecules, simulation box volume, temperature, pressure and component's chemical potential; whereas N_{tot} and V_{tot} are the total number of molecules and total volume, respectively, as two simulation boxes are used in Gibbs ensemble each representing a different phase (liquid or gas).

Chapter 2

Multi-Scale Reservoir Flow Simulators

As stated earlier, the expensive computational cost is the main obstacle researchers have to face when applying molecular simulation in estimating fluids thermophysical properties needed for subsurface reservoir flow simulators. The main challenge resides in coupling two different time scales. As flow simulators will often be idle waiting for MD or MC routines to finish their cycles and provide estimates. The strategy adopted, to tackle this problem, is based on creating offline table consisting of a large number of pre-generated MC Markov chains at various uniformly spaced thermodynamic conditions. That offline table will serve as a database to compute the relevant thermophysical quantities and feed it to the flow simulator. In this context, two approaches were developed to facilitate this coupling and make it more feasible. At the first level, a set of various early rejection schemes are presented to accelerate the MC simulations and reduce the computational time needed for generating the offline table. At the second level, and in order to fill the gaps within the table, extrapolation schemes through reweighting and reconstruction of the MC Markov chains are introduced. At this stage, all these methods have been developed and tested for single-site Lennard-Jones (LJ) particles. Hence, a set of optimized single-site LJ parameters is recommended by accelerating MC simulations via polynomial chaos technique. In

the following sections of this chapter, each proposed method and its corresponding results are thoroughly discussed.

2.1 Early Rejection Schemes

One way of accelerating MC molecular simulations is by modifying the conventional Metropolis algorithm [9] itself. For this purpose, several techniques have been introduced [10–14]. In particular, early rejection schemes serve as a typical example for such techniques. Early rejection methods are capable of saving part of the computational cost by reaching the rejection decision for the undesired or "doomed" MC trials at earlier stages than the conventional algorithm. However, it is important for such algorithms to preserve the accuracy while reducing the required simulation time. In literature, the bond formation early rejection scheme [15] exists at which an MC trial is rejected whenever two particles get too close to each other, forming a hypothetical "bond". To overcome some of the drawbacks of this method, we have first introduced the conservative early rejection scheme. In that work [16], we have shown that the bond formation method is inapplicable for some MC trials and it does not produce the same Markov chain as the conventional method. Although the conservative method solves these problems, it is still, in general, slower than the bond formation scheme. We therefore further introduced a hybrid scheme, for Gibbs ensemble simulations, in which the bond formation technique is used for MC trials where it is successful, and the conservative early rejection scheme in the moves where the bond formation technique fails.

To elaborate more on this point, the conservative early rejection scheme is based on the fact that the Lennard-Jones (LJ) model, which is used to evaluate potential energy between interacting particles, has a global minimum value (u_{min}). While looping over all the particles in the system to evaluate changes in energy due to an

MC trial performed, the unevaluated interactions are assumed to be equal to u_{min}. In other words, all the particles waiting for their contribution to be evaluated are assumed to be at that exact distance corresponding to u_{min} away from the displaced particle. Throughout the loop, the rejection condition is checked at every step and whenever it is true, the loop is terminated and the move is directly rejected. Based on this assumption, the produced Markov chain is guaranteed to be exactly the same as the one produced by the conventional algorithm.

Later on, we have proposed the non-conservative early rejection scheme at which only few particles (N_{min}) are assumed to be at that exact distance corresponding to u_{min}. In fact, this assumption, as will be shown later, is more realistic. However, results showed slight modifications in the generated Markov chains, hence the method is called "non-conservative". Nonetheless, these modifications are too small to cause major changes in the final outputs such as canonical averages, while the simulation times were significantly reduced.

In this section a large number of numerical simulations are presented in order to compare outputs from all the different methods mentioned. In all scenarios, the non-conservative method was much faster than the conservative scheme. It also performed comparable to, and sometimes faster than, the bond formation scheme. In comparison to the conventional algorithm, the non-conservative early rejection scheme has reduced the simulation time by 60 % for both canonical and NVT-Gibbs ensembles. These simulation time reductions were shown to be dependent on simulation thermodynamic conditions, temperature and density in specific.[1]

[1]The content of this section appears in Molecular Physics [16] and Molecular Simulation [17] journals.

2.1.1 Simulation Algorithms

2.1.1.1 MC conventional algorithm

At first, the canonical ensemble is considered for introducing the conventional algorithm [9]. In canonical ensemble, a group of N particles is simulated in an imaginary box with volume V at a certain temperature T, thus it is often referred to as "NVT ensemble". Usually NVT ensemble is used to study single-phase behavior where the system's pressure (p) is the main output. For such systems, performing elemental changes to the original configuration creates new ones. For instance, when simulating structureless particles the only possible elemental change in NVT ensemble is displacing a randomly chosen particle. The criterion below (Equ. 2.1), used to decide whether the new configuration is to be granted or not, is dependent on the energy difference between the new configuration and the previous one,

$$\exp\left(-\beta\Delta U\right) \geq R. \tag{2.1}$$

Here, β is the Boltzmann's factor, ΔU is the change in system's energy and R is a randomly generated number between 0 and 1. If the relation in Equ. 2.1 is true then the MC trial is accepted and a new configuration is formed, otherwise the trial is rejected and the previous configuration is retrieved. As a matter of fact, the change in the system's energy due to the particle displacement is related to the change in the energy associated with the displaced particle, say the i^{th} particle, itself. Therefore, Equ. 2.1 can be rewritten as below:

$$\exp\left[-\beta\left(\sum_{j=1,j\neq i}^{N} u_{ij}^{n} - \sum_{j=1,j\neq i}^{N} u_{ij}^{o}\right)\right] \geq R, \tag{2.2}$$

where u_{ij}^{n} accounts for the energy associated with the displaced i^{th} particle at the new position due to its interaction with the j^{th} one, while u_{ij}^{o} stands for the interaction

energy between the same particles (i and j) but prior to particle i displacement. Therefore, the total energy associated with the displaced i^{th} particle can be computed by summing all the interactions with the rest $N-1$ particles in the system. By applying the logarithmic operation to both sides followed by few rearrangements, the condition in Equ. 2.2 can be reformulated as follows:

$$\sum_{j=1,j\neq i}^{N} u_{ij}^{n} \leq \sum_{j=1,j\neq i}^{N} u_{ij}^{o} - \frac{\ln(R)}{\beta}. \tag{2.3}$$

Finally, Equ. 2.3 will be used as the acceptance condition in canonical ensemble. In addition to the canonical ensemble, this paper deals with the Gibbs ensemble. In particular, the NVT-Gibbs ensemble [18–21] at which two simulation boxes are simulated each representing a different phase. Simulations in NVT-Gibbs ensemble are executed under constant total number of particles of both phases ($N = N_1 + N_2$), total volumes ($V = V_1 + V_2$) and T. In order to reach equilibrium, at least three different types of MC trials are required. These trials are particle displacement within a certain phase to insure thermal equilibrium, particle transfer from one phase to the other to achieve chemical equilibrium and finally volume changes to attain mechanical equilibrium. For more details about the Gibbs ensemble and the conditions used to decide on accepting new configurations, reference [8] is recommended. Similar rearrangements to the previous canonical ensemble equations have been done for the conditions here to take the following forms for particle displacement and transfer, respectively:

$$\sum_{j=1,j\neq i}^{N_p} u_{ij}^{n} \leq \sum_{j=1,j\neq i}^{N_p} u_{ij}^{o} - \frac{\ln(R)}{\beta}, \tag{2.4}$$

$$\sum_{j=1,j\neq i}^{N_{in}} u_{ij}^{n} \leq \sum_{j=1,j\neq i}^{N_{out}} u_{ij}^{o} - \frac{\ln\left[\frac{R\,(N_{in}+1)\,V_{out}}{N_{out}V_{in}}\right]}{\beta} + (N_{out}-1)\,E_{out}^{tail} - (N_{in}+1)\,E_{in}^{tail}. \tag{2.5}$$

In order to explain the equations above, it is important to highlight that the interface between gas phase and liquid phase does not exist in Gibbs ensemble. This implies that the particles in the two simulation boxes of the Gibbs ensemble (one for liquid and the other for gas phase) only interact with the other particles residing the same box. Therefore whenever a particle is displaced, a loop over the particles sharing same box (N_p) is only required as appears in Equ. 2.4. In the case of particle transfer from one box to another, subscripts "*out*" and "*in*" refer to the box the particle was taken out from and the box it was introduced into, respectively. N and V refer to the number of particles and volume before the transfer has taken place. We note that, particle transfer trial leads to change in number of particles in both boxes, hence tail energy correction functions (E^{tail}) need to be re-evaluated.

In summary, at every MC step in either canonical or Gibbs ensemble the relevant relation out of the Equations (Equs. 2.3, 2.4, or 2.5) is used to determine whether the introduced elemental change is to be accepted or not. These relations are arranged such that, the right-hand side is known from the previous configuration in addition to the prescribed simulation thermodynamic conditions. On the other side, the left part of these inequalities is computed by looping over the N or N_p particles to evaluate the energy associated with the particle i at its new position. After the loop is over, the relevant relation is tested to determine whether to accept or reject the trial.

2.1.1.2 Bond formation algorithm

Apparently, based on the conventional scheme, one needs to loop over all the particles in order to reach a decision whether to accept or reject the move. However, if it was possible to reach this decision earlier, a considerable amount of simulation time might be saved.

The first type of such methods to be discussed in this work is the bond formation early rejection one. The bond formation early rejection method is introduced based on the work of Wang and Swendsen [15]. In this scheme, an MC trial is rejected whenever a "bond" is formed between the displaced molecule and any of the other existing molecules in the simulated system. This means, MC trials are only accepted if no "bond" is developed between any couple of molecules due to displacement. The criterion used to decide whether a "bond" is formed or not is called the bond formation probability (P_{bond}). It is a function of the energy difference due to the interaction between the displaced molecule, say the i^{th} molecule, at its new and old position with respect to another j^{th} molecule. If the energy difference was negative then no "bond" is formed; otherwise, the bond formation probability is computed as follows

$$P_{bond} = 1 - \exp\left(-\beta \Delta u_{ij}\right). \tag{2.6}$$

Therefore, whenever a molecule i is moved, the difference in energy due to its interaction with another molecule j at new position versus the old one ($\Delta u_{ij} = u_{ij}^{n} - u_{ij}^{o}$) is used to compute P_{bond} (Equ. 2.6). As stated earlier, if the energy difference was negative, bonds could not form and the loop would continue to check with the rest of the molecules. As soon as a positive change is detected, P_{bond} is evaluated and compared against a random number ($0 < R < 1$). Whenever P_{bond} exceeds the R, the move is immediately rejected assuming that the two molecules have got too close to each other.

This method significantly accelerates the MC molecular simulation, as will be shown in the results section; however it has two main drawbacks. First, the bond formation concept requires the knowledge of the interaction between the investigated molecules before implementing the MC trial. Unfortunately, such knowledge is not always available. For instance, in Gibbs ensemble the molecule transfer from one box to another is an essential MC move. In such transfer, the molecule subjected to this trial interacts with completely two different sets of molecules. The first group belongs to the box it was originally residing, and the other belongs to the one it was transferred to. As a result, the bond formation concept of comparing energies due to interactions with same molecule at old and new position fails. The second drawback is an expected consequence of using a different algorithm than the Metropolis one [9]. The bond formation scheme suggests a new criterion for accepting/rejecting MC trials; thus it generates a Markov chain different from the conventional one. We note that, however, results show that such modifications are, at most of the times, tolerable.

2.1.1.3 Conservative early rejection scheme

The conservative early rejection scheme [16] was developed as an attempt to accelerate MC molecular simulations while overcoming the drawbacks of the bond formation scheme. Hence, the scheme is designed to conserve the same Markov chain as the conventional Metropolis algorithm and, more importantly, to be applicable for the different types of MC trials in the various existing ensembles.

This scheme is based on the fact that the energy due to interaction between any two Lennard-Jones (LJ) sites has a global minimum (Fig. 2.1). The following paragraphs are devoted to illustrate the details of this method. A convenient starting point for this discussion is to highlight the LJ model. It is a widely used model to compute the energy due to dispersion-repulsion (short range) interaction between a

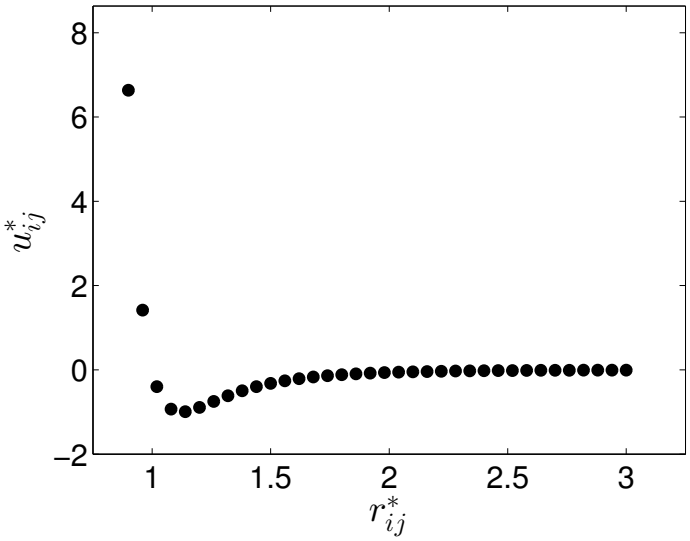

Figure 2.1: Normalized potential energy (u^*_{ij}) due to the interaction between a couple of particles computed using LJ model (Equ. 2.20) as a function of the normalized separating distance (r^*_{ij}).

pair of particles

$$u_{ij} = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{2.7}$$

In Equ. 2.19, u_{ij} stands for potential energy due to dispersion-repulsion interaction between the i^{th} and j^{th} particles, and r_{ij} is their separating distance. ε is the potential well depth and σ is the "hard sphere" radius. Assuming all the simulated particles are identical, u_{ij} and r_{ij} can be normalized with respect to ε and σ such that, $u^*_{ij} = u_{ij}/\varepsilon$ and $r^*_{ij} = r_{ij}/\sigma$ and this leads to the dimensionless form of the LJ model

$$u^*_{ij} = 4 \left[\left(\frac{1}{r^*_{ij}} \right)^{12} - \left(\frac{1}{r^*_{ij}} \right)^{6} \right]. \tag{2.8}$$

Fig. 2.1 displays the profile of u^*_{ij} with respect to r^*_{ij}. It can be easily noticed that u^*_{ij} has a global minimum value of -1. It is straightforward to mathematically prove that, by finding r^*_{ij} at which the first derivative of u^*_{ij} vanishes. Doing so, it is found that $r^*_{ij} = 2^{1/6}$ corresponds to the minimum energy $u^*_{ij}\,(2^{1/6}) = -1$. To sum up, the fact that the interaction energy between any two molecules has a global minimum is

the most important factor to reach an early rejection decision as will be explained.

Now by recalling that Equ. 2.3 represents the acceptance condition for the translational MC trial move, then failing to satisfy that relation leads to the trial rejection. Thus, the move would be rejected if, and only if, this inequality was true

$$\sum_{j=1,j\neq i}^{N} u_{ij}^{n} > \sum_{j=1,j\neq i}^{N} u_{ij}^{o} - \frac{\ln(R)}{\beta}. \tag{2.9}$$

At this point, and whenever a particle is displaced all the terms in the rejection condition (Equ. 2.9) are known except the energy associated with the particle at its new position. In the conventional algorithm, this term is evaluated by looping over all the $(N-1)$ particles existing in the simulated system. As the loop is completed, Equ. 2.9 is tested and the decision to whether accept or reject the trial is reached. On the other side, in the conservative scheme still the same criterion in Equ. 2.9 is used, however the inequality is tested at every step of the loop instead of checking it only once after the loop is finished. Meanwhile, the interactions at the new position are separated into two terms. The first term collects the evaluated interactions so far, and the second accounts for the unevaluated interactions yet to come. To guarantee that the proposed algorithm is reaching exactly to the same decision taken by the conventional algorithm, the unevaluated interactions are assumed to have the value of the smallest energy possible between any two LJ particles denoted by u_{min}. Thus, the condition takes the following form:

$$\sum_{j=1,j\neq i}^{k} u_{ij}^{n} + u_{min}(N-k) > \sum_{j=1,j\neq i}^{N} u_{ij}^{o} - \frac{\ln(R)}{\beta}. \tag{2.10}$$

We point that, k stands for the number of the evaluated interactions, so far, in the loop. By the same logic, the acceptance conditions used for particle displacement and transfer in the Gibbs ensemble can be modified into the conservative scheme rejection

conditions, respectively:

$$\sum_{j=1,j\neq i}^{k} u_{ij}^{n} + u_{min}(N_p - k) > \sum_{j=1,j\neq i}^{N_p} u_{ij}^{o} - \frac{\ln(R)}{\beta}, \tag{2.11}$$

$$\begin{aligned}\sum_{j=1,j\neq i}^{k} u_{ij}^{n} + u_{min}(N_{in} - k) > \sum_{j=1,j\neq i}^{N_{out}} u_{ij}^{o} - \\ \frac{\ln\left[\frac{R\,(N_{in}+1)\,V_{out}}{N_{out}V_{in}}\right]}{\beta} + (N_{out}-1)\,E_{out}^{tail} - (N_{in}+1)\,E_{in}^{tail}.\end{aligned} \tag{2.12}$$

To summarise this section, both the conventional and the conservative scheme use exactly the same criterion to check whether a move is accepted or not. However, each of them reaches the decision based on a different workflow. The conventional scheme computes all the interactions before checking the acceptance condition, while the conservative early rejection scheme checks the condition at every step of the loop assuming that the unevaluated interactions contribute with the specific minimum energy possible u_{min}. Based on this assumption, it is guaranteed that whenever the rejection relation is satisfied there is no way that it can be broken as none of the unchecked particles could contribute with energy less than u_{min}. As a consequence, the decision of rejection is often reached even before the loop is completed; hence the simulation time is reduced.

2.1.1.4 Non-conservative early rejection scheme

As has been stated earlier, assuming that all the un-investigated particles reside at the exact distance from the displaced particle i where their mutual energy is at its minimum (u_{min}) is non-realistic. Nonetheless, for serving the purpose of the conservative scheme this assumption is essential to guarantee that the exactly same

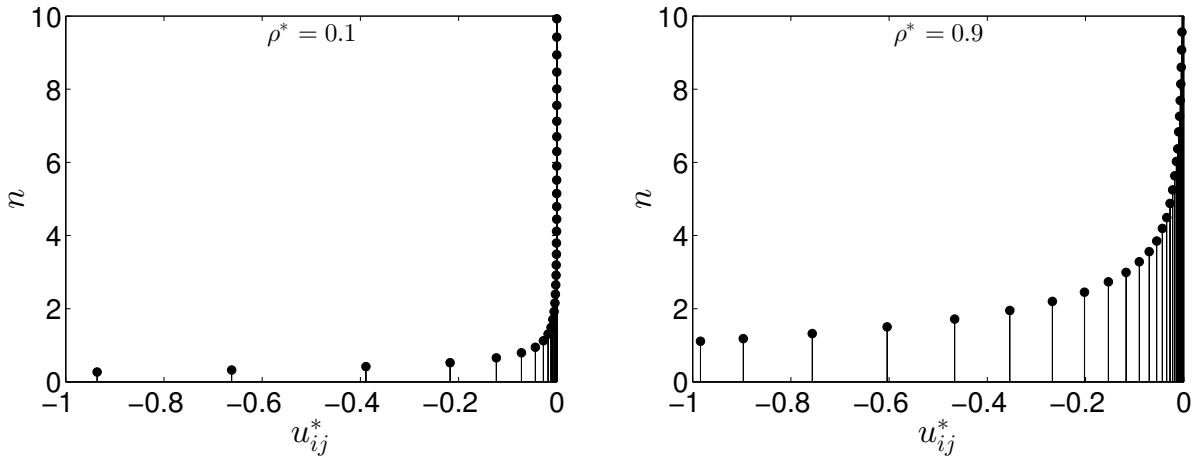

Figure 2.2: Distribution of number of particles per MC step (n) contributing with u^*_{ij} energy towards a randomly displaced particle i. These distributions correspond to a system of $N = 512$ particles at $T^* = 10$ with $\rho^* = 0.1$ (Left) and 0.9 (Right).

Markov chain as the one generated by the conventional algorithm is produced. This argument suggests that it is possible to violate that criterion by assuming that only few particles, instead of all, are actually residing at that specific distance. This statement is the milestone in developing the non-conservative early rejection scheme.

As an attempt to understand how the simulated particles interact, Fig. 2.2 is created for systems at normalized $T^* = 10$ with $N = 512$ particles at normalized number densities $\rho^* = 0.1$ and 0.9, simultaneously. To explain this figure, it is important to know that the particles in the system can interact in three ways towards the particle i. Most of the particles are located at a distance further than the normalized cutoff radius ($r^*_c = r_c/\sigma = 0.49L^*$, where $L^* = L/\sigma$ is the normalized simulation box side length) from i at which the interaction is truncated; this leads to an interaction of $u^*_{ij} = 0$. The second type of particles contribute with $u^*_{ij} > 0$. These particles are usually few and contribute with large values of u^*_{ij} as they are located at the vicinity of the particle i. For convenience and while creating the distributions, these contributions are counted as $u^*_{ij} = 0$ instead of their actual values. Finally, the rest of the particles contribute with $u^*_{ij} < 0$. Therefore, to create these distributions, at each MC step whenever a randomly chosen particle i is displaced, the distances separating

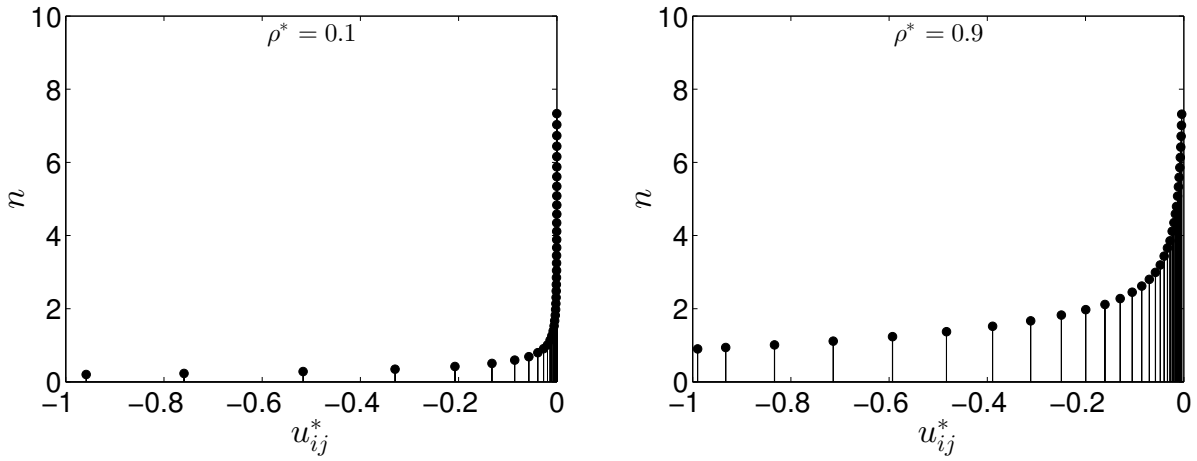

Figure 2.3: Same as Fig. 2.2 but with $N = 256$.

i from other particles are computed and collected in a histogram with binning value of $\Delta r/L = 0.01$. This process is repeated for every MC step. At the end, the counted particles are divided by the total number of MC steps to get the values of "n". Then the distances are transferred into normalized energy values (u^*_{ij}) using the LJ model in Equ. 2.20.

These generated distributions show a clear dependence on the system's density. As ρ^* increases, more particles start to contribute with negative u^*_{ij}. This is due to the fact that for dense systems particles get closer and less truncated interactions are found, hence a shift towards negative contributions is seen. In both cases ($\rho^* =$ 0.1 and 0.9), it is obvious that none of the particles in the system is contributing with the exact normalized minimum energy ($u^*_{min} = -1$) and even very few are contributing with values lower than -0.8. This agrees with the argument that the conservative scheme is indeed "too conservative". A final note on this discussion, simulations at different temperatures than $T^* = 10$ lead to same results with slight, almost unnoticeable, changes that can be safely neglected.

To take this analysis a step further, similar distributions were collected but at two different N values 256 (Fig. 2.3) and 1024 (Fig. 2.4). According to these figures, the distributions are also dependent on N and not merely ρ^*. To have a collective

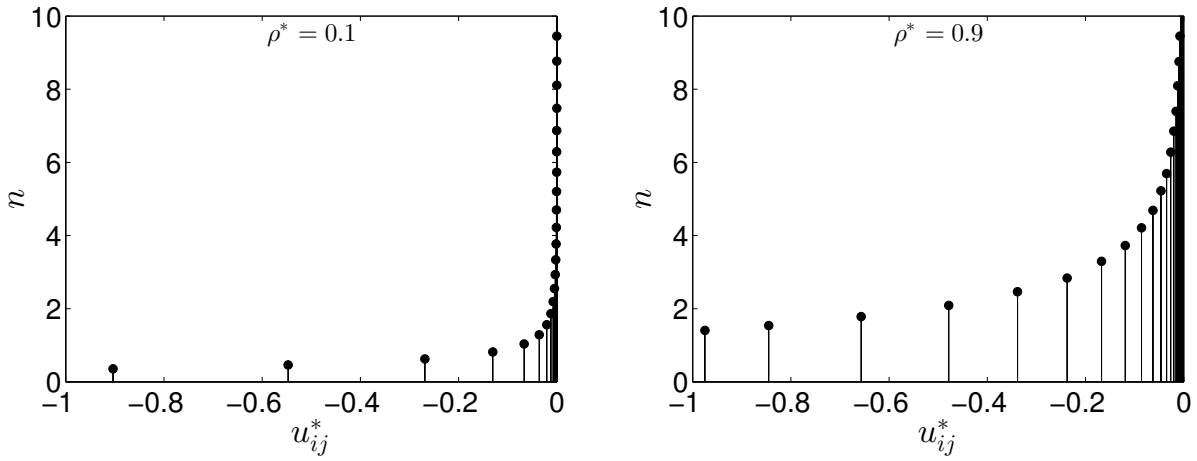

Figure 2.4: Same as Fig. 2.2 but with $N = 1024$.

perception of all the different distributions, the average contribution energy ($\overline{u}^*_{ij}$) of a particle j in the system towards the displaced particle i is computed and plotted against different ρ^* values corresponding to systems made of different number of particles (Fig. 2.5). Based on this figure, $\overline{u}^*_{ij}$ is linearly dependent on ρ^* while slope changes as N does.

So far, the above interpretation highlights the dependence of $\overline{u}^*_{ij}$ on both ρ^* and N. Nonetheless, this relation could be better represented by a single variable; which is the normalized volume ($V^* = N/\rho^*$). By transforming all the three lines with respect to the new suggested variable V^*, interestingly, all the lines collapse into a single one (Fig. 2.6). In particular, the following linear relation with respect to $1/V^*$ is reached:

$$\overline{u}^*_{ij} = -10.6/V^*. \tag{2.13}$$

Based on these findings, the rejection condition in Equ. 2.10 is now modified into Equ. 2.14 assuming that only N_{min} particles out of the $(N - k)$ unchecked particles in the system contribute with u_{min} while the rest of the particles contribute with the

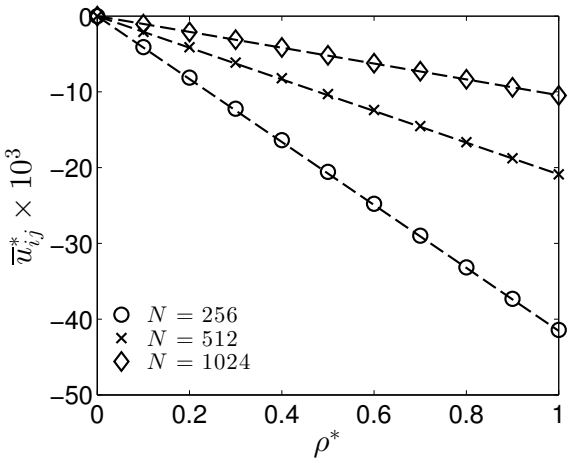

Figure 2.5: The normalized average energy ($\overline{u}^*_{ij}$) contribution by any particle j towards the randomly displaced particle i as a function of system's normalized number density (ρ^*). Plots of three different systems are shown, each at a different number of particles ($N = 256$, 512 and 1024). The dashed lines are added to emphasize the linear relation between $\overline{u}^*_{ij}$ and ρ^*.

average energy $\overline{u}_{ij} = \varepsilon \overline{u}^*_{ij}$ that can be calculated using Equ. 2.13

$$\sum_{j=1,j\neq i}^{k} u^n_{ij} + \overline{u}_{ij}(N - k - N_{min}) + N_{min}u_{min} > \sum_{j=1,j\neq i}^{N} u^o_{ij} - \frac{\ln(R)}{\beta}. \tag{2.14}$$

To properly choose how many particles are assigned to N_{min}, the both sides of Equ. 2.13 are multiplied by N:

$$N\overline{u}^*_{ij} = -10.6\rho^*, \tag{2.15}$$

the left hand side of this equation ($N\overline{u}^*_{ij}$) can be used to estimate the total contribution of all the existent particles in the system towards i per configuration. In the paper's examples, ρ^* values range from 0.1 to 1. By plugging the maximum value of density ($\rho^* = 1$) in Equ. 2.15, the estimated total energy contribution per MC step towards i is $N\overline{u}^*_{ij} = -10.61$. Therefore, we choose to fix 5 particles at $u^*_{min} = -1$ such that $N_{min}u^*_{min} = -5$ (which is around the half of the maximum energy expected -10.61 at the highest $\rho^* = 1$). For all the cases studied in this paper in both single and two-

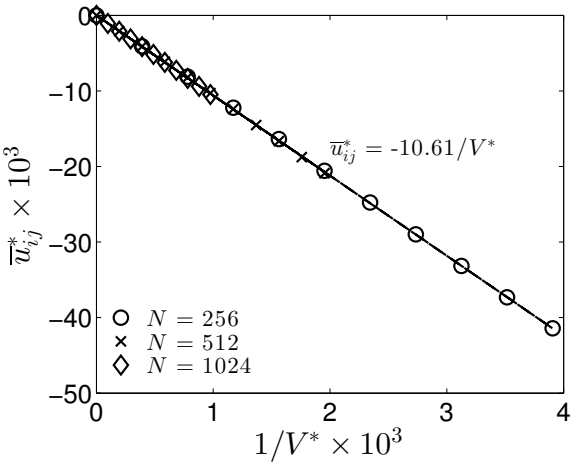

Figure 2.6: Same as Fig. 2.5 but this time showing the relation between $\overline{u}^*_{ij}$ and the inverse of the system's normalized volume ($1/V^*$). The dashed line represents the best linear fit, so that $\overline{u}^*_{ij}$ and $1/V^*$ are related with Equ. 2.13.

phase analysis $N_{min} = 5$ is used. Finally, it is worth to address that it is not necessary for N_{min} to be 5, other values are also possible when simulating other systems or even for this set of simulations. Therefore, N_{min} could be optimized based on the system to be studied.

In similar fashion, the rejection conditions in the Gibbs ensemble for the particle displacement and particle transfer are modified to have the forms in Equs. 2.16 and 2.17, respectively. We note that, phase volumes are allowed to change during the Gibbs simulation. As a result, at each time these volumes are changed new $\overline{u}_{ij}$ values need to be computed using Equ. 2.13.

$$\sum_{j=1,j\neq i}^{k} u^n_{ij} + \overline{u}_{ij}(N_p - k - N_{min}) + N_{min}u_{min} > \sum_{j=1,j\neq i}^{N_p} u^o_{ij} - \frac{\ln(R)}{\beta}, \tag{2.16}$$

$$\sum_{j=1,j\neq i}^{k} u_{ij}^{n} + \overline{u}_{ij}(N_{in} - k - N_{min}) + N_{min}u_{min} > \sum_{j=1,j\neq i}^{N_{out}} u_{ij}^{o} - \frac{\ln\left[\frac{R\,(N_{in}+1)\,V_{out}}{N_{out}V_{in}}\right]}{\beta} + (N_{out}-1)\,E_{out}^{tail} - (N_{in}+1)\,E_{in}^{tail}. \tag{2.17}$$

To sum up, the key difference between the non-conservative and the conservative early rejection schemes is by assuming only N_{min} particles contribute with u_{min} energy towards the displaced particle i instead of all the rest $(N-1)$ particles. The generated distributions helped in deciding on how many of these particles were close to u_{min}. Actually, only few of them were observed to be close at the whole spectrum of thermodynamic conditions investigated. As only N_{min} particles are assumed to contribute with u_{min}, the rest un-investigated $(N - k - N_{min})$ particles are assigned to the average contribution energy $\overline{u}_{ij}$ that can be computed from Equ. 2.13.

2.1.2 Results and Discussion

2.1.2.1 Canonical ensemble results

In the first group of experiments, the speeding up introduced by the proposed non-conservative early rejection method is compared with all of the other three methods discussed in the "Simulation Methods" section. In this part, the comparison is conducted in the single-phase region simulated in canonical ensemble. Numerical experiments are designed to show the dependence of the performance of all the four methods at various simulation conditions. For that purpose, simulations were run for three different number of particles $N = 256$, 512 and 1024. For each of these systems, different thermodynamic conditions corresponding to $\beta^* = 1/T^* = 0.1$ and 0.7 with

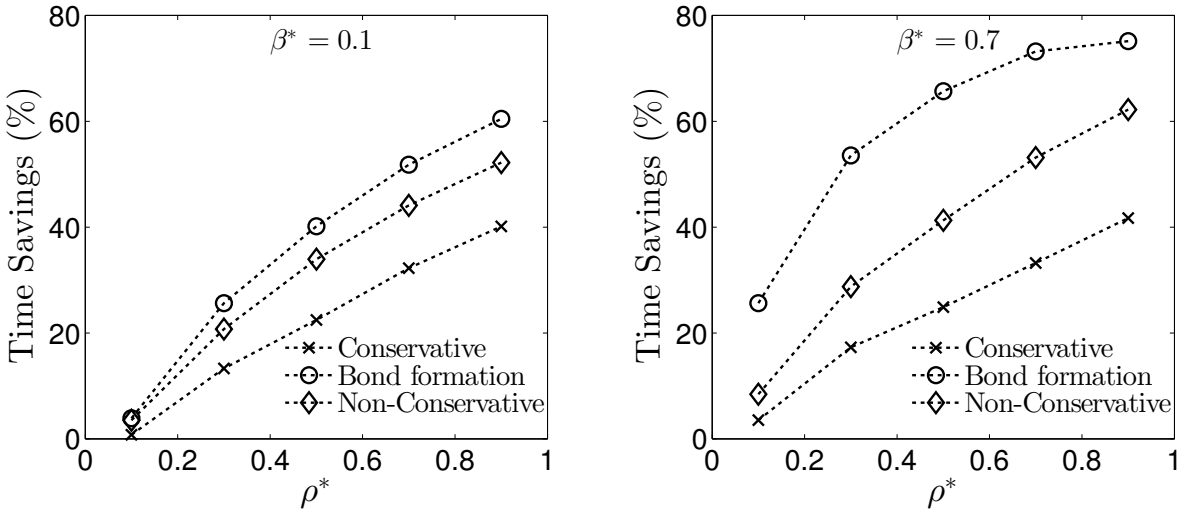

Figure 2.7: Simulation time savings by each of the bond formation and conservative early rejection schemes in addition to the proposed non-conservative method. Results in this figure correspond to a system with $N = 256$ at $\beta^* = 0.1$ (Left) and 0.7 (Right) in canonical ensemble.

$\rho^* = 0.1$, 0.3, 0.5, 0.7 and 0.9 were used. We note that, all the numerical experiments conducted start from the exact same initial configuration, at which particles are uniformly distributed in the simulation box.

For instance with $N = 256$ at $\beta^* = 0.1$, five simulations per method were conducted at the different five ρ^* values. According to Fig. 2.7 (Left), the time saving by each of the three early rejection methods increases as ρ^* increases. We note that, the time saving is calculated by the following formula:

$$t_s(\%) = \frac{t_0 - t_1}{t_0} \times 100. \tag{2.18}$$

In this formula, t_s stands for the percentage of the time saved, t_0 is the time needed by the conventional algorithm, and t_1 is the simulation time taken by each of the three early rejection schemes. In general, the bond formation scheme is the fastest of all at the different densities used. It achieves a time saving of more than 60 % at $\rho^* = 0.9$, while the non-conservative and the conservative achieve 50 % and 40 % time saving, respectively. The same numerical experiments were repeated but this

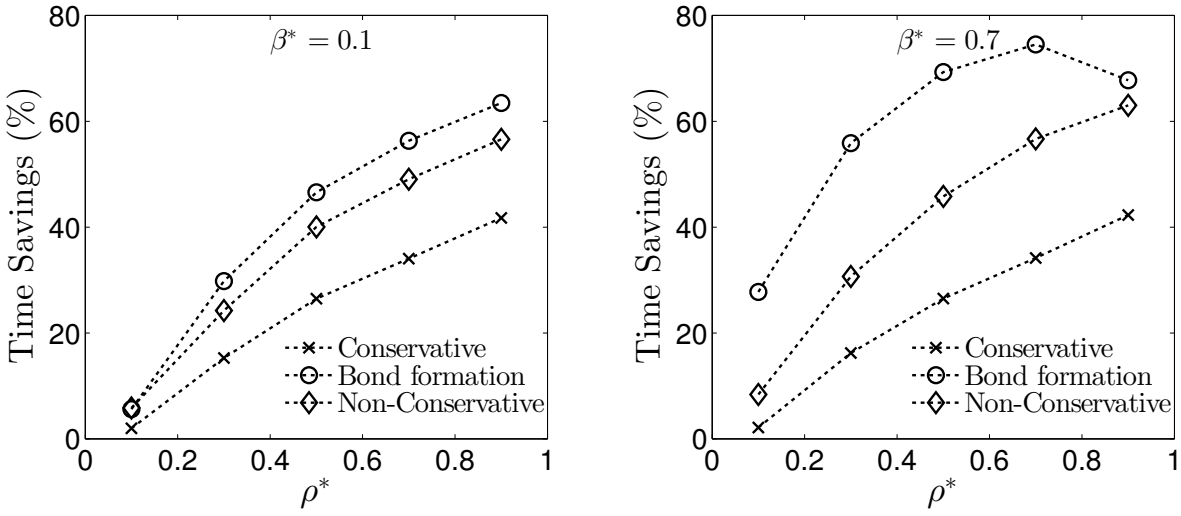

Figure 2.8: Same as Fig. 2.7 but for $N = 512$.

time at $\beta^* = 0.7$ (Fig. 2.7, Right). The results are consistent in both groups with the realization that, in general, all the three methods performed better at higher β^* especially the bond formation method. Further analysis is done by repeating the same experiments above, but this time for $N = 512$ (Fig. 2.8) and $N = 1024$ (Fig. 2.9). In comparison to the previous set (Fig. 2.7), the same conclusions can be reached.

To complete the simulation time analysis, the simulation times saved by each of the early rejection methods are plotted in Fig. 2.10 against MC trials rejection rates. According to this chart, the simulation time saved increases as the rejection rates of the performed MC trials get higher. The highest time savings achieved were at rejection rates close to 1. For example, the bond formation early rejection scheme can save up to around 75 % of the original simulation time while the non-conservative and the conservative ones save up to around 65 and 45 %, respectively.

In addition to the rejection rates, it is also important to check the dependence of the presented method on the cutoff radius (r_c). As mentioned earlier, $r_c^* = 0.49L^*$ was assigned for all the conducted simulations in this section. Knowing that, L^* was changing from one case to another, so was the value of r_c^*. In order to study the effect of r_c^* on the simulation times, a group of five simulations were run for each of the

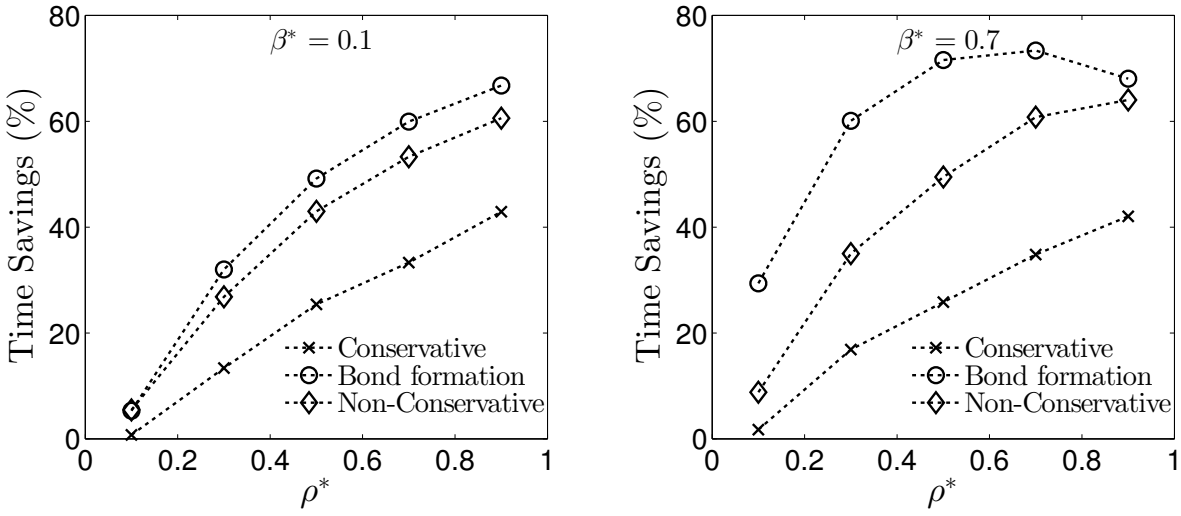

Figure 2.9: Same as Fig. 2.7 but for $N = 1024$.

methods discussed earlier. All these simulations shared the same $\beta^* = 0.7$, $N = 256$ and $r_c^* = 2.2$ (The minimum r_c^* that could be used is in the range of 2.2 to 2.5 as can be seen in Fig. 2.1). By comparing the plots in Fig. 2.7 (Right) and Fig. 2.11 (Left), it is clear that the proposed non-conservative method in addition to the other two early rejection schemes have maintained their speed with slight decrease at low ρ^* values. Moreover and in agreement with the results reported in Fig. 2.10, Fig. 2.11 (Right) emphasizes that the speeding up is primarily dependent on the MC trials rejection rates.

Finally, all the early rejection algorithms used to obtain the simulation results of this part are validated against the conventional algorithm outputs. As expected (see Table B.2 in Appendix B), the normalized pressure obtained by all the early rejection methods is in good agreement with the conventional algorithm values. The statistical uncertainties reported in Table B.2 in Appendix B are estimated using the bootstrap method [22, 23].

To summarize the results obtained in the single phase section, the non-conservative early rejection method was shown to be faster than the conservative scheme, and almost matching the speed of the bond formation one. Moreover, it introduces less

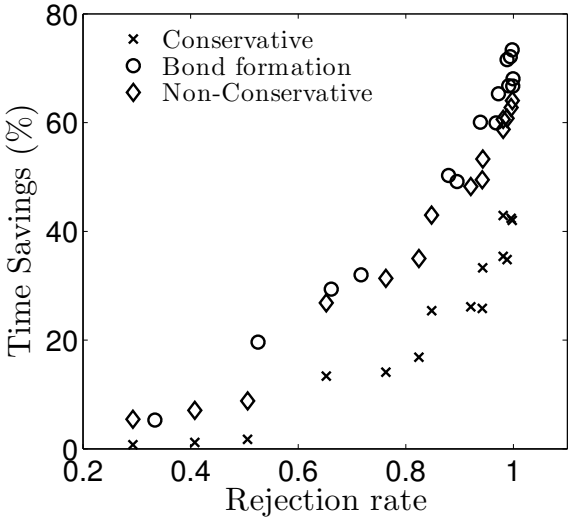

Figure 2.10: Time savings (%) by each of the three early rejection methods studied in the canonical ensemble against the rejection rates of the MC trials performed.

modification to the original Markov chain when compared to the bond formation method. In general, the key factor in determining the significance of the speeding is the rejection rate of the MC trials performed.

2.1.2.2 Gibbs ensemble results

In an earlier paper [16], a hybrid early rejection method was proposed in addition to the conservative one for generating the Gibbs ensemble results. The hybrid early rejection scheme uses the conservative scheme for the particle transfer MC trial while uses the bond formation scheme for the particle displacement one. The reason for proposing such a scheme was to take advantage of the speed of the bond formation method in particle displacement. Therefore, in this paper the non-conservative scheme simulation times are analyzed in comparison to both the conservative and the hybrid schemes.

For that purpose, a total of nine numerical experiments were designed each at a different normalized temperature ranging from $T^* = 0.8$ up to 1.2, with a 0.05 temperature increment. Fig. 2.12 plots all the simulation times required by the conventional

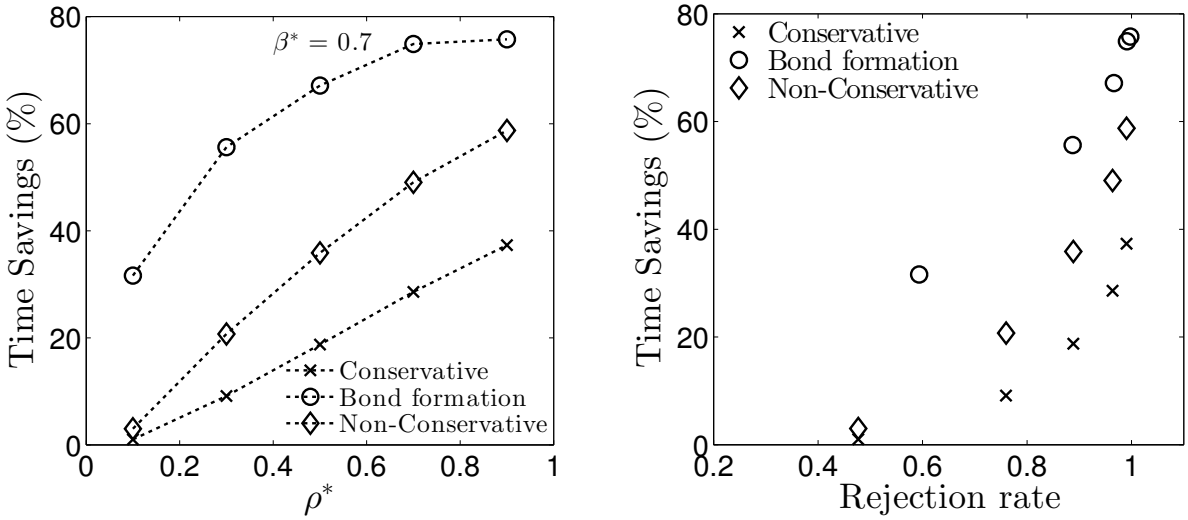

Figure 2.11: Simulation time savings by each of the three early rejection methods at $r_c = 2.3\sigma$ (Left). Relation between these time savings and the MC trials rejection rates (Right).

algorithm in addition to each of the early rejection methods; namely the conservative, hybrid and non-conservative, for all the different T^* values. By examining Fig. 2.12, it is clear that the non-conservative method is the fastest of all regardless of the simulation thermal condition. In general, as the system is set closer to the critical temperature more simulation time is required except for the conventional method. This behavior is a well-known phenomenon, and relates to the way the particles are split into the two phases.

To highlight the significance of the reduction in the simulation time, Fig. 2.13 is plotted. Along the whole range of temperatures used, the non-conservative scheme was successful on saving between 30 and 65 % of the time needed by the conventional method. More interestingly, this range is higher than the maximum reduction achieved by the conservative scheme, occurring at the lowest temperature ($T^* = 0.8$) used in the two-phase analysis.

Another important perspective to look at, is the dependence of simulation time reduction with respect to the average rejection rate of the two MC trials implemented (particle displacement and transfer). According to Fig. 2.14 and similar to previous

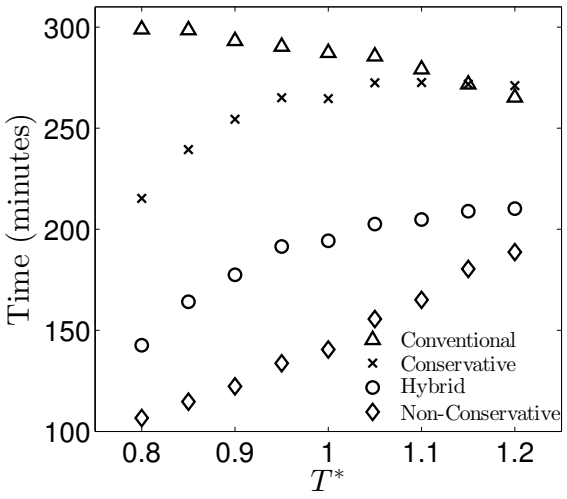

Figure 2.12: Simulation times required by each of the bond formation and conservative early rejection schemes in addition to the proposed non-conservative one, by 1 Intel Xeon processor on Dell PC. Results correspond to systems with T^* ranging from 0.8 to 1.2 in Gibbs ensemble.

realizations, all the early rejection schemes perform better at higher rejection rates. Moreover, this figure reveals a crucial difference between the non-conservative and the bond formation approaches. By comparing the rejection rates reached by each of the three methods, the rejection rates achieved by the conservative and the non-conservative are completely consistent for all data point sets along the same temperature. In contrary, the hybrid scheme, due to bond formation method involvement, reaches higher rejection rates in most of the cases (This can be clearly seen for simulations at $T^* = 1.2$). This result is in agreement with the results reported earlier in Fig. 2.10. This implies that, part of the simulation time saved by the bond formation early rejection scheme is due to the fact that more MC trials are rejected than usual. In other words, the bond formation scheme accelerates MC molecular simulations by not only rejecting "doomed" trials at an earlier stage, but also by rejecting more successful trials than normal. Of course, this is another advantage for the proposed non-conservative early rejection scheme over the bond formation one.

Finally and in order to validate the results obtained by all of the algorithms used for the early rejection schemes in this section, Table B.3 in Appendix B reports the

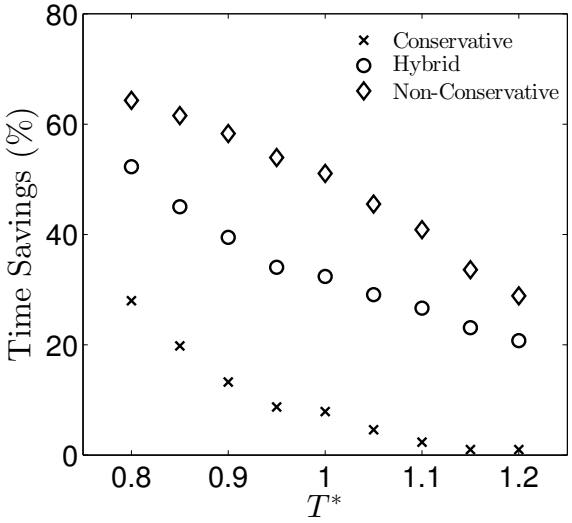

Figure 2.13: Simulation time savings by each of the bond formation and conservative early rejection schemes in addition to the proposed non-conservative scheme at different T^* in Gibbs ensemble.

normalized number density values corresponding to the vapor (ρ_v^*) and liquid (ρ_l^*) phases computed from the different simulations. For all the early rejection methods presented, the results obtained are in good agreement with the values obtained from the conventional algorithm. Again here the bootstrap method [22, 23] was used for statistical errors estimation.

2.1.3 Conclusions

To conclude, in this study two early rejection schemes have been introduced the conservative early rejection scheme and the non-conservative one. The proposed non-conservative method succeeded in outperforming the conservative one by saving up to around 60 % of simulation times needed by conventional algorithm in both canonical and Gibbs ensembles. Moreover, it overcame the main problem with the bond formation method by being applicable to all kinds of MC trials, while performing with almost the same of its speed and even faster in some occasions. Therefore, the non-conservative early rejection scheme combines the strength elements of each of the other two methods.

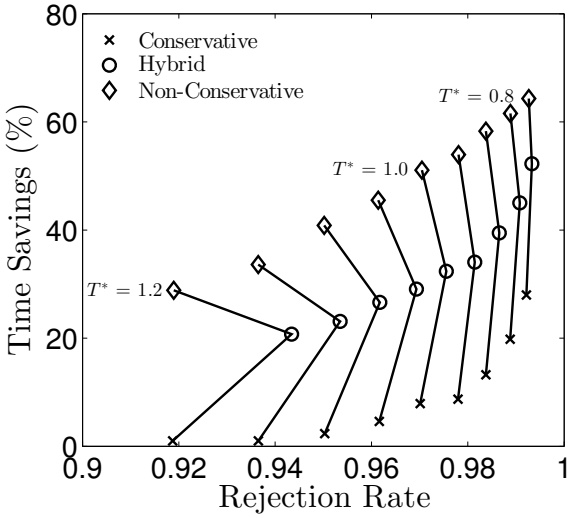

Figure 2.14: Time savings (%) by each of the three early rejection methods studied in the Gibbs ensemble against the average rejection rates of the particle displacement and transfer trials performed. Solid lines connect simulations sharing the same normalized temperature (T^*).

At the end, these methods has a great potential to elevate the performance of MC molecular simulations by using the non-conservative early rejection scheme for more complicated systems and applications such as simulating single polymer chain or systems with multiple ones. The fact that the performance of such schemes is enhanced by the increase in rejection rates creates a great motivation to test this technique in investigating such systems. Keeping in mind that the extension and implementation of the proposed algorithm is programmer-friendly and straightforward.

2.2 Polynomial Chaos MC Simulations

In this part, the time consuming MC molecular simulations were replaced by fast surrogate models created via Polynomial Chaos (PC) expansions. PC expansions provide an approximate representation of the solution of the forward problem, which can be used to reduce the computational cost of repetitive model evaluations. PC works by representing the model variables and parameters in terms of a spectral expansion in

an orthogonal polynomial basis according to their probabilistic distributions. The resulting cost effective surrogate models can then be used to efficiently reproduce the solutions and the statistical properties of Quantities of Interest (QoIs) [24–27]. The PC-based model provides a complete probabilistic representation of the outputs in terms of the random inputs. PC may suffer from 'the curse of dimensionality' [28] that limits the application of PC to only a moderate number of stochastic parameters for surrogate model construction. Nonetheless, PC methods have become one of the standard approaches for solving stochastic problems, to propagate and quantify uncertainties in various disciplines including physical [29–31], chemical, and geophysical systems [32–35].

So far, only few studies have applied PC expansions to molecular simulations. In fact, those studies have worked with Molecular Dynamics (MD) instead of MC molecular simulations. For instance, Rizzi et al. have demonstrated a successful coupling between PC and MD in studying force field parameters of water molecules [36,37] in addition to concentration driven ionic flow in nano-pores [38,39]. Later on, similar coupling is adopted in order to investigate flow at nanoscale [40] and quantify parametric uncertainty in multi-scale simulations [41].

In this study, a novel combination between PC expansions and MC molecular simulations is proposed. First, two independent PC surrogate models were constructed to replace the canonical and the NVT-Gibbs ensembles, respectively. These surrogates are capable of accurately reproducing the normalized supercritical isotherms and the two-phase envelope of the Lennard-Jones (LJ) fluid. The surrogate models estimations match the MC results within a margin of $\pm$ 3% deviation in the whole thermodynamic range studied, while requiring a tiny fraction of the computational time needed by the MC molecular simulations. Later on, a large-scale optimization for force field parameters of different molecules is conducted via the pre-constructed PC surrogates. The optimization outcome leads to a recommended set of single-site

LJ models for a group of molecules. An excellent fit is achieved for argon (Ar), krypton (Kr), xenon(Xe), methane (CH_4), nitrogen (N_2), and carbon monoxide (CO). Nonetheless, and not surprisingly, other bigger and/or polar molecules show larger misfits due to the simplicity of the LJ model used.[2]

2.2.1 Simulation Methods

2.2.1.1 MC molecular simulation

The essence of MC molecular simulation is to predict macroscopic properties of fluids by simulating the interactions among a large group of molecules (statistical ensemble) representing the fluid. Each statistical ensemble is characterized by the different physical constraints imposed in order to capture the thermo-physical conditions at which these fluids exist. Fortunately, nowadays a wide range of ensembles are available to be used in MC molecular simulations. The decision of choosing the suitable ensemble to work with mainly relays on the simulation output needed. For instance, in this work, pressures under supercritical conditions and two-phase saturated densities were needed to search for adequate LJ parameters for different molecules. Therefore, the canonical (NVT) ensemble was used to predict single-phase pressures while the NVT-Gibbs ensemble was adopted to compute the two-phase saturated densities.

In the canonical ensemble, a certain number of molecules (N) is trapped in a fixed hypothetical volume (V), leading to a constant system density ($\rho = N/V$) throughout the simulation. In addition, the system's temperature (T) is fixed as well. As a result, the fluid's pressure (P), under the prescribed ρ and T, can be predicted. The NVT-Gibbs ensemble [18–21] is used to study phase equilibria between two phases.Two separate simulation boxes, each containing a certain number of molecules, are created to present each of these phases. Similar to the canonical ensemble, the temperature is kept fixed throughout the simulation. In order to reach

[2]The content of this section is published in Journal of Chemical Physics [42].

equilibrium between the two phases, molecules are randomly transferred between the two boxes (chemical equilibrium) while the individual volumes of each box are allowed to change (mechanical equilibrium) such that the total volume is conserved. Finally thermal equilibrium is reached by performing translational and configurational MC trials on individual molecules. For more details about the Gibbs ensemble and the conditions used to decide on accepting new configurations, see reference [8].

The statistical nature of the MC molecular simulation arises from the random MC trials performed on the simulated ensembles. MC trials (e.g. molecules displacement, molecules transfer between simulation boxes, and boxes volume change) introduce elemental changes to the system to create a large number of different molecular configurations. These configurations are sampled to construct the so-called MC Markov chain. The decision of accepting these elemental changes or rejecting them while creating the new configurations is related to the interactions among the simulated molecules. Thus, a key component of any MC molecular simulation is the potential model describing molecular interactions.

In this study, the LJ model was used to represent the molecular interactions, in specific, the van der Waals long-range attractive force, and the repulsive force resulting from the Pauli exclusion principle that prevents the collapse of molecules. The LJ model relates the energy between any pair of particles to their separating distance as follows,

$$u_{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{2.19}$$

where u_{ij} stands for potential energy due to dispersion-repulsion interaction between particle i and particle j with a separating distance r_{ij}. The LJ parameters ε_{ij} and σ_{ij} are the potential well depth and the "hard sphere" radius, respectively. This work focuses only on pure component single-phase and two-phase systems, thus all particles are identical and the LJ parameters can be reduced to ε and σ. For convenience, both u_{ij} and r_{ij} are normalized by ε and σ such that, $u^*_{ij} = u_{ij}/\varepsilon$ and $r^*_{ij} = r_{ij}/\sigma$, leading

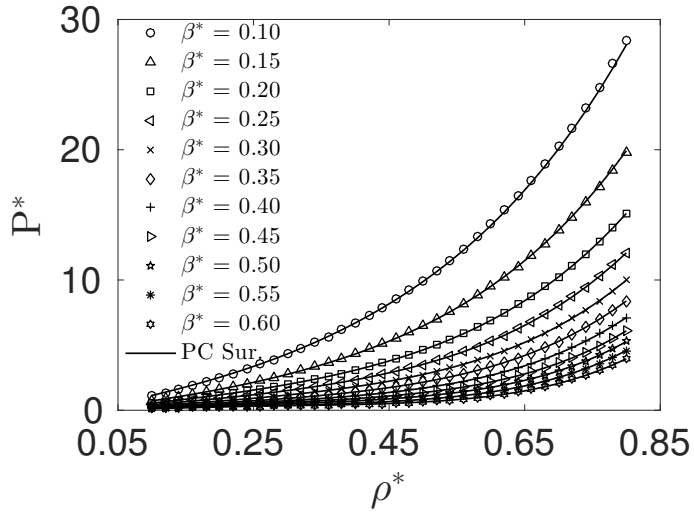

Figure 2.15: Normalized LJ fluid isotherms at supercritical thermodynamic conditions ($0.1 \leq \beta^* \leq 0.6$) generated by MC canonical ensemble simulations (symbols) in comparison with the results obtained by the pre-constructed single-phase PC surrogate model (lines). For this case, a polynomial of degree 6 was used with 64 total quadrature nodes.

to the dimensionless LJ model form:

$$u_{ij}^* = 4\left[\left(\frac{1}{r_{ij}^*}\right)^{12} - \left(\frac{1}{r_{ij}^*}\right)^{6}\right]. \tag{2.20}$$

2.2.1.2 The polynomial chaos framework

The general framework of PC technique works by treating input variables uncertainties, parameters, and model state as stochastic variables. Hence, the system state can be represented by a random vector, $x(t)$, and its time evolution is expressed through the following differential equation:

$$\dot{\mathbf{x}}(t, \boldsymbol{\Theta}) = \mathbf{f}(t, \boldsymbol{\Theta}, \mathbf{x}, \mathbf{u}), \quad \mathbf{x}(t_0) = \mathbf{x}_0. \tag{2.21}$$

In the equation above, $\boldsymbol{\Theta}$ represents the uncertain parameter vector and $\mathbf{u}$ the assumed deterministic forcing terms. The initial state estimate, $\mathbf{x}_0$, may also be uncertain but here it is assumed known. The polynomial chaos method enables

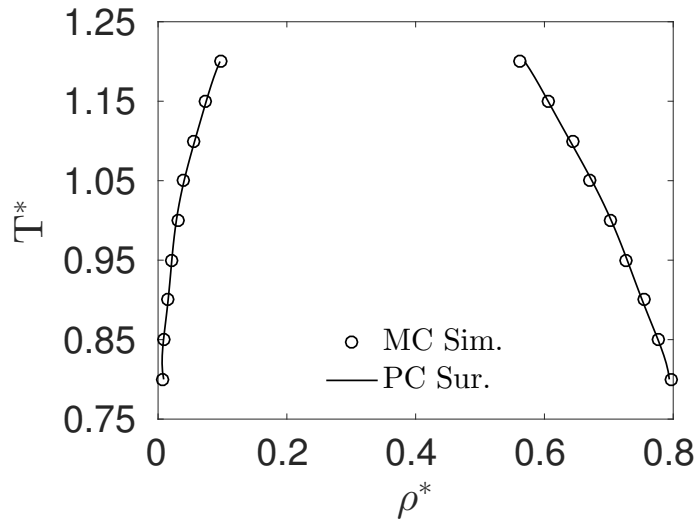

Figure 2.16: Normalized LJ fluid two-phase envelope at subcritical thermodynamic conditions ($0.8 \leq T^* \leq 1.2$) generated by NVT-Gibbs ensemble simulations (circles) in comparison with the results obtained by the pre-constructed two-phase PC surrogate model (lines). For this case, a polynomial of degree 6 and 20 quadrature nodes were used.

to compute the full pdf characterizing the space-time evolution of x_k, allowing to quantify the uncertainty of the system outputs with respect to the parameters.

Recently, new uncertainties propagation methods have been developed based on the polynomial chaos theory. Generalized Polynomial Chaos (gPC) is an extension of the homogeneous chaos idea of Wiener [43]. The idea is based on separating the random variables from deterministic ones while solving the stochastic differential equation. Then the random variables are expanded using a suitable polynomial expansion.

The gPC approach assumes that the model outputs $\mathbf{X} \in L^2(\Omega^*)$ to admit a spectral expansion of the form

$$\mathbf{X} = \sum_{k=0}^{\infty} c_k \Psi_k \doteq \sum_{k=0}^{P} c_k \Psi_k(\boldsymbol{\xi}), \tag{2.22}$$

where $\xi_i \overset{iid}{\sim}$ are canonical random variables used to parameterize random inputs, Ψ_k is the d-variate Legendre polynomials and c_k represents the PC coefficients. The

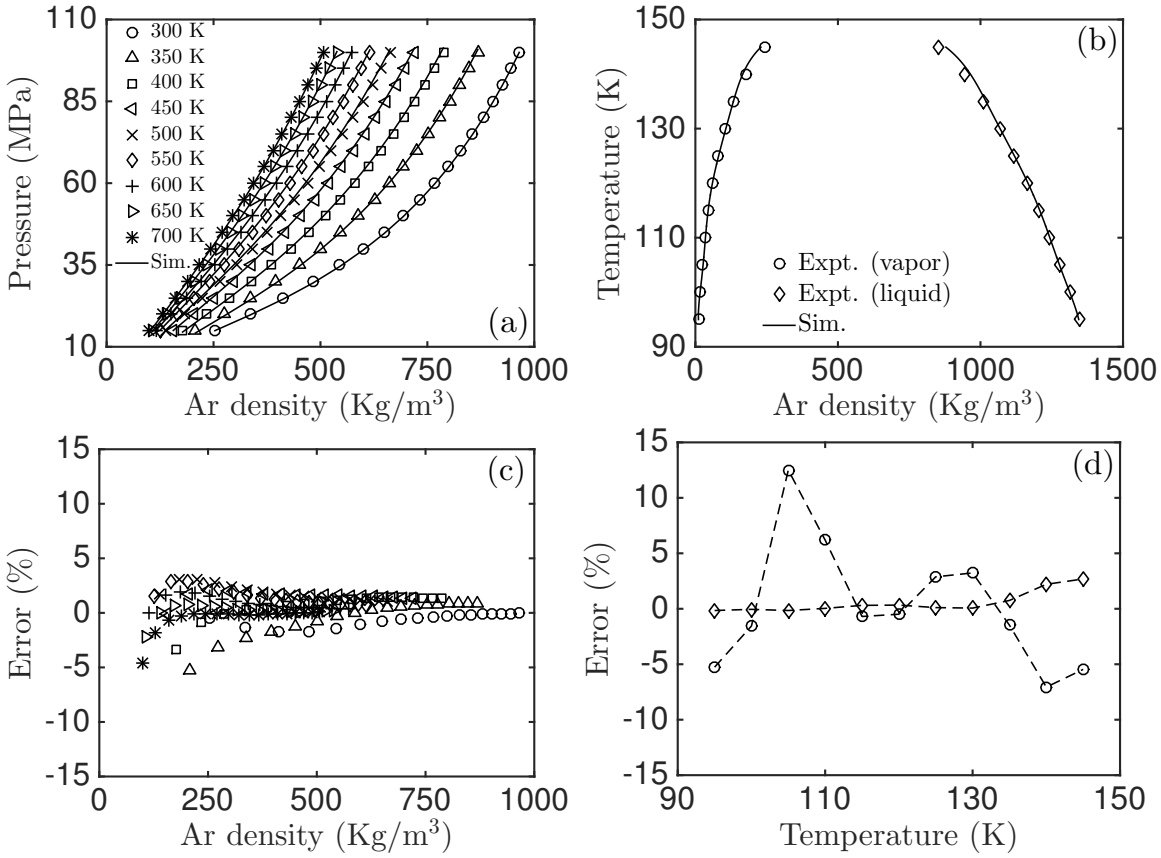

Figure 2.17: (a) Argon's experimental pressure values along supercritical isotherms ranging from 300 K to 700 K (symbols) fitted by the PC surrogate model of the MC canonical ensemble simulations (lines). (b) Argon's two-phase envelope (circles and diamonds represent experimental saturated vapor and liquid densities, respectively) fitted by the PC surrogate model of the NVT-Gibbs ensemble simulations (lines). (c) Deviation percentages from the true experimental pressure values along the fitted supercritical isotherms. (d) Deviation percentages from the true experimental saturated density in the two-phase region; dashed lines are added for eye guidance.

order of truncation P_{tr} depends on the stochastic dimension d and expansion order, p, as

$$P_{tr} = \frac{(d+p)!}{d!p!} - 1. \tag{2.23}$$

Since $\{\Psi_k\}_0^{P_{tr}}$ forms an orthogonal system, we can write the following dot product

$$(X, \Psi_k) = c_k(\Psi_k, \Psi_k), \tag{2.24}$$

with

$$c_k = \frac{\langle X, \Psi_k \rangle}{\langle \Psi_k^2 \rangle}. \tag{2.25}$$

In Equ. 2.25, the moments $\langle \Psi_k^2 \rangle$ of the multivariate Legendre polynomials can be analytically evaluated [30] while $\langle X, \Psi_k \rangle$ requires more elaborate computation, as we note that

$$\langle X, \Psi_k \rangle = \int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_\xi(\mathbf{s}), \quad k = 0, ..., P_{tr}.$$

Hence, evaluating c_k involves computing the values of a set of $P_{tr}+1$ integrals over $\Omega^* \subseteq \mathbb{R}^d$, which can be discretized as finite sums using an appropriate quadrature formula:

$$\int_{\Omega^*} X(\mathbf{s}) \Psi_k(\mathbf{s}) dF_\xi(\mathbf{s}) \doteq \sum_{j=1}^{N_q} w_j X(\boldsymbol{\xi}_j) \Psi_k(\boldsymbol{\xi}_j). \tag{2.26}$$

In the equation above, $\boldsymbol{\xi}_j \in \Omega^*$ and w_j are the nodes and weights of the quadrature rule. In this approach, the main computational burden essentially amounts to the deterministic evaluation of the model over the sample set of the quadrature nodes. Let $\mathbf{\Pi} \in \mathbb{R}^{(P_{tr}+1) \times N_q}$ be the NISP matrix,

$$\mathbf{\Pi}_{k,j} = \frac{w_j \Psi_k(\boldsymbol{\xi}_j)}{\langle \Psi_k^2 \rangle}, \quad k = 0, ..., P_{tr}, \quad j = 1, ..., N_q$$

also let $\boldsymbol{\zeta}$ be the vector with coordinates $\zeta_j = X(\boldsymbol{\xi}_j)$. Then the vector $\mathbf{c}$ can be expressed as $\mathbf{\Pi}\boldsymbol{\zeta}$, or in component form:

$$c_k = \sum_{j=1}^{N_q} \mathbf{\Pi}_{kj} \zeta_j = \sum_{j=1}^{N_q} \mathbf{\Pi}_{kj} X(\boldsymbol{\xi_j}), \quad k = 0, ..., P_{tr}. \tag{2.27}$$

With a fully tensorized quadrature rule, the complexity of NISP scales with $N_q = n^d$, the number of nodes n and number of uncertain parameters d. Hence, this approach is computationally limited to a moderate number of parameters. For detailed mathematical formulation of the NISP method, the reader is referred to [28].

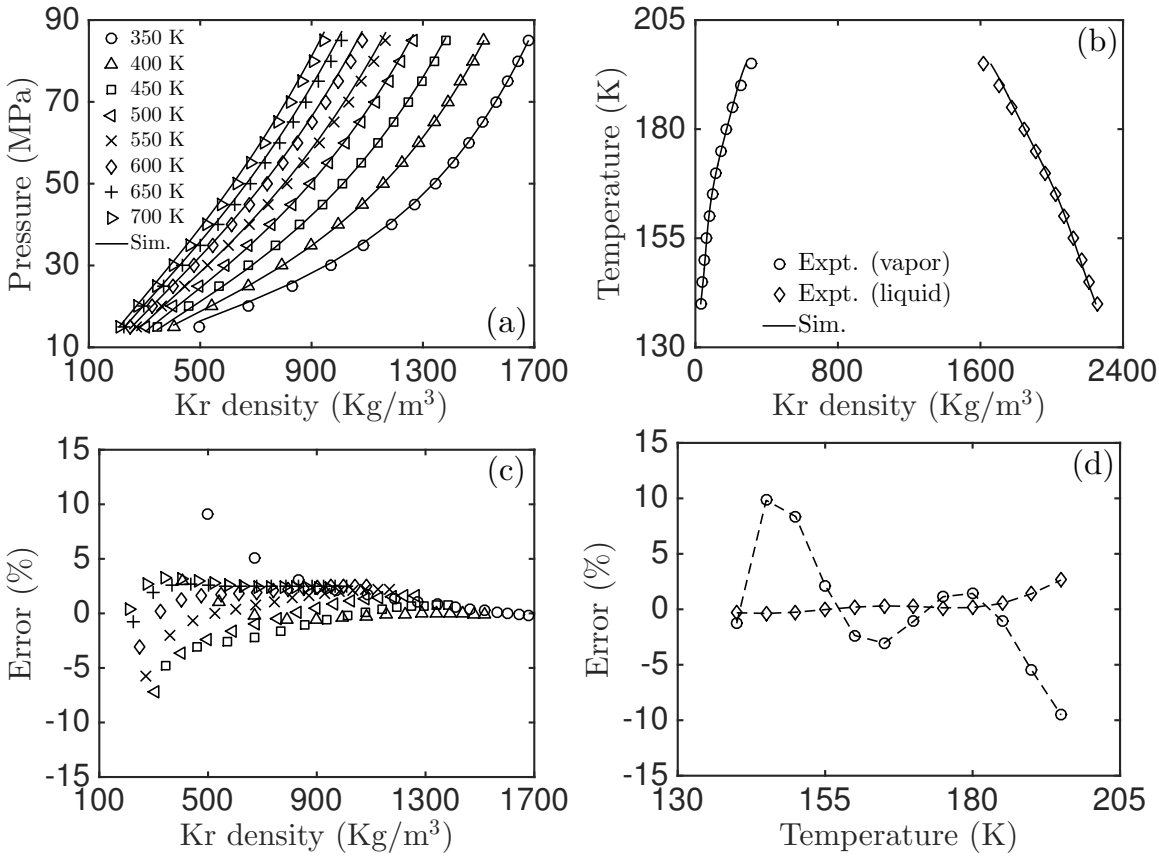

Figure 2.18: Same as Fig. 2.17 but for krypton with supercritical isotherms ranging from 350 K to 700 K.

2.2.2 Results and Discussion

2.2.2.1 Surrogate models construction

The construction of the surrogate models through PC expansions, which are then used to produce the MC simulation results, was conducted as follows. At first, let $\mathbf{m} = (m_1, m_2, ..., m_n)^{\mathrm{T}}$ be the vector of random model inputs having uniform distribution. Specifically, the inputs m_i were parameterized by ξ_i through

$$m_i(\boldsymbol{\xi}) = \mu_i + \sigma_i \xi_i, \quad i = 1, 2, ..., n \tag{2.28}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_n)^{\mathrm{T}}$, $\mu_i = (1/2)(a_i + b_i)$, and $\sigma_i = (1/2)(b_i - a_i)$. For a given vector of random inputs $\mathbf{m}(\boldsymbol{\xi})$ the output is denoted by $\mathbf{X}(\boldsymbol{\xi}) = \mathbf{A}(\mathbf{m}(\boldsymbol{\xi}))$.

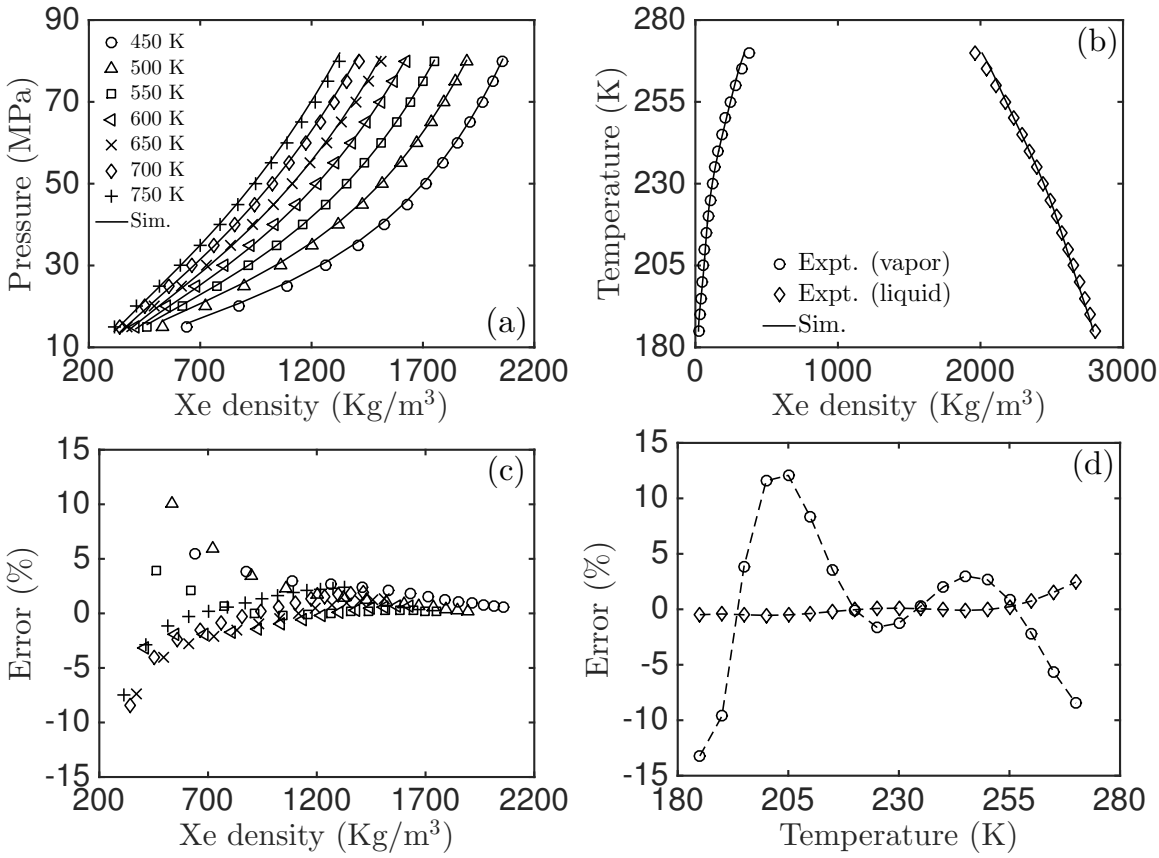

Figure 2.19: Same as Fig. 2.17 but for xenon with supercritical isotherms ranging from 450 K to 750 K.

Note that $\mathbf{A}(\mathbf{m}(\boldsymbol{\xi}))$ is the output of the MC simulation solved with the set of the input parameters $\mathbf{m}(\boldsymbol{\xi})$. $\mathbf{X}$ may correspond to any subset of the QoIs. In the single-phase case, $\mathbf{X}$ corresponds to the normalized equilibrium pressure (P^*), while in the two-phase scenario it corresponds to the normalized liquid- (ρ^*_{liq}) and vapor-phase saturation density (ρ^*_{vap}). The stochastic variables in the single-phase simulation are both the normalized Boltzmann's factor (β^*) and system number density (ρ^*). In the two-phase simulation, β^* is the only stochastic variable.

For this study, two independent surrogate models were generated to replace single-phase (canonical ensemble) and two-phase (NVT-Gibbs ensemble) MC simulations, simultaneously. MC simulations in canonical ensemble were run using 216 particles with 1×10^6 equilibrium MC steps, followed by 5×10^6 production MC steps. On the

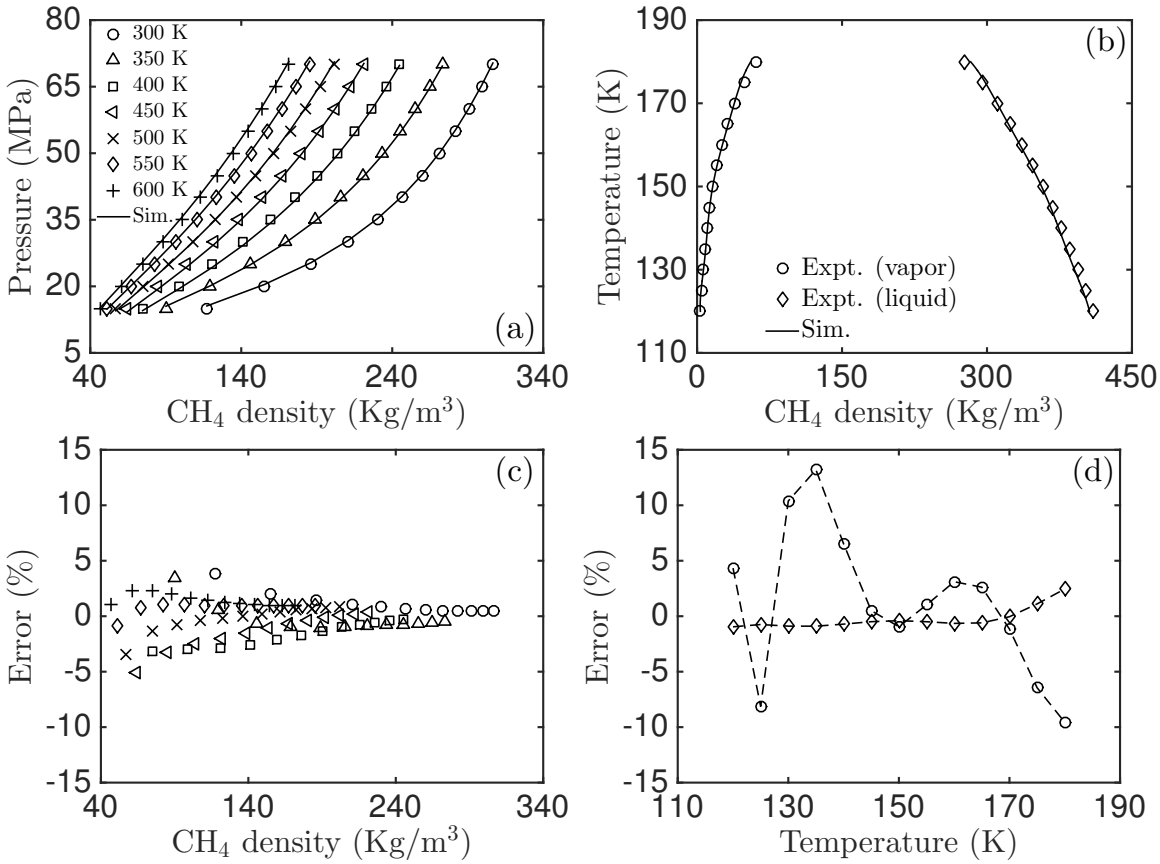

Figure 2.20: Same as Fig. 2.17 but for methane with supercritical isotherms ranging from 300 K to 600 K.

other hand, a total of 686 particles (343 particles uniformly distributed in each box at the initial configuration) were simulated in NVT-Gibbs ensemble with 10×10^6 and 20×10^6 equilibrium and production steps, respectively. In both ensembles, a cutoff radius $r_c = 0.49L$ was imposed while energy and pressure tail correction functions were introduced to account for the truncated interactions beyond r_c. Regarding the computational cost, each MC canonical ensemble run lasts for around 20 minutes using an Intel Xeon processor on a Dell workstation. On the same machine, a single NVT-Gibbs ensemble simulation requires around 150 minutes of computational time.

In the single-phase scenario, the Gaussian quadrature formula was employed to build the NISP matrix. A total of only 64 quadrature points (8 in each input parameters direction β^* and ρ^*) was used in order to construct the surrogate model that can

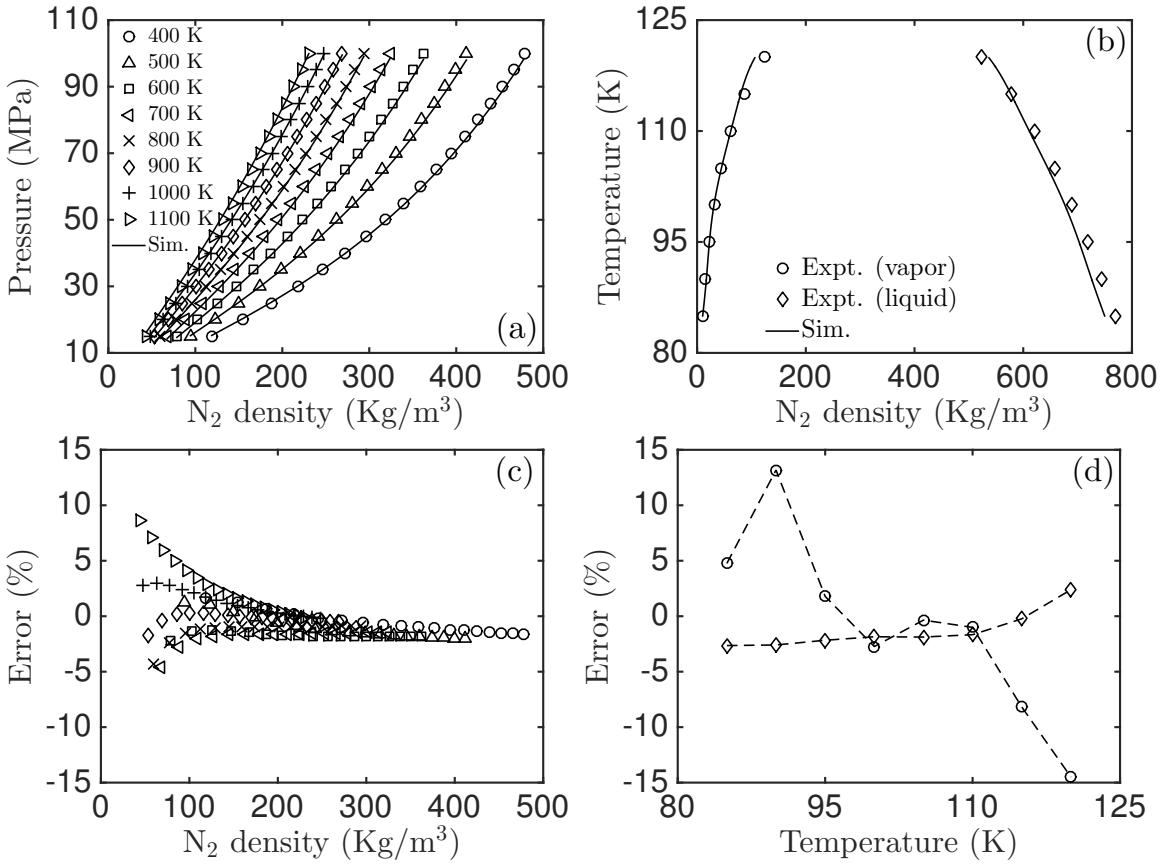

Figure 2.21: Same as Fig. 2.17 but for nitrogen with supercritical isotherms ranging from 400 K to 1100 K.

reproduce the MC canonical ensemble results over a wide range of simulation conditions. According to this, the time needed to construct the surrogate model is 64 times the time needed by a single forward run of the canonical ensemble. However once the surrogate model is constructed, any output can be evaluated in few seconds by running the surrogate model itself. Similarly, only 20 quadrature points were needed to generate the surrogate model that can produce the NVT-Gibbs ensemble results within seconds.

According to Fig. 2.15, the results obtained by the single-phase PC surrogate model are in excellent agreement with the P^* values directly calculated from the MC canonical ensemble simulations. In fact, this surrogate model can reproduce the MC canonical results within $\pm$ 3% error along the whole range of $0.1 \leq \beta^* \leq 0.6$ and

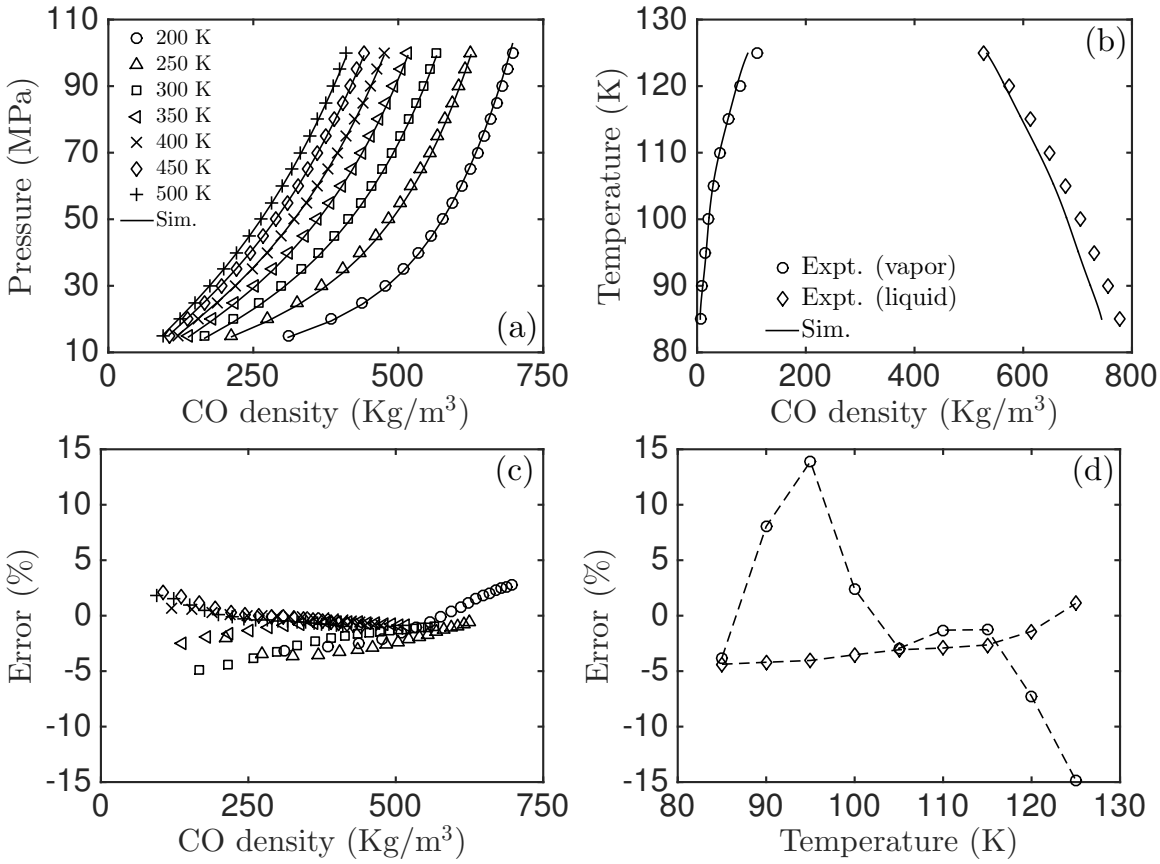

Figure 2.22: Same as Fig. 2.17 but for carbon monoxide with supercritical isotherms ranging from 200 K to 500 K.

$0.1 \leq \rho^* \leq 0.8$. Similarly, the two-phase surrogate model can accurately estimate both ρ^*_{liq} and ρ^*_{vap} generated by NVT-Gibbs ensemble (Fig. 2.16) in the normalized temperature range of $0.8 \leq T^* \leq 1.1$, with a relative error of $\pm$ 2%.

2.2.2.2 Single-site LJ parameters optimization

As a direct application to the coupling presented in this article between PC and MC methods, a large-scale optimization for single-site LJ parameters was conducted for several small molecules. The recommended parameters (ε_{rec}, σ_{rec}) were obtained by minimizing the difference between various predictions from the surrogate against their corresponding experimental measurements. All the experimental data used here were retrieved from NIST database [44].

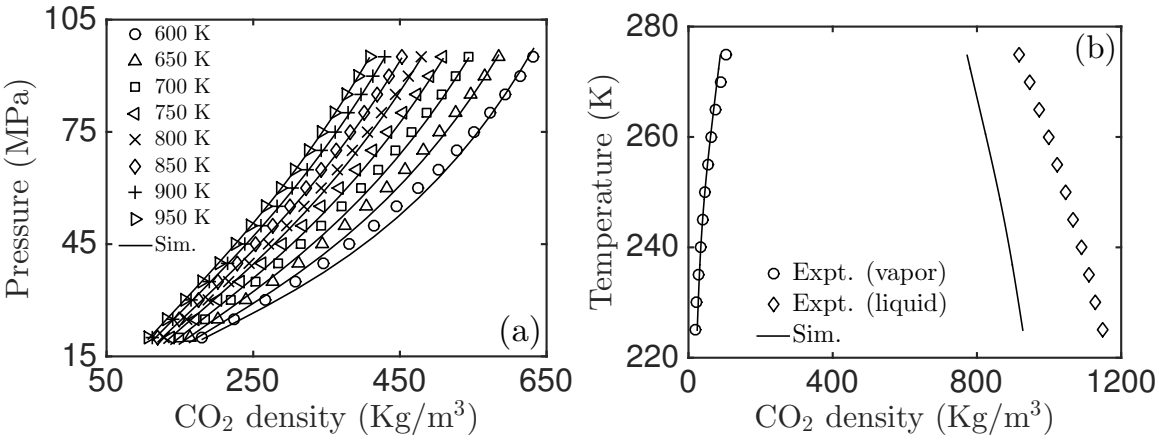

Figure 2.23: (a) Supercritical isotherms of carbon dioxide data (symbols) ranging from $T = 600$ K to $T = 950$ K fitted by the single-phase PC surrogate model (lines). (b) Large deviation in two-phase region between experimental data (symbols) in comparison with the PC model fit (lines). In contrary to the previous molecules, representing CO_2 molecule by merely an LJ sphere is not sound.

For the starting point of the optimization, a physically valid guess is needed for both ε and σ. For this purpose, formulas obtained by fitting virial coefficient data were used [45]. These equations are dependent on the critical properties of the component itself:

$$\varepsilon_0 = \frac{T_c k_B}{1.35} \text{ and } {\sigma_0}^3 = \frac{0.35M}{\rho_c N_A}, \tag{2.29}$$

where k_B is the Boltzmann's constant, N_A is the Avogadro's number, and M is the molecular weight, whereas T_c and ρ_c are the critical temperature and mass density, respectively.

The optimization process was performed using the MATLAB built-in function "fmincon", starting with the initial guesses $(\varepsilon_0, \sigma_0)$ from Equ. 2.29. During the fitting process, the tuning parameters were subjected to certain lower and upper limits such that they can only attain values satisfying these constraints. Typical ranges used were $\pm$ 10 K for ε/k_B and $\pm$ 0.25 Å for σ. These constraints were forced for two main reasons. First, is to make sure that the optimized parameters still agree with the theoretically expected quantities. Second, is to stay within the limits of the constructed PC functions.

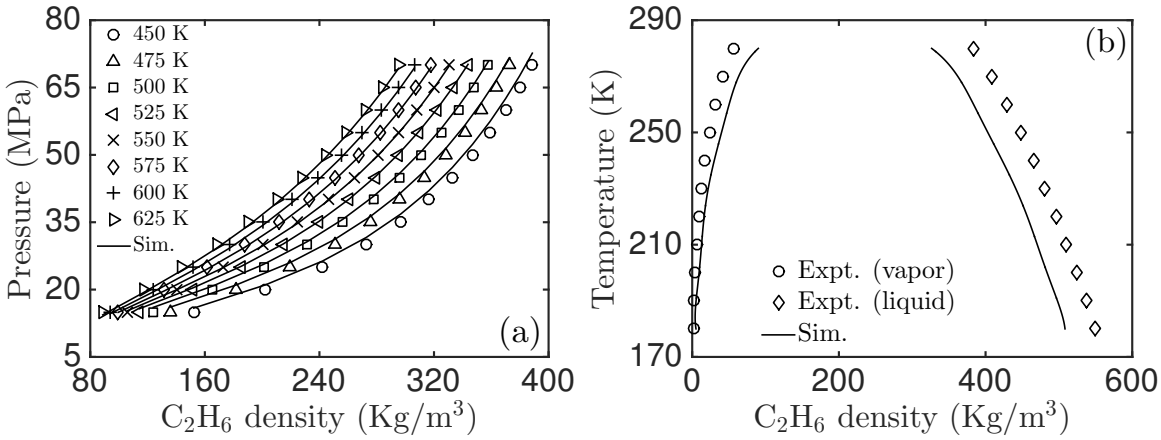

Figure 2.24: (a) Supercritical isotherms of ethane data (symbols) ranging from 450 to 625 K fitted by the single-phase PC surrogate model (lines). (b) Large deviations in two-phase region similar to CO_2 case.

The optimization scheme aims to minimize the objective function defined as the difference between the surrogate model predictions and corresponding experimental data. The data used for matching were the single-phase vapor pressures along several supercritical isotherms ($T > T_c$ and $P > P_c$), in addition to the two-phase (liquid and vapor) saturation densities at different temperatures. Therefore, the objective function (J) is formulated as the discrete norm over the experimental data used:

$$J = \left\| \frac{P^{\text{sur}} - P^{\text{exp}}}{P^{\text{exp}}} \right\|_2 + \left\| \frac{\rho_{\text{vap}}^{\text{sur}} - \rho_{\text{vap}}^{\text{exp}}}{\rho_{\text{vap}}^{\text{exp}}} \right\|_2 + \left\| \frac{\rho_{\text{liq}}^{\text{sur}} - \rho_{\text{liq}}^{\text{exp}}}{\rho_{\text{liq}}^{\text{exp}}} \right\|_2 . \tag{2.30}$$

Based on the described workflow above, LJ parameters for several molecules were optimized such that a good fit with experimental data was reached. At first, LJ parameters for three noble gases, namely Ar, Kr, and Xe, were obtained. In Figs. 2.17a and 2.17b, experimental data used for tuning Ar parameters were plotted against the simulated data generated by the recommended values ($\varepsilon_{\text{rec}}/k_B$ and σ_{rec}). In addition, the relative misfits (in %) between each of these points in each phase were respectively reported (see Figs. 2.17c and 2.17d). According to these results, the MC simulation outputs for Ar can be reproduced in few seconds using the constructed PC surrogates

within ± 5% deviation for the majority of the investigated data points. It is also evident that at supercritical conditions, larger misfits were found at lower pressures; in other words, at temperatures closer to the critical temperature. In similar fashion, higher deviations were observed while estimating the gas phase densities in the two-phase envelope. Such behavior can be attributed to the misfits inherited by the PC surrogates from the high statistical uncertainties of the MC simulations in the vapor phase.

Similar trends, but generally with higher misfits, were reached for both Kr and Xe (Figs. 2.18 and 2.19, respectively) when compared to Ar case. Furthermore, experimental data of bigger but rather simple molecules were matched too. For instance, results from PC surrogates showed excellent agreement with CH_4 data (Fig. 2.20). The resulting deviation was less than ±5 % for all data points studied, except for few vapor phase densities.

In addition to CH_4, both CO and N_2 were modeled as single-site LJ particles. Interestingly, the data for both molecules were reproduced with the recommended parameters in rather good agreement, despite the simplicity of the representing molecular model. Clearly, the match between experimental data and simulation results for CO and N_2, especially in the two-phase region, was not as good as the match obtained for any of the noble gases considered or CH_4.

A further, more challenging, step was taken trying to fit experimental data for CO_2 and C_2H_6. However, and as expected, the spherical LJ model failed to capture the complex nature of such molecules (Figs. 2.23 and 2.24). This misfit was dramatically manifested in the two-phase region. The reason for such failure is due to the limitation of the single-site model itself, as both CO_2 and C_2H_6 are linear molecules instead of spherical. Moreover, CO_2 partial charges, which play a vital role in determining its thermodynamic behavior, were not considered by the LJ model used.

To sum up, the large-scale optimization process for the LJ parameters was success-

fully demonstrated for several noble gases, in addition to CH_4, CO and N_2 molecules. On the other side, attempting to use the spherical non-polar LJ model to represent more complex molecules has failed. Therefore, the recommended LJ parameters for only the well-fitted molecules were reported (Table B.4 in Appendix B).

2.2.3 Conclusions

A novel approach combining MC molecular simulation with the PC expansions method was presented. This combination is particularly successful because MC simulations require very long computational time while having few input variables, which suits well the PC expansion formulation. The resulting surrogate models built via PC expansion functions allowed extremely fast and sufficiently accurate reproduction for both single- and two-phase data of LJ fluids. In addition, an application example for the combination of these two methods was demonstrated through conducting a large-scale optimization of LJ model parameters for few molecules. These recommended parameters are available in Table B.4 (Appendix B).

2.3 Reweighting and Reconstruction Method

Monte Carlo (MC) Markov chain collects the detailed configurational information of the molecules at the microscopic level throughout the simulation. For instance, ensemble averages such as system's energy or pressure are found by simply averaging these chains. In addition, fluctuations around average can be used to estimate second derivative properties such as heat capacity and fluid compressibility. Therefore, another way to accelerate MC molecular simulations is by extracting more information from these Markov chains such that a merely single simulation can replace other multiple ones. This could be an alternative path to reduce the computational cost instead of trying to modify the Metropolis algorithm itself.

For that purpose, researchers have developed several methods in order to retrieve more macroscopic information from the collected Markov chains. For example, histogram reweighting [46–48] and thermodynamic scaling [49–56] are two well-established methods so far. In the first, histograms generated per simulation are grouped to form a collective distribution. The created distribution is then reweighted by the proper factors in order to estimate ensemble averages at different conditions than the simulated ones. As for thermodynamic scaling it refers to a family of scaling methods based on non-physical umbrella sampling for several ensembles and various applications.

In this context, a novel simple and self-consistent reweighting and reconstruction scheme is proposed. In contrast to the reported methods in literature, where non-Boltzmann distributions are utilized and some tuned parameters are required, the proposed extrapolation schemes retain physically meaningful distributions and do not contain any tuned parameters. The suggested method is, so far, designed for structureless identical LJ particles in Canonical ensemble.

Coming paragraphs are devoted to highlight the theoretical basis and the work-flow of the proposed scheme. In addition, the frontiers opened by this method by coupling MC results with compressible single-phase single-component fluid flow in porous media.[3]

2.3.1 Theoretical Basis

2.3.1.1 Ensemble averages and second derivatives

Initially and for easier presentation, three configurational quantities u, v and w are defined. For convenience and as all the simulated particles are assumed identical, these quantities can be normalized with respect to ε and σ. For instance, The normalized

[3]The material in this part appears in Journal of Computational Physics [57] and Procedia Computer Science [58–60].

potential energy of a configuration (u^*) is computed by Equ. 2.31 (see below); where, u is normalized by ε and distance r_{ij} is normalized by σ ($r^*_{ij} = r_{ij}/\sigma$) and further divided by the normalized simulation box length ($L^* = L/\sigma$) such that $r^*_{ij} = L^* s^*_{ij}$. It is easy to see that s^*_{ij} represents the normalized distance separating the i^{th} and j^{th} particles relative to the normalized cubic simulation box length (L^*) with values ranging from 0 to 1. Similar normalization scenario is applied to the other required quantities in Equs. 2.32 and 2.33 below:

$$u^* = \frac{u}{\varepsilon} = 4\sum_{i<j}\left[\left(\frac{1}{L^* s^*_{ij}}\right)^{12} - \left(\frac{1}{L^* s^*_{ij}}\right)^{6}\right] \tag{2.31}$$

$$v^* = \frac{v}{\varepsilon} = 8\sum_{i<j}\left[2\left(\frac{1}{L^* s^*_{ij}}\right)^{12} - \left(\frac{1}{L^* s^*_{ij}}\right)^{6}\right] \tag{2.32}$$

$$w^* = \frac{w}{\varepsilon} = 8\sum_{i<j}\left[10\left(\frac{1}{L^* s^*_{ij}}\right)^{12} - 3\left(\frac{1}{L^* s^*_{ij}}\right)^{6}\right] \tag{2.33}$$

Canonical averages and fluctuations over the total sampled configurations are used in order to find certain quantities under interest. In particular, four quantities are considered as a starting point to test and validate the reweighting and reconstruction concept. These quantities are the normalized system's energy per particle ($\overline{E}^*$), the system's normalized pressure (p^*), the normalized isochoric heat capacity (C^*_v) and the inverse of the normalized isothermal compressibility ($c_T^{-1^*}$) and can be evaluated as follows (see Appendix A for detailed derivations of these quantities):

$$\overline{E}^* = \frac{E}{N\varepsilon} = \frac{3}{2\beta^*} + \frac{1}{N}\Big\langle u^* \Big\rangle_{can} + \frac{1}{N}E^*_{tail} \tag{2.34}$$

$$p^* = \frac{p\sigma^3}{\varepsilon} = \frac{\rho^*}{\beta^*} + \frac{1}{V^*}\Big\langle v^* \Big\rangle_{can} + p^*_{tail} \tag{2.35}$$

$$C_v{}^* = \frac{C_v}{k_B N} = \frac{3}{2} + \frac{1}{N}\beta^{*2}\left(\left\langle u^{*2}\right\rangle_{can} - \left\langle u^*\right\rangle_{can}^2\right) \tag{2.36}$$

$$c_T{}^{*-1} = \frac{c_T{}^{-1}\sigma^3}{\varepsilon} = \frac{\rho^*}{\beta^*} + \frac{1}{V^*}\left\langle w^*\right\rangle_{can} - \frac{\beta^*}{V^*}\left(\left\langle v^{*2}\right\rangle_{can} - \left\langle v^*\right\rangle_{can}^2\right) + c^*_{tail} \tag{2.37}$$

where N is the total number of particles, β^* is the normalized Boltzmann's factor and equivalent to the reciprocal of the normalized temperature ($T^* = T/(\varepsilon/k_B)$), k_B is the Boltzmann's constant and ρ^* is the normalized number density ($\rho^* = \rho\sigma^3$). Whereas, $\left\langle X\right\rangle_{can}$ stands for the canonical average of property X over all the sampled configurations and subscript "*tail*" corresponds to the tail correction functions used to count for truncated interactions beyond the imposed cutoff radius ($r_c^* = 0.49L^* = s_c^* L^*$). These tail correction functions can be computed using the following equations (see Appendix A for derivations):

$$E^*_{tail} = \frac{E_{tail}}{\varepsilon} = \frac{8}{3}\pi N\rho^*\left[\frac{1}{3}\left(\frac{1}{L^* s_c^*}\right)^9 - \left(\frac{1}{L^* s_c^*}\right)^3\right] \tag{2.38}$$

$$p^*_{tail} = \frac{p_{tail}\sigma^3}{\varepsilon} = \frac{16}{3}\pi\rho^{*2}\left[\frac{2}{3}\left(\frac{1}{L^* s_c^*}\right)^9 - \left(\frac{1}{L^* s_c^*}\right)^3\right] \tag{2.39}$$

$$c^*_{tail} = \frac{c_{tail}\sigma^3}{\varepsilon} = \frac{16}{3}\pi\rho^{*2}\left[\frac{10}{3}\left(\frac{1}{L^* s_c^*}\right)^9 - 3\left(\frac{1}{L^* s_c^*}\right)^3\right] \tag{2.40}$$

2.3.1.2 Reweighting and reconstruction workflow

The general case for applying the reweighting and reconstruction scheme is to predict an ensemble average $\left\langle A^*\right\rangle_{can}$ at certain temperature and density that are different from the ones used to generate the MC Markov chain. In other words, the general case will be to extrapolate ensemble averages along paths of changing temperature and

density at same time. This is possible by first reconstructing and then reweighting the original sampled Markov chain using the following relation

$$\left\langle A^*\left(\beta_n^*, L_n^*\right)\right\rangle_{can} = \frac{\sum_{k=1}^{M} \exp\left(-W_k\right) A_k^*\left(\beta_o^*, L_n^*\right)}{\sum_{k=1}^{M} \exp\left(-W_k\right)}, \tag{2.41}$$

$$W_k = \beta_n^*\left[u_k^*\left(\beta_o^*, L_n^*\right) + E_{tail}^*\left(L_n^*\right)\right] - \beta_o^*\left[u_k^*\left(\beta_o^*, L_o^*\right) + E_{tail}^*\left(L_o^*\right)\right]. \tag{2.42}$$

In the equations above, the reweighting factors "$\exp(-W_k)$" are related to the temperature and energy difference between the reference case and the targeted one. A can be replaced by any of the configurational quantities u, u^2, v, v^2 or w with "o" and "n" subscripts representing the original thermodynamic conditions and the new ones where the targeted point exists. Subscript "k" counts for the sampled configurations (M) out of the total created configurations forming the Markov chain.

Based on the general Equ. 2.41, extrapolation along isothermal and isochoric paths are also possible where both reconstruction and reweighting are required for the first, and only the latter is needed for the second. In case of extrapolating along isotherms both Equs. 2.41 and 2.42 reduce to

$$\left\langle A^*\left(\beta_o^*, L_n^*\right)\right\rangle_{can} = \frac{\sum_{k=1}^{M} \exp\left(-W_k\right) A_k^*\left(\beta_o^*, L_n^*\right)}{\sum_{k=1}^{M} \exp\left(-W_k\right)}, \tag{2.43}$$

$$W_k = \beta_o^*\left[u_k^*\left(\beta_o^*, L_n^*\right) + E_{tail}^*\left(L_n^*\right) - u_k^*\left(\beta_o^*, L_o^*\right) - E_{tail}^*\left(L_o^*\right)\right], \tag{2.44}$$

while for extrapolation along isochors, Markov chains are only reweighted as simulation box length is preserved (L_n^*=L_o^*). Consequently, Equs. 2.41 and 2.42 can be

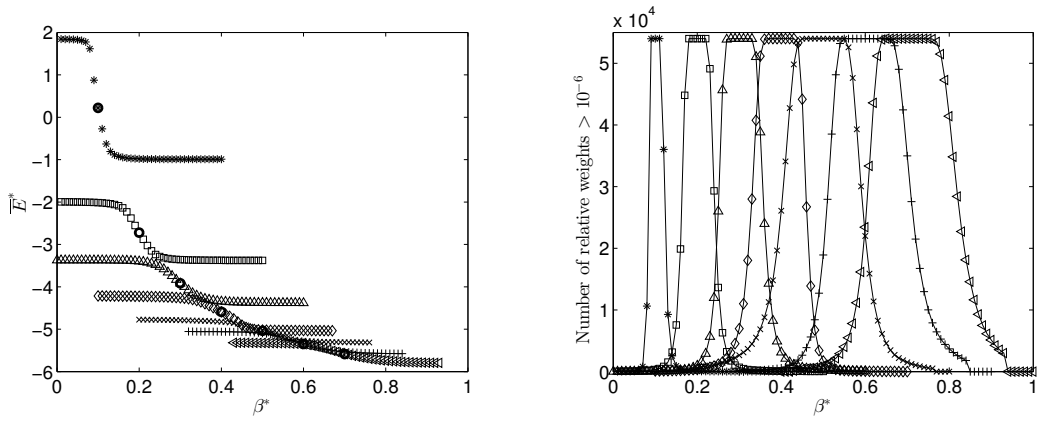

Figure 2.25: Estimated $\overline{E}^*$ (symbols) for long extrapolation ranges in comparison to simulated $\overline{E}^*$ (circles) along $\rho^* = 0.9$ isochor (Left). On right, number of relative weighting factors which are greater than 10^{-6} sharply damps as the targeted point resides a further β^*. Note that the ideal part of $\overline{E}^*$ is not included.

simplified to take the following forms:

$$\left\langle A^*\left(\beta_n^*, L_o^*\right)\right\rangle_{can} = \frac{\sum_{k=1}^{M} \exp\left(-W_k\right) A_k^*\left(\beta_o^*, L_o^*\right)}{\sum_{k=1}^{M} \exp\left(-W_k\right)}, \tag{2.45}$$

$$W_k = \left(\beta_n^* - \beta_o^*\right)\left[u_k^*\left(\beta_o^*, L_o^*\right) + E_{tail}^*\left(L_o^*\right)\right]. \tag{2.46}$$

2.3.1.3 Prediction limits

It is obvious that the suggested extrapolation scheme will have certain limits for its prediction ranges. For this purpose the change of $\overline{E}^*$ with respect to β^* along $\rho^* = 0.9$ isochor was partly reproduced (Fig. 2.25 left panel). Circles correspond to the simulated values obtained from generated MC chains for a system of 512 LJ particles, while the other symbols represent the attempt to estimate the neighboring $\overline{E}^*$ by reweighting the nearest simulated point. As can be seen, the extrapolation fails as the distance from the simulated point increases to reach a constant value. This value represents the tail correction function (E_{tail}^*). In order to better understand this result, right panel of Fig. 2.25 was created. In that figure, the number of significant

relative weights used to reweight the residual part (u_k^*) of $\overline{E}^*$ damps as the distance between the simulated point and the extrapolated one increases. Due to this sharp damping, the residual part vanishes and $\overline{E}^*$ approaches the constant value of E_{tail}^*. For more information on the prediction limits of the proposed method, the reader is referred to our papers [57, 58].

2.3.2 NVT-NpT Switching Scheme Results

In order to be able to couple the reweighting and reconstruction scheme with Darcy-scale flow simulator, a switching scheme that allows solving an NPT problem using NVT Markov chains was developed [59]. Before describing the switching mechanism from the NVT to the NpT ensemble, the application of the classical reweighting and reconstruction scheme in predicting p^* is first explained. Starting from Equ. 2.35, the following relation shows how p^* at a new temperature ($\beta_n^* = 1/\mathrm{T}_n^*$) and density ($\rho_n^*$) can be estimated using the information from a Markov chain generated at neighboring β_o^* and ρ_o^*.

$$p^*\left(\beta_n^*, L_n^*\right) = \frac{\rho_n^*}{\beta_n^*} + \frac{1}{V^*}\Big\langle v^*\left(\beta_n^*, L_n^*\right)\Big\rangle_{can} + p_{tail}^*\left(L_n^*\right). \tag{2.47}$$

In the equation above, the ideal part and the tail correction function are directly modified to adhere with the new prescribed conditions (β_n^* and ρ_n^*=$N/{L_n^*}^3$). However, the residual part $\langle v^*\left(\beta_n^*, L_n^*\right)\rangle_{can}$ is computed as explained in Equ. 2.41; such that, the reweighting factors are evaluated as exactly shown in Equ. 2.42.

$$\Big\langle v^*\left(\beta_n^*, L_n^*\right)\Big\rangle_{can} = \frac{\sum_{k=1}^{M} \exp\left(-W_k\right) v_k^*\left(\beta_o^*, L_n^*\right)}{\sum_{k=1}^{M} \exp\left(-W_k\right)}. \tag{2.48}$$

So far the problem is forward, as the only targeted unknown is p^*. Nonetheless, another way to visualize this problem is by assuming that the new pressure $p^*\left(\beta_n^*, L_n^*\right)$ and β_n^* are given and to look for their corresponding ρ_n^*. In standard MC simulations, such new formulation requires designing a new experimental setup which, as men-

tioned earlier, is called the NpT ensemble. In other words, the numerical experiment has to be rerun again with the same expensive computational cost. On the other hand, this work shows that such a switch between the NVT and NpT ensembles is fairly simple when using the reweighting and reconstruction scheme. Basically, the same Equ. 2.47 is used but this time to search for L_n^* value that successfully retains the prescribed pressure $p^*\left(\beta_n^*, L_n^*\right)$. In the following section, the proposed scheme will be further elaborated with more details.

The extensive analysis conducted to test and validate the proposed approach, is based on a set of 750 Markov chains generated by MC molecular simulations in NVT ensemble. This set spans the range of $0.05 \leq \beta^* \leq 0.75$ and $0.02 \leq \rho^* \leq 1$. The β^* range is chosen such that the system is always at a temperature higher than the critical temperature to avoid entering the two-phase envelope. On the other hand, ρ^* values are decided such that cases from sparse to highly dense systems are considered. In generating the 750 chains, a system of 256 uniformly distributed structureless LJ particles in a cubic cell with dimensions fulfilling the prescribed system's ρ^* is initially used. Starting from these initial configurations, a sequence of successive configurations is created by introducing elemental changes to the system, known as MC trials. For this particular case, particle displacement was the only possible MC trial. Therefore, at each MC step a particle is randomly chosen and displaced. For each run, a total of 61,440,000 MC trials was performed to generate the same number of configurations. The first 10 % of these configurations were discarded to account for equilibration. Out of the 90% left, only one configuration was sampled every 1024 steps. As a result, each of these 750 Markov chains is made of 54,000 sampled configurations. As a matter of fact, for the reweighting and reconstruction method only the sum of the short range $(1/s^{12})$ and long range $(1/s^6)$ interactions need to be separately collected in the Markov chains.

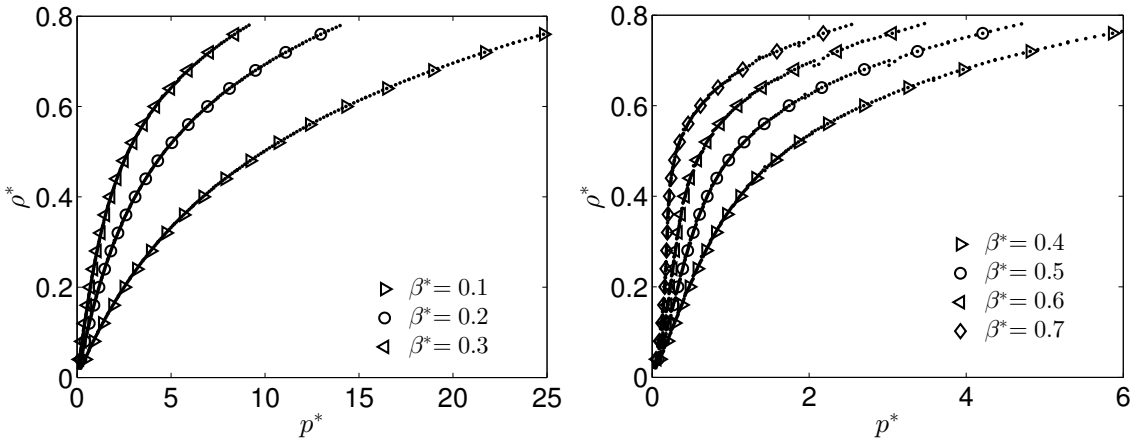

Figure 2.26: Extrapolated ρ^* $(\cdot)$ along seven different isotherms ($\beta^* = 0.1$, 0.2 ... and 0.7). Markov chains used as source points for extrapolation are plotted as symbols. Each source is used to extrapolate the adjacent five dots from each side. The plots are split into two for clarity.

2.3.2.1 Solving the NpT problem

After conducting the necessary MC simulations in NVT ensemble, a data bank of MC Markov chains is available. In order to solve the NpT problem, p^* is first computed for each chain corresponding to specific β^* and ρ^* using Equ. 2.35. Therefore, for every chain in the generated data bank the three thermodynamic states $\beta^* = 1/T^*$, ρ^* and p^* are known. Assuming that we want to estimate ρ^* at a given p_x^* and β_x^*, the first step is to determine the suitable MC Markov chain from the generated set of chains as the source of extrapolation. Considering the case of extrapolation along isothermal paths, the chain with β^* equal to β_x^* and p^* closest to the targeted p_x^* is chosen as the source point of extrapolation.

In Fig. 2.26, seven isotherms were reproduced. Along each isotherm only 19 Markov chains were used as source points for extrapolation. From each source point, five adjacent ρ^* values are predicted from each side at different p^* values such that the extrapolations from two consecutive source points meet at the middle. In fact, these predictions were made using Equ. 2.47 while solving for L_n^* as both $p^*\,(\beta_n^*, L_n^*)$ or p_x^* and β_n^* or β_x^* at extrapolation points are known. According to Fig. 2.26, all

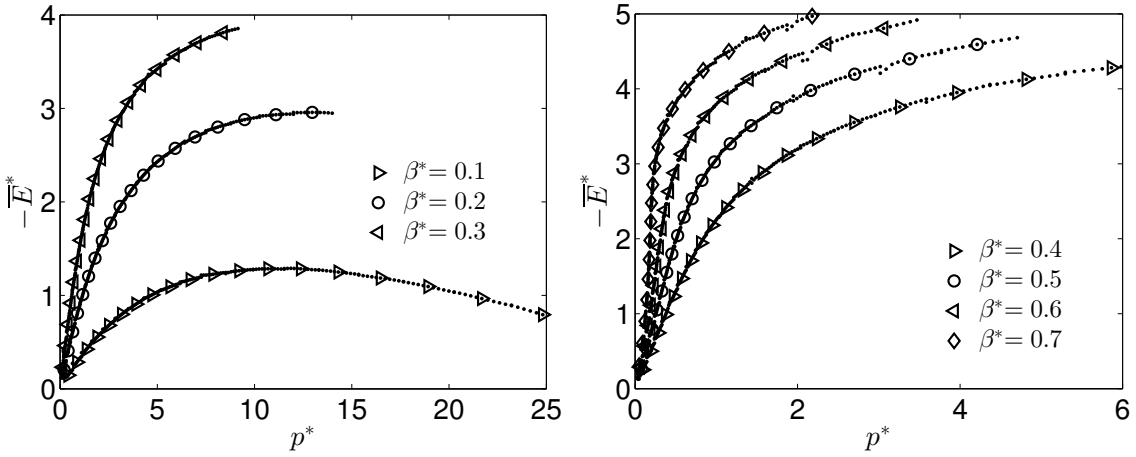

Figure 2.27: Same as Figure 2.26 but for extrapolating $-\overline{E}^*$.

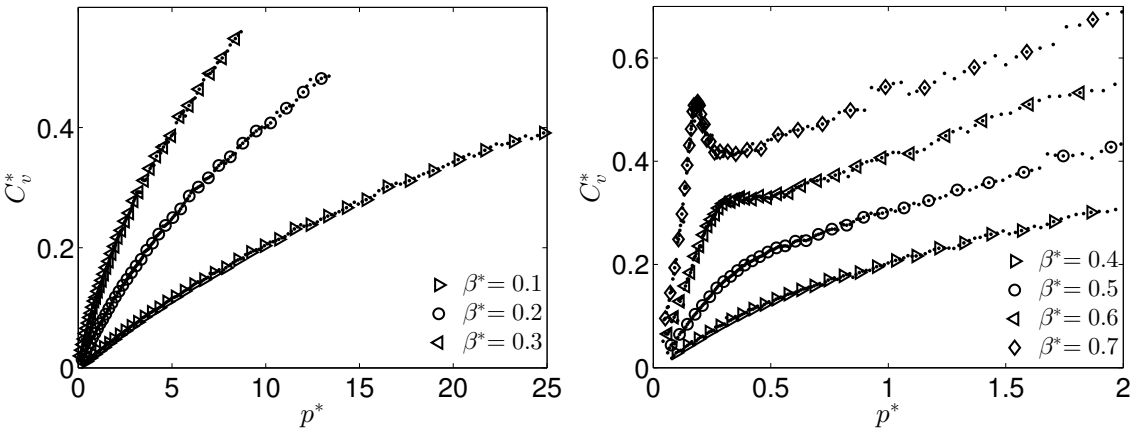

Figure 2.28: Same as Fig. 2.26 but for extrapolating C_v^*.

the isotherms are well reproduced as the transitions from one extrapolation set to another are, in most occasions, smooth. However, in agreement with results reported in our previous papers [57, 58], the ranges of extrapolation are dependent on the thermodynamic states of the source points of extrapolation. It is observed that at low temperatures (high β^*) the ranges get narrower as the system approaches the critical temperature.

2.3.2.2 Switching back to the NVT ensemble

As the ρ_x^* values associated with different pressures (p_x^*) of interest are all known, it is possible now to shift back to the regular forward reweighting and reconstruction

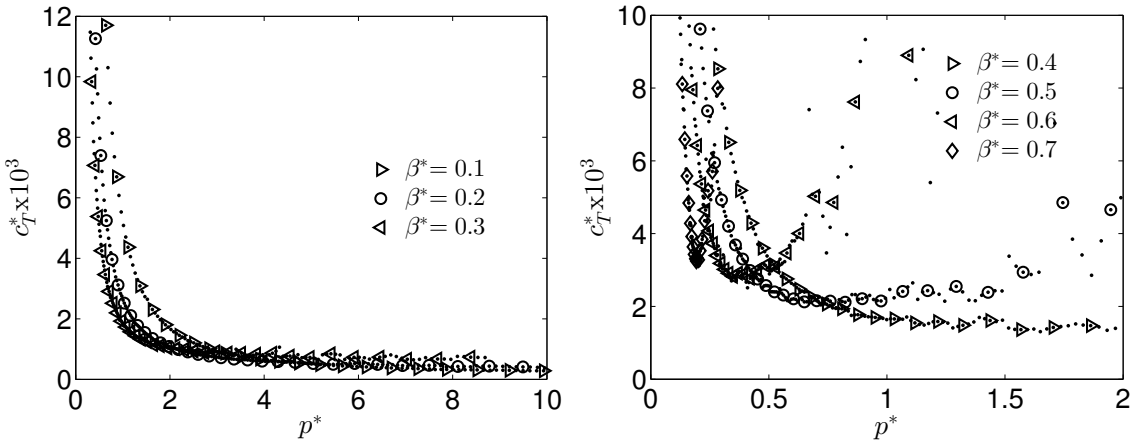

Figure 2.29: Same as Fig. 2.26 but for extrapolating c_T^*.

scheme to solve for the other three properties considered using Equs. 2.34, 2.36 and 2.37. In this regard, three more figures are generated reproducing isotherms of $-\overline{E}^*$ (Fig. 2.27), C_v^* (Fig. 2.28) and c_T^* (Fig. 2.29). Again here, the plots were split into two sets for clarity. Similar to ρ^*, extrapolation ranges diminish as source points approach critical temperature. In addition, it is evident that for second derivative quantities narrower ranges of extrapolation are expected when compared to primary quantities. To sum up, the MC Markov chains generated from NVT ensemble simulations were used to solve for density at a given pressure and temperature. To solve the inverse problem, the MATLAB built-in non-linear solver "fsolve" was used. Whenever densities are determined, the rest of the parameters are evaluated using the forward scheme.

2.3.2.3 Extrapolation along paths of changing β and p

To take the method a step further, in fact, the extrapolation process does not have to solely be along isotherms. It is possible to extrapolate from a certain source point to other neighboring points at both different temperature and pressure conditions. As an example (Fig. 2.30), source points ($\circ$) along $\beta^* = 0.3$ were used to extrapolate points ($\cdot$) at different pressures along two different isotherms $\beta^* = 0.25$ ($\triangleright$) and 0.35

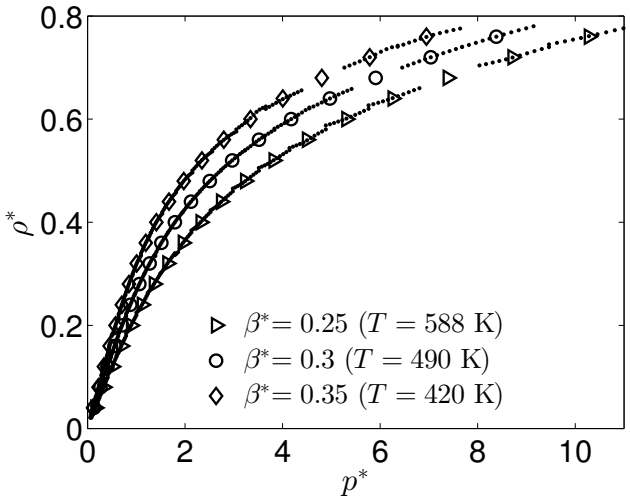

Figure 2.30: ρ^* ($\cdot$) values are extrapolated using Markov chains along $\beta^* = 0.3$ ($\circ$) as source points. Each source point is used to extrapolate along the same isotherm and the other two plotted ones. To better visualize the extrapolation area along different temperature and pressure conditions, the points extrapolated by one of the source points are removed.

($\diamond$). The symbols ($\triangleright$ and $\diamond$) are only plotted to show how well the extrapolation was. In addition, the extrapolated points by one of the source points were removed to help visualize the extrapolation area covered by each source point.

Moreover and in order to get a clear idea about the extrapolation limits of the proposed scheme, the normalized quantities (β^*) are transformed into temperatures (T) with absolute units. By referring to Fig. 2.30, the source points along $T = 490$ K isotherm ($\circ$) were successful in reproducing the other two isotherms with more than 70 K temperature difference from both sides. Again, for more details about extrapolation limits, our two earlier papers are recommended [57, 58].

2.3.2.4 Validation

Finally, the algorithms and extrapolation schemes used in this work are all validated by comparing the data from direct simulations and extrapolated ones with respect to methane experimental data [44]. In specific, density values (Fig. 2.31) are used in comparison. Moreover, and for better comparison, quantities in absolute units are

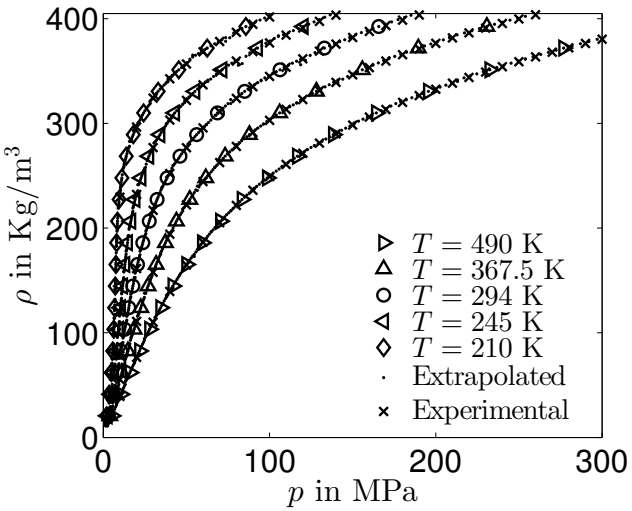

Figure 2.31: Comparison between ρ experimental values ($\times$) against simulated (Symbols) and extrapolated ($\cdot$) ones. Only five of the previous isotherms are reported as for the other two, for the best of our knowledge, no experimental data exists.

reported instead of the normalized ones. We note that, for converting quantities into real units $\varepsilon/k_B = 147$ K and $\sigma = 3.722$ Å were used [57].

2.3.3 Multi-Scale Compressible Single-Phase Flow Simulator

In this part, a successful coupling between Darcy-scale flow simulator and MC molecular simulation is demonstrated. In particular, a compressible single-phase flow case is considered. MC molecular simulation, via reweighting and reconstruction method and the switching scheme, replaces the equation of state in terms of estimating the necessary thermophysical quantities needed by the flow simulator.

2.3.3.1 Flow governing equations

The cell centered finite difference method for compressible single-phase single-component system was used. The system was reduced to a set of linear equations in which the pressure field was the primary unknown. Then using MATLAB's linear solver, these equations were implicitly solved employing a constant time step. Simulations with uniform and non-uniform rectangular mesh were both considered.

The governing equations that describe the system start with the mass conservation law, which equates the difference between the input and the output fluxes to the accumulation term in the unit cell:

$$\frac{\partial(\phi\rho)}{\partial t} + \nabla.(\rho\mathbf{u}) = q. \tag{2.49}$$

In Equ. 2.49, ϕ is the rock porosity, ρ is the mass density, $\mathbf{u}$ is the Darcy velocity, q is the mass source and t is the time. Based on the definition of the isothermal compressibility, the mass conservation law can be formulated in terms of pressure instead of density as follows:

$$\phi\rho c_T \frac{\partial p}{\partial t} + \nabla.(\rho\mathbf{u}) = q, \tag{2.50}$$

where p is the pressure and c_T is the isothermal compressibility. On the other hand the Darcy velocity can be expressed as follows:

$$\mathbf{u} = -\frac{k}{\mu}\nabla p. \tag{2.51}$$

under the assumption of no gravitational field effect, such that k and μ are the rock permeability and fluid viscosity, respectively. As for the production well model, the well known Peaceman radial well model [61] was employed:

$$q = WI\,(p_{bh} - p_c)\,, \tag{2.52}$$

$$WI = \frac{2\pi\rho\sqrt{k_x k_y}h_z}{\mu ln(\frac{r_e}{r_w})}. \tag{2.53}$$

In the Peaceman model, WI is the well index, k_x and k_y are the permeabilities in the x and y directions respectively, h_z is the vertical thickness, p_{bh} is the well bottom

hole pressure, p_c is the cell average pressure, r_e is the cell equivalent radius and r_w is the well radius.

2.3.3.2 Reservoir fluid properties

At first, the natural gas correlation by Lee et al. [62] was adopted in order to compute methane viscosity as function of pressure and temperature:

$$\mu_g = K_1 e^{X\rho^Y}, \tag{2.54}$$

$$K_1 = \frac{(0.00094 + 2 \times 10^{-6} M)^{1.5}}{(209 + 19M + T)}, \tag{2.55}$$

$$X = 3.5 + \frac{986}{T} + 0.01M, \; Y = 2.4 - 0.2X. \tag{2.56}$$

In the equations above, μ_g is the viscosity in cp, ρ is the density in g/cm^3, p is the pressure in psia, T is the temperature in R and M is the gas molecular weight. Finally, the reweighting and reconstruction method in addition to the NVT-NpT schemes were used in order to have good estimations of the fluid density and isothermal compressibility.

2.3.3.3 Production scenarios

The proposed coupling was tested with various production scenarios from methane gas reservoirs (Methane LJ parameters are: $\epsilon/\mathrm{k_B}$= 147.4 K and σ = 3.722 Å) at isothermal condition (T = 368.5 K). These reservoirs have different configurations with a 2D domain covering 240 m × 240 m area of 0.2 porosity with Neumann no flow boundary condition. In all the studied scenarios, the production lasted for 200 days, while the initial reservoir pressure and the well down-hole pressure were 200 and 100 atm, respectively. The proposed multi-scale coupling flow chart between MC

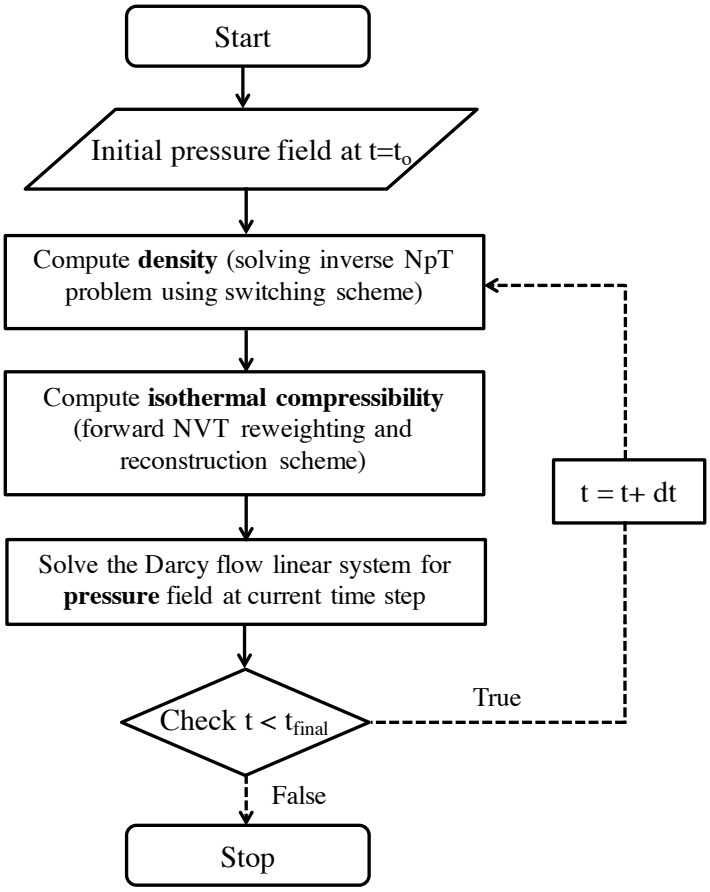

Figure 2.32: MC simulation and Darcy scale coupling algorithm.

molecular simulation and Darcy flow, used in simulating the studied cases in this study, is given in Fig. 2.32.

In the first scenario, a homogeneous permeability field of 100 md was imposed throughout the reservoir domain which was discretized as 80 × 80 cells. A single production well was placed at the center. Fig. 2.33 shows the contour plot of pressure field with the velocity streamlines and the production rate profile relative to initial production. As expected, the pressure and velocity fields are symmetric around the wellbore because of the homogeneity of the domain properties.

In the second scenario, a heterogeneous permeability field with values ranging from less than 1 md to about 200 md was imposed throughout the reservoir domain with a single production well at the center with 50 × 50 cells discretized domain.

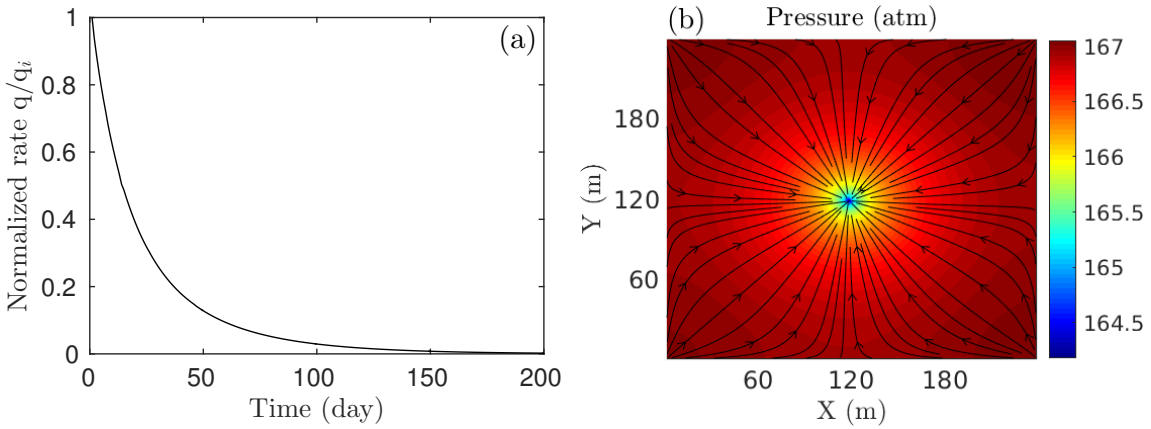

Figure 2.33: Homogeneous reservoir case: (a) Normalized production rate profile. (b) Pressure contour map with darcy velocity streamlines after 10 days of production.

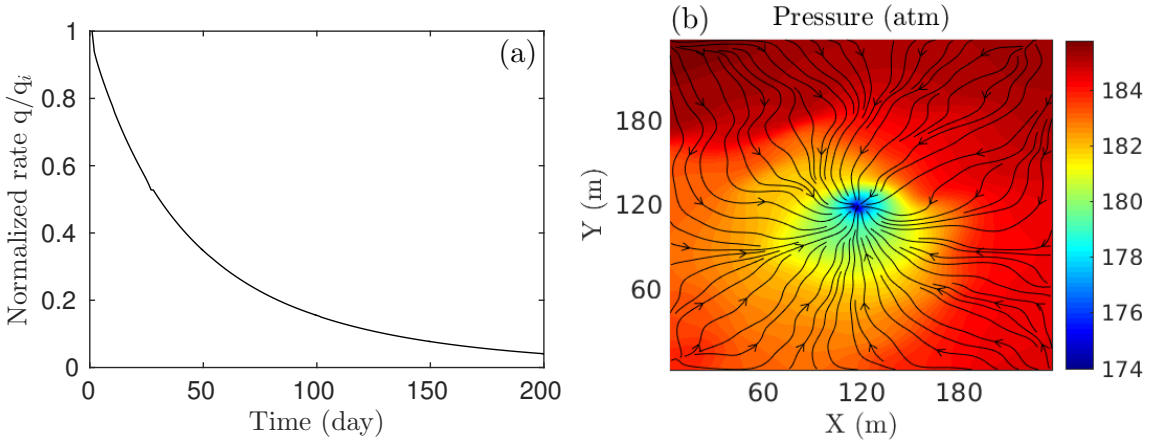

Figure 2.34: Heterogeneous reservoir case: (a) Normalized production rate profile. (b) Pressure contour map with Darcy velocity streamlines after 10 days of production.

The effect of the heterogeneous medium is clearly seen in the irregular shape of the streamlines and the pressure contour plot (Fig. 2.34.b). In this case the production rate decline is slower than the homogeneous case (Fig. 2.33.a and 2.34.a) although the overall average rock permeability in the two cases is almost the same. This is mainly because of the presence of dead zones in the heterogeneous reservoir that have very low permeability and as a result, the reservoir is depleted at slower rates. In our paper [60], more production scenarios are reported.

2.3.4 Conclusions

An efficient coupling between MC molecular simulation and Darcy flow simulation in porous media was developed using the reweighting and reconstruction method along with NVT-NpT switching scheme. The proposed combination was tested on methane system in different reservoir configurations. The accurate description of thermodynamic behavior of the simulated reservoir fluids using MC Molecular simulation decreases the uncertainty in reservoir simulation predictions. The reweighting and reconstruction technique allows extrapolation in seconds instead of running the typical molecular simulation for hours. In conclusion, the implementation of the reweighting and reconstruction method using few offline, pre-computed Markov chains has made it computationally feasible to benefit from the molecular simulation accuracy in determining phase properties needed by the reservoir simulator.

Chapter 3

Shale Gas Sorption and Transport Modeling

The global natural gas market has been significantly affected by the large increase in shale gas reservoirs exploitation. Shale gas reservoirs are categorized as unconventional energy resources. They are usually characterized by extremely small grain sizes, low permeability, and heterogeneity in chemical composition (Organic and inorganic permeable media). These factors have increased the complexity towards understanding the underlying mechanisms for gas sorption, transport and phase behavior in such systems.

Naturally occurring clay minerals determine the key physical traits such as permeability and certain chemical properties such as gas adsorption of shales [63–66] and provide a distinctive material for carbon dioxide sequestration [67–70] and selective sorption [71–74]. Shale mineralogy varies widely such that some shales are silt-rich or carbonate-rich whereas others are dominated by clay minerals, for example, illite, chlorite, smectite, and kaolinite [64, 75, 76]. Processes such as Knudsen diffusion, slip flow (Klinkenberg effect) and adsorption at the solid matrix can affect permeabilities and Darcy-type flow would be disturbed in case of deviation from the laminar flow [77]. However, none of the above models predict the correct behavior of shale gas production data. Thus in spite of the recent improved success in shale gas production [78],

the underlying adsorption and flow mechanisms in shale media are not well understood [79,80]. In nature, smectite such as montmorillonite is one of the most common type of swelling clay minerals [81–87]. To date, there have been numerous simulation studies on water adsorption and swelling in montmorillonites [81, 83, 84, 88–97]. The stable basal d-spacing is around 10 Å for dry clays and increases upon interaction with water to the range 11.5-12.5 Å forming a fully saturated monolayer (1W) water arrangement [83, 88, 90–93, 98]. Due to intake of more water, the d-spacing can increase to the next stable state (14.5-15.5 Å) where water forms a bilayer (2W) structure. Likewise, measured basal d-spacings for three layers of water (3W) are in the range 18.0-19.1 Å. In recent years, molecular modeling of the structure and dynamics of the water-methane mixture in the interlayer region of smectites has attracted interest [79, 99–103]. These simulations demonstrated that montmorillonite surfaces facilitate methane hydrate crystallization from aqueous solution in agreement with experiments [104]. Geologic sequestration of carbon dioxide has recently emerged as an alternative for reducing greenhouse emissions [67–70, 86, 103, 105–110].

According to experimental investigations [67, 69], carbon dioxide adsorption capacity of clay minerals is comparable to that of coal. The swelling of organoclays in the supercritical CO_2 medium is also important in nanocomposite preparations and applications [111, 112]. Preferential adsorption of ions and molecules on clay minerals such as montmorillonite can be used for ion exchange [71] and gas selectivity [72–74,79,113]. The main component of shale gas is methane, but the composition of the shale gas depends on multiple factors, most of which have geological and geochemical origins [114]. An important issue relevant to methane recovery [64, 65] and CO_2 storage capacity [67, 69] in clays involves the effect of the presence of preadsorbed water, which cannot be avoided owing to the hydrophilic nature of the samples. The presence of such adsorbed solvents leads to, for example, a striking increase of gas solubility in porous solids with pore size in the range of nanometers [115] and

modified sorbate intake [116]. Experimental studies have considered adsorption on dry [64–67, 70, 73] as well as on moisture equilibrated [64, 65, 67, 69, 86, 108, 117] clays. The mixture adsorption data is typically predicted from pure component adsorption measurements using the capability of approximate methods such as the ideal adsorbed solution theory (IAST) [118, 119]. In common practice, IAST can give fair predictions of gas mixture adsorption in many zeolites [120] and metal-organic frameworks [121].

In the first study of this chapter, we chose Na-, Cs-, and Ca-montmorillonite for the intercalation of CO_2 and CH_4 molecules into wet clays. This is not only with a view to getting new insight into the effect of ionic size and charge on adsorption behavior of CO_2 and CH_4 molecules, but also because they have already been elaborately investigated for pure water adsorption in clays, both experimentally and in molecular simulations. The adsorption of methane or CO_2 upon montmorillonite from mixture containing species (impurities) other than hydrocarbons and water has been rarely studied [107, 122, 123]. The objective of our present work was to systematically investigate using grand canonical Monte Carlo (GCMC) simulations the adsorption of pure sorbate molecules such as methane, carbon dioxide, nitrogen and hydrogen sulfide as well as their mixture upon clay minerals in the presence of water. The number of adsorbed water molecules is kept constant in our simulations upon subsequent gas adsorption, because typical hydrophilic pores indicate that water remains adsorbed upon carbon dioxide and methane adsorption [65, 86, 107, 124]. The adsorption isotherms and density profiles of these four molecules, at different basal d-spacings and densities of preadsorbed water in the interlayer of clays, are first obtained from GCMC simulations in the μVT ensemble, where μ is the chemical potential of the adsorbing molecule. For this study, we chose basal spacings d = 12, 15, 18, and 30 Å covering, for example, the swelling states (at fully saturated water arrangement) $\sim$ 1W, $\sim$ 2W, $\sim$ 3W, and $>$ 3W, respectively. The simulated d-spacings are similar in spirit to previous studies [83, 84, 90, 95, 96] and consistent

with those known for montmorillonites in the atmospheric environment containing sorbate molecules [65, 67, 69, 107].

When using pre-determined interlayer distances to study the adsorption of binary and ternary mixtures with one of their components (water) preadsorbed at certain concentrations, the final equilibrated compositions of the mixtures do not necessarily correspond to the thermodynamically stable states. Recent simulations employed similar pre-determined interlayer distances for GCMC simulations of adsorption of binary mixtures (e.g., H_2O/CO_2, H_2O/CH_4, and CO_2/CH_4) in clays [79,103,113,125]. The adsorption simulation using a large basal d-spacing (30 Å) would represent, for example, gas adsorption onto an external montmorillonite surface. Since the experimental data of ternary mixtures are not yet available, we assume that the stable states for ternary mixtures (e.g., $H_2O/CO_2/CH_4$) are also close to the above selected d-spacings ($\sim$ 1W, $\sim$ 2W, $\sim$ 3W etc.) over the studied conditions. In addition, IAST based on fitting curves of single component adsorption isotherms is used to predict mixture isotherms, and compared with the corresponding results of our direct molecular simulations.

In the second part, molecular dynamics (MD) simulations were used to investigate the structural and transport properties of CO_2, methane, and their mixture in hydrated Na-montmorillonite at preadsorbed water content and 298.15 K. Our results cover the experimental RH region, where swelling and shrinking normally occur. For this study, we chose the same basal spacings as earlier (d = 12, 15, 18, and 30 Å). These initial steps towards molecular understanding of water and CO_2 in montmorillonite interlayers are important for determining impacts from the long-term exposure of carbon dioxide to geological formations [69, 126, 127]. All GCMC and MD simulations conducted in this chapter were performed using MedeA software from Materials Design.

3.1 MC Study of CH_4, CO_2 and Water Adsorption in Montmorillonite

Using grand canonical Monte Carlo (GCMC) simulations, we study the adsorption behavior of CH_4, CO_2, and their mixtures at 298.15 K and pressures up to 50 bar in Na-, Cs-, and Ca-montmorillonite clays in the presence of water. Montmorillonite clays in the presence of preadsorbed water, preferentially adsorb CO_2 over CH_4 during both pure component and mixture adsorption. We observe the general trend that the presence of increasing preadsorbed water content in the clay interlayers, reduced adsorption amounts of pure CH_4 and CO_2 molecules. With a relatively large basal spacing (d= 30 Å), the favorability of adsorption of CO_2 by montmorillonite at relatively low pressures and intermediate water contents has been demonstrated using simulations. GCMC simulation is also used to assess the effect of water on the adsorption of N_2/CH_4, H_2S/CH_4, CO_2/N_2, and CO_2/H_2S binary mixtures in Na-montmorillonite clay. The ideal adsorbed solution theory is shown to agree well with the observed adsorption capacities and selectivities of mixtures in Na-montmorillonite clay.[1]

3.1.1 Simulation Details

3.1.1.1 Simulation cell

The model of clay used is the Wyoming-type montmorillonite [81,83,88,90,93,97,103], which contains tetrahedral and octahedral substitutions. Wyoming-type montmorillonite is commonly referred as Na-enriched montmorillonite occurring in nature. Texas-type (Ca-enriched) montmorillonite also has tetrahedral and octahedral substitutions. The montmorillonite modeled here has a layer charge of -0.75e, ap-

[1] The following material corresponding to this section is published in Microporous and Mesoporous Materials journal [128].

proximately 66.6% of which is in the octahedral layer, and the unit cell formula is $M_{0.75/n}(Si_{7.75}Al_{0.25})(Al_{3.5}Mg_{0.5})O_{20}(OH)_4$, where M represents a counterion (Na^+, Cs^+, or Ca^{2+}) and n represents the charge on the ion. This clay model is based on the pyrophyllite unit cell structure ($Si_8Al_4O_{20}(OH)_4$) with the position of the atoms in the unit cell of the clay taken from Skipper et al. [81]. In order to minimize finite size effect, the simulation box is made up of two tetrahedral-octahedral-tetrahedral (TOT) clay layers [83], each containing replicated pyrophyllite unit cells ($8 \times 4 \times 1$) with dimensions 42.24, 36.56, and 6.56 Å along x, y, and z directions, respectively. The simulation supercell of orthorhombic symmetry, thus consists of two TOT clay layers made up by a total of 64 unit cells.

Periodic boundary conditions were employed in all three spatial dimensions. The two TOT clay layers are parallel to each other and considered to be rigid molecules. At the start of each simulation, the water molecules and counterions were placed at random within the interlayer region. The sorbate molecules are permitted to move in and out of the simulation box, while the number of preadsorbed water molecules and counterions are kept constant during GCMC simulations. Note, however, that the dynamics of preadsorbed water molecules and counterions are allowed. The z dimension of the simulation box $L_z = 2d$, where d is the basal spacing which is the sum of the clay layer thickness and the interlayer space. The distance along z direction between planes through the centers of the inner oxygen atoms of two adjacent TOT layers defines the interlayer spacing. As in previous simulations [83, 90, 103], the clay layer of Wyoming-montmorillonite was constructed by introducing in each clay sheet 16 isomorphic substitutions of trivalent Al atoms by divalent Mg atoms in the octahedral sheet, 8 isomorphic substitutions of tetravalent Si atoms by trivalent Al atoms in the tetrahedral sheets, and 24 compensating monovalent sodium or cesium ions in the interlayer region (12 in the case of Ca-montmorillonite). All these substitutions were assigned in a fashion to obey the Loewenstein's rule (neighboring

Al-O-Al avoidance).

3.1.1.2 Force fields

All atoms in the system interact via the pairwise additive Lennard-Jones (LJ) 12-6 function representing the van der Waals energy term [8, 93],

$$U_{\mathrm{LJ}}(r_{ij}) = 4\varepsilon_{ij}\Big[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\Big], \tag{3.1}$$

where r_{ij} is the distance between the centers of i and j atoms. The parameter ε_{ij} controls the strength of the short-range interactions, and the LJ diameter σ_{ij} is used to set the length scale. The LJ parameters σ_{ij} and ε_{ij} are deduced from the conventional Lorentz-Berthelot combining rules [8]:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}, \tag{3.2}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}. \tag{3.3}$$

The charged atoms are interacting with each other via the unscreened Coulomb potential

$$U_{\mathrm{Coul}}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{3.4}$$

where q_i and q_j are the partial charges of the atoms i and j, respectively, and ε_0 is the dielectric permittivity of vacuum. The CLAYFF force field [93] that consists of nonbonded (electrostatic and van der Waals) terms predicts thermodynamic properties of clays in fair agreement with experiments [66, 93, 96, 97]. The CLAYFF parametrization incorporates structural and spectroscopic data for a variety of simple hydrated compounds. The partial charges were obtained by the electrostatic potential (ESP) analysis of the density functional theory results [93]. Thus in the present work, the charges of the clay atoms and their Lennard-Jones parameters are taken

from Cygan et al. [93]. Water is represented by the extended simple point charge (SPC/E) model, which consists of three interaction sites and a rigid structure with a tetrahedral H-O-H bond angle and 1.0 Å O-H bond lengths [129]. The CLAYFF force field adapted to the SPC/E water model provides flow properties in reasonable agreement with experiment [96]. Na^+ and Cs^+ cations are modeled using the parameters proposed by Smith and co-workers [92], and the LJ parameters of Ca^{2+} are taken from Koneshan et al [130]. For sorbate molecules considered as rigid, several different force fields are utilized that exhibit a very good ability to predict thermodynamic properties. Methane (single-site) is represented by the TraPPE force field [131]. The force field parameters for nitrogen (three-site) [132] and hydrogen sulfide (four-site) [133] were taken from the literature. The nitrogen model involves two LJ force centers separated by a fixed distance of 1.098 Å as well as two negative electrostatic charges located on the atomic centers and one positive charge (q_1) on the center of mass. The rigid H_2S model has the standard H-S bond length (1.34 Å) and H-S-H angle (92°). In this model, an additional partial charge (q_1), located at a point on the bisector of the H-S-H angle, is taken into account. The LJ parameters for carbon dioxide are obtained from the flexible version of the EPM2 model suggested by Harris and Yung [134]. This model approximates carbon dioxide by three sites with each site represented by a LJ sphere with an embedded central point charge, and the bond angle was allowed to fluctuate according to the bending potential

$$U_{\text{angle bend}}(\theta) = \frac{1}{2} k_\theta (\theta - \theta_0)^2, \tag{3.5}$$

where k_θ, θ, and θ_0 are the force constant, bending angle, and equilibrium bending angle, respectively. The CO_2 molecules are taken with a fixed C-O bond length of 1.15 Å, $k_\theta = 1236$ kJ/mol/rad^2, and $\theta_0 = \pi$ rad. The LJ parameters and charges used in this study are presented in Table B.5 in Appendix B.

3.1.1.3 Simulation methods

We used grand canonical Monte Carlo (GCMC) technique to simulate adsorption of various sorbate molecules in the presence of water and counterions by rigid montmorillonite in the μVT ensemble. In these calculations, sorbate molecules are allowed to move in and out of the simulation box, while the number of preadsorbed water molecules and counterions are kept constant. Note, however, that the dynamics of preadsorbed water molecules and counterions are allowed during GCMC simulations. Thus we simulate a clay system in equilibrium with an infinite reservoir of molecules that fixes the chemical potential μ_{A} of each species (A = CH_4, CO_2, N_2 or H_2S) at temperature T. The Monte Carlo trial moves consist of displacements and/or rotations of different molecules in the interlayer, and deletion or insertion of sorbate molecules. The new configuration resulting from the translational or rotational move is accepted following the standard Metropolis acceptance criterion [8]. However, due to the high densities of molecules in the interlayer, efficient sampling of sorbate such as CO_2 in the grand canonical ensemble requires a rotational-bias insertion technique [83]. In this method, a CO_2 molecule is inserted at a random position in the system and the selection of its orientation is biased. The acceptance probability for the insertion step reads

$$P_{\mathrm{acc}}^{N\to N+1} = \min\left[1, \exp\left(-\beta\left[\triangle U + \ln\left(\frac{ZV}{N+1}\right) + \beta^{-1}\ln(kP_j)\right]\right)\right], \tag{3.6}$$

and for the deletion

$$P_{\mathrm{acc}}^{N\to N-1} = \min\left[1, \exp\left(-\beta\left[\triangle U + \ln\left(\frac{N}{ZV}\right) - \beta^{-1}\ln(kP_{j=1})\right]\right)\right], \tag{3.7}$$

where P_j is the probability of selecting the jth orientation from k randomly generated orientations, $\beta = 1/(k_{\mathrm{B}}T)$, $Z = \exp(\beta\mu)/\Lambda^3$, Λ is the de Broglie wavelength, V

the volume of the simulation box, N the number of molecules, μ the chemical potential, and $\triangle U$ the change in energy resulting from the molecule insertion or deletion. During the simulations, a 4:1:1 combination of the transfer (insertion/deletion), rigid translation, and rigid rotation MC moves, respectively, was found to be the most efficient choice. To bias insertion and deletion moves, $k = 20$ trial orientations were used.

A cutoff radius of $L_{\min}/2$, where $L_{\min}$ is the minimum dimension of the simulation box was applied to the LJ interactions, and the long-range electrostatic interactions were handled using the Ewald summation technique [8, 83]. In this work, the Ewald summation parameters with real space cutoff $r_c = L_{\min}/2$, convergence parameter $\alpha = 5.6/L_{\min}$ and $k_{\max} = 8$ for the sum in the reciprocal space were used. Since the adsorbent was assumed to be a rigid structure, for computational efficiency the interaction energies between an adsorbed molecule and the clay were pretabulated over a grid (typical resolution is ≈ 0.2 Å in each direction) that is overlaid on the simulation unit-cell. We used linear interpolation between the nearest eight nodes of the grid to determine energy in a given location. Each GCMC simulation consisted of 4×10^7 Monte Carlo steps to guarantee equilibration followed by a production run of 2×10^7 steps. For the estimation of the statistical uncertainty, each production run was divided into 10 consecutive blocks, and the standard deviation of these block averages was calculated [8]. Typically the uncertainties in the average number of adsorbed molecules and total energy were estimated to be from 1 to 2%. The chemical potentials of sorbate molecules needed in the GCMC simulations are computed from the NPT ensemble Monte Carlo simulations, using the Widom's insertion method [8]. The Widom's method was used on a system of no less than 512 particles. The averaging of measured fugacity (Widom's method) in the clay system was performed by an automated convergence analysis based on a block average method. The imposed fugacities of adsorbates are within the measured uncertainties (typically in our

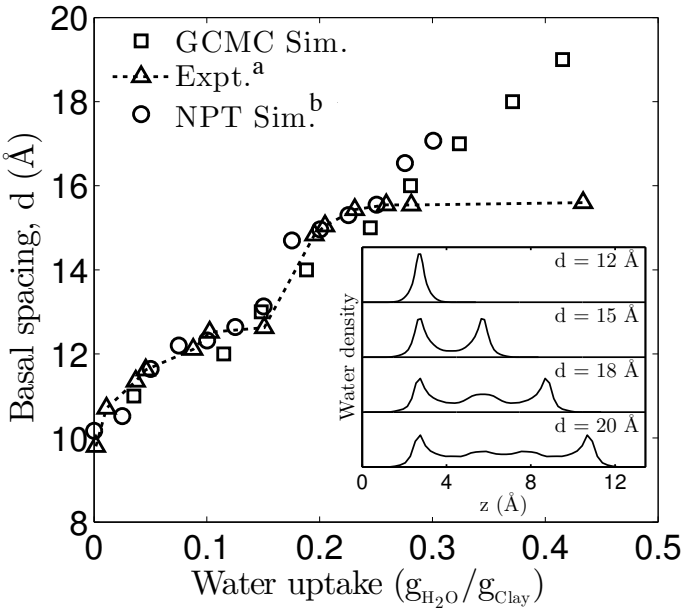

Figure 3.1: Swelling curve for the hydration of Na-montmorillonite derived from GCMC simulations. Inset: z-axis probability distribution functions for water oxygen atoms ([a]Fu et al., 1990; [b]Boek et al., 1995).

simulations, within 2, 5, 1 and 8% for CH_4, CO_2, N_2, and H_2S, respectively).

3.1.2 Results and Discussion

3.1.2.1 Simulation model validation

The originally developed CLAYFF force field parameters are compatible with the flexible simple point charge water and clay models, and uses a slightly different mixing rule than ours [93]. To test the transferability of the CLAYFF parameters while using SPC/E water model and rigid montmorillonite, we performed GCMC simulation in the $\mu_{H_2O}VT$ ensemble. As shown in Fig. 3.1, the swelling behavior of Na-montmorillonite under saturated vapor pressure of water (298.15 K and 1 bar) obtained from the GCMC simulation is in agreement with experimental data [98] and previous simulation results [88]. The density profiles for water molecules (see inset of Fig. 3.1) display various well-defined layers (monolayer, two layer, three layer etc. hydrated states) as a function of basal d-spacing [83, 88]. Montmorillonites are

also known to exhibit oscillatory thermodynamic behavior associated with stepwise, crystalline swelling [83, 90, 92, 94]. The relative stabilities of multiple stable states in a clay system under conditions of constant T, N, and applied pressure can be determined from the minima in the swelling free energy [83, 90, 92, 94]. As the water content is increased in the interlayer, density profiles show that sodium ions are able to hydrate thereby less effectively screening the negatively charged, mutually repelling clay surfaces [83, 88, 95]. While calcium ions [90, 95] also favor clay swelling, cesium ions [92, 95] migrate to and bind to the clay surface and act as a clay swelling inhibitor, i.e., Cs-montmorillonites form stable monolayer hydrates in water. The Poisson-Boltzmann treatment to describe the ionic density profile is not valid in the low states of clay hydrations [135]. The various counterions can form, such as, inner-sphere surface complex consisting of ions which are strongly bound to the tetrahedral substitutions, and outer-sphere surface complex that consists of ions loosely associated with the octahedral substitution [88].

3.1.2.2 Single component adsorption isotherms

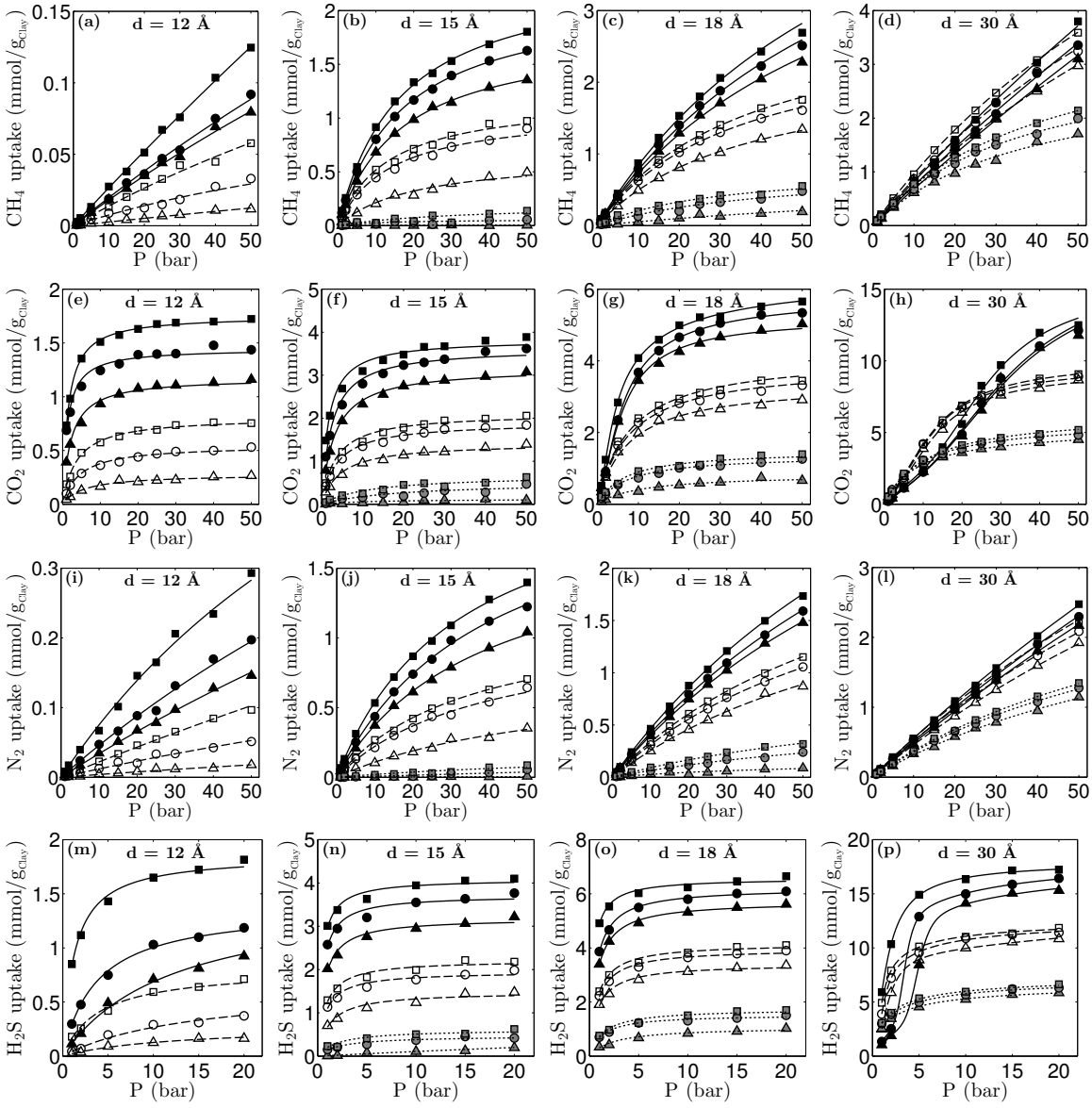

Figure 3.2: Single-component adsorption isotherms for CH_4, CO_2, N_2, and H_2S molecules on Ca- (squares), Na- (circles), and Cs-montmorillonite (triangles) in the presence of 0.2 (black-filled symbols), 0.4 (open symbols), and 0.6 g/cm^3 (gray-filled symbols) of preadsorbed water at 298.15 K as computed from GCMC simulation. The lines are fitting results to the GCMC simulation data.

GCMC simulations were performed to calculate single-component adsorption isotherms for methane, carbon dioxide, nitrogen and hydrogen sulfide in variably hydrated montmorillonite at 298.15 K. The computed isotherms at basal *d*-spacings of 12, 15, 18

and 30 Å are shown in Fig. 3.2 as symbols. In addition, density of water is reported as the mass of water molecules in an interlayer in g per interlayer volume (simulation cell area $\times$ $(d - 6.56)$ Å) in cm^3. For example, in ambient conditions and for basal d-spacings of 12, 15, and 18 Å, we have obtained $\approx$ 0.52, 0.71, and 0.80 g/cm^3 of saturated water densities in the interlayer space of Na-montmorillonite clay, respectively (Fig. 3.1). Over the basal d-spacings and pressure range considered in our study, the presence of increasing water molecules in the clay typically reduced the sorbate adsorption. Fig. 3.2 shows not only the data of our GCMC simulations but also the best fits (lines) to these isotherm curves. Fits of the single-component isotherms using equations, such as Langmuir model [119], are needed to predict the adsorption isotherms of gas mixtures using Ideal Adsorbed Solution Theory (IAST).

The adsorbed amounts of sorbate molecules in the clay-water system under identical conditions show the following general order: $H_2S > CO_2 > CH_4 > N_2$ (see Fig. 3.2). A similar trend was noted in recent experimental studies dedicated to montmorillonite clays [122]. The adsorption isotherms for zeolites also displayed a similar behavior [136]. Polar H_2S has much more affinity to the hydrophilic montmorillonite framework than any other sorbate molecules considered in this work. In this study, we assume that H_2S is adsorbed physically to the clay [136,137]. CO_2 molecule essentially by virtue of its, e.g., quadrupole moment which is about three times larger than N_2 [120, 136], interacts more strongly with the electrostatic field defined by the clay hydrate. CH_4 has zero quadrupole moment, but higher polarizability than N_2 which is responsible for their above observed order [136, 138]. The above adsorption order is affected, however, by the selection of a relatively small basal spacing (e.g., for the case of $d = 12$ Å, $N_2 > CH_4$, possibly due to the comparatively significant quadrupole moment of the former).

In the absence of intermolecular interactions and surface energetic heterogeneity, the isosteric heat [139] should be constant at all loadings. The isosteric heats of

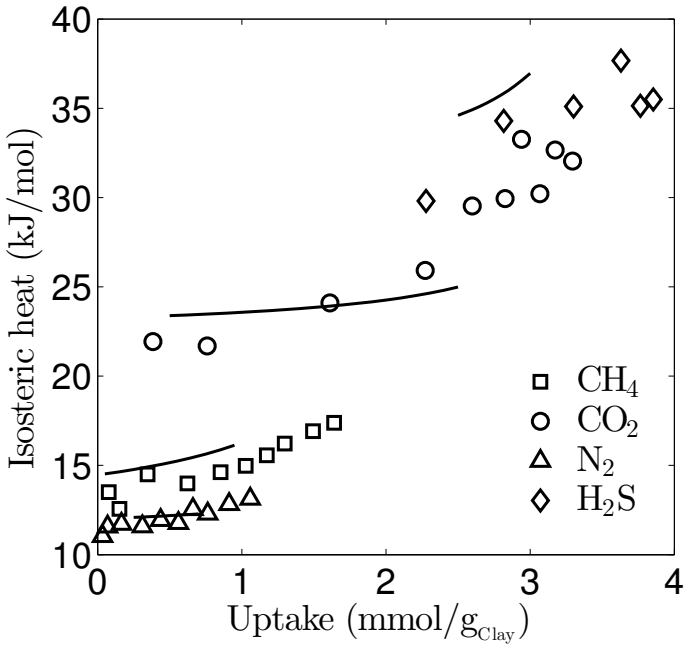

Figure 3.3: Isosteric heats of adsorption for single-component adsorption of CH_4, CO_2, N_2, and H_2S molecules on Na-montmorillonite at 298.15 K, computed using GCMC simulations. The basal *d*-spacing and the preadsorbed water content are 18 Å and 0.4 g/cm^3, respectively.

adsorption of pure adsorbate molecules calculated by using the Clausius-Clapeyron relation (lines) and the fluctuations of the energy (symbols), at a basal *d*-spacing of 18 Å and water content of 0.4 g/cm^3 in the Na-montmorillonite, are shown in Fig. 3.3. It is clear from Fig. 3.3 that the isosteric heats decrease in the order of $H_2S > CO_2 > CH_4 > N_2$, which is attributed mainly to weakening of fluid-solid interaction. This result is also consistent with the preferentially adsorbed order, according to the isotherms presented in Fig. 3.2. We also observe a monotonic increase of the isosteric heats of adsorption of various adsorbates as a function of their loading. This is typically due to favorable adsorbate-adsorbate intermolecular interactions and a relatively homogeneous adsorbent surface. Thus, the stronger the adsorbate-adsorbate interaction, the quicker the curve rising trend.

Other than high uptake capacities, clays exhibit type IV isotherms in the International Union of Pure and Applied Chemistry (IUPAC) classification, for CO_2 and H_2S

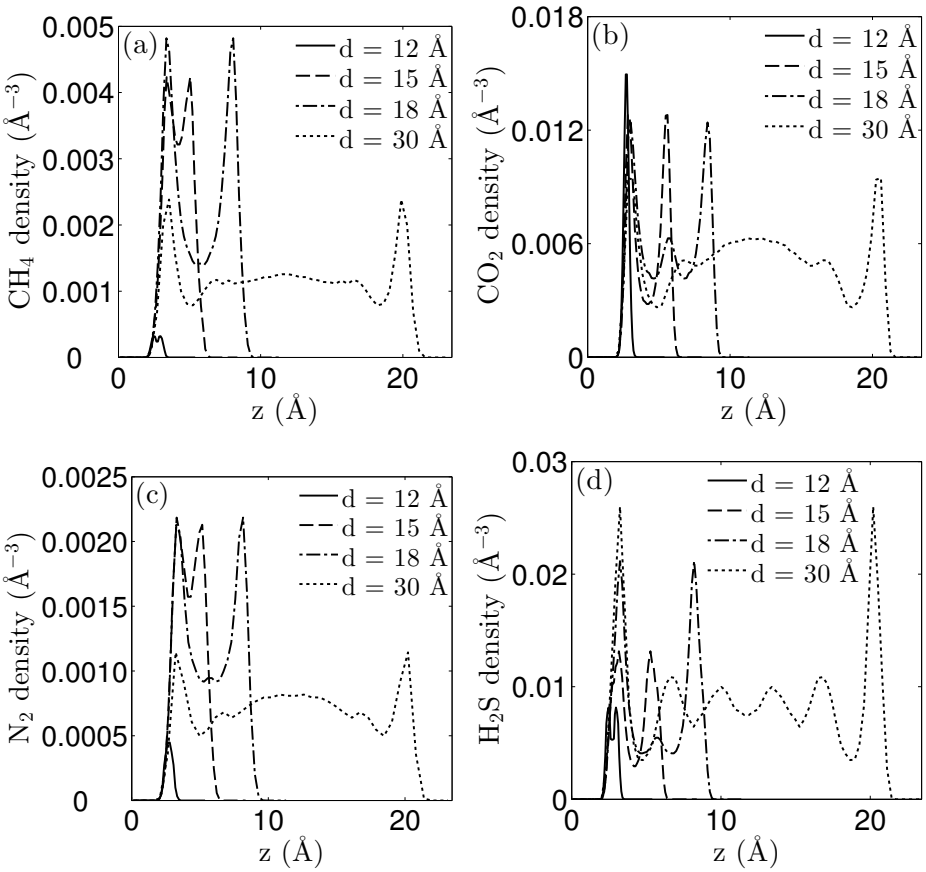

Figure 3.4: Equilibrium distributions of (a) CH_4, (b) CO_2 (carbon atoms), (c) N_2 (nitrogen atoms), and (d) H_2S (sulfur atoms) molecules in Na-montmorillonite at $T = 298.15$ K and $P = 20$ bar as computed from GCMC simulation. The origin corresponds to the clay surface oxygen. The preadsorbed water content is 0.4 g/cm^3.

molecules over the studied pressure range at a basal d-spacing of 30 Å, as illustrated in Figs. 3.2h and 3.2p, respectively. In contrast, adsorption isotherm behavior of CH_4 and N_2 at this basal d-spacing was not significantly different, in comparison with the corresponding ones at small separations (basal d-spacings of 12, 15 or 18 Å). Adsorption isotherm, such as the IUPAC type IV, indicates multilayer formation followed by condensation in the pore [119, 140]. Therefore, to explore the distribution of the various species in the interlayer space of the clay, number density profiles were calculated at bulk pressure $P = 20$ bar and with 0.4 g/cm^3 of preadsorbed water. Fig. 3.4 displays the resulting average density profiles of various sorbate molecules computed along the z-axis (perpendicular to the Na-montmorillonite surface). The distributions

of sorbate molecules indicate layered structures similar to pure hydration states such as 1W, 2W, etc. A very high peak is obtained for the first-layer adsorption in almost all cases, which indicates that the strong solid-fluid interaction causes the molecules to pack much closer to the clay surface. In general, the profiles of CH_4 (Fig. 3.4a) and N_2 (Fig. 3.4c) are very similar over the studied conditions. As seen in Figs. 3.4b and 3.4d, pronounced changes occur in the density profiles of CO_2 and H_2S with increasing basal d-spacing. The distributions indeed confirm that at a basal d-spacing of 30 Å, CO_2 and H_2S molecules significantly accumulate in the middle of the interlayer region of the clay. H_2S molecules form well-defined layered structures and CO_2 molecules also adsorb in multilayers here. The profiles obtained for CH_4 and CO_2 in clays are in good qualitative agreement with previous simulations [103, 106, 108, 113].

Very interestingly, we found that the presence of a high water content in the clay is also favorable for the adsorption of carbon dioxide, for a basal d-spacing of 30 Å (see Figs. 3.2h). The obvious trend in this plot is that, at relatively high pressures, the effect of increased water is to reduce CO_2 uptake. The opposite effect is observed only at relatively low pressures, and even then only the intermediate amounts of water show this enhancement. These predictions could open new possibilities for tuning the adsorption in clays for CO_2 sequestration and other applications. The multilayer adsorption observed in clay hydrates, for example, could potentially provide a mechanism by which to adsorb CO_2 selectively from gas mixtures at conditions close to those existing in subsurface geological reservoirs and cap rocks. Note that, the relevant P/T conditions based on lithostatic pressure and geothermal gradient significantly differ from those used in this work [79, 101, 106, 141]. A similar behavior is predicted for H_2S at a basal d-spacing of 30 Å (see Fig. 3.2p). In graphene-oxide systems, however, the same phenomena have been reported using combined experiment and simulation studies [142]. Jin and Firoozabadi also observed using GCMC molecular simulations that when $P < 40$ bar, CO_2 sorption in montmorillonite clays

at a basal *d*-spacing of 46.56 Å, is comparable for preadsorbed water contents of 0.4 g/cm^3 and 0.6 g/cm^3, and slightly higher sorption is observed at water content of 0.6 g/cm^3 [103]. Loring et al. [143] reported that the sorbed CO_2 concentration in montmorillonite clays increased dramatically at sorbed H_2O concentrations from 0 to 4 mmol/g. Sorbed CO_2 then decreases as sorbed H_2O increases above 4 mmol/g. Adsorption is a process that is a balance between energetic and entropic contributions. Such arguments have been previously used to describe competitive CO_2/CH_4 adsorption in nanoporous materials of varying porosity [144]. Consistent with the above observed adsorption preference for the four adsorbates in terms of polar interactions with the adsorbent, the presence of some water favors adsorption of polar molecules. This factor dominates at low loading, where not much adsorbate is present. However at high loadings, entropic factors dominate because the presence of water occupies pore volume, now not available to the adsorbates.

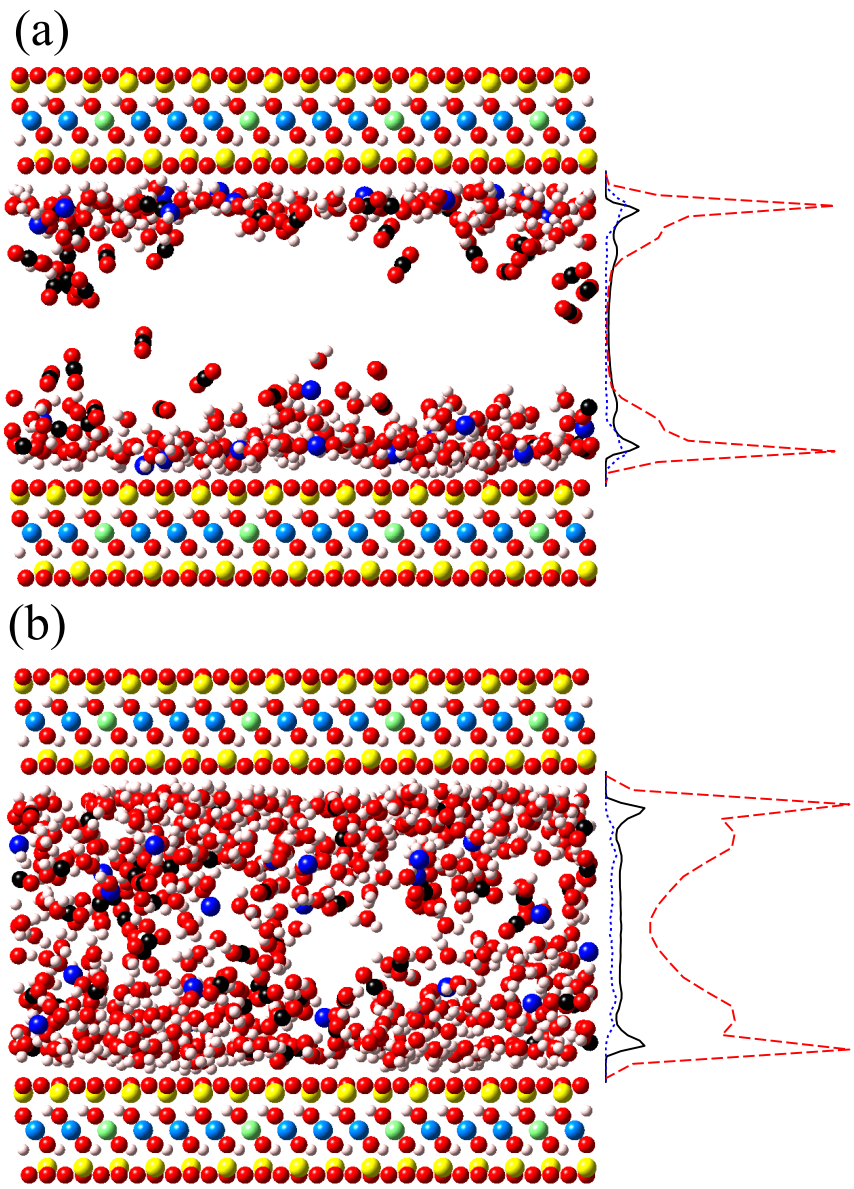

Figure 3.5: Equilibrium snapshots and density profiles of clay mineral-water-CO_2 system with preadsorbed water contents of (a) 0.2 and (b) 0.4 g/cm^3 at $T = 298.15$ K and $P = 10$ bar as obtained from GCMC simulation. In the density profiles, the solid, dashed, and dotted lines represent carbon, water oxygen, and sodium atoms, respectively. Each snapshot shows only one of the two interlayers belonging to the central simulation box, for the purpose of clarity. Color code: O, red; H, white; Si, yellow; Al, light blue; Mg, light green; Na, dark blue; C, black. The basal d-spacing is 30 Å.

To gain some insight into why preadsorbed water molecules increase CO_2 adsorption, we examined the number density profiles from the GCMC simulations. The

distributions of water, CO_2 and counterions in the interlayer of Na-montmorillonite at a basal d-spacing of 30 Å, $P = 10$ bar, and with 0.2 and 0.4 g/cm^3 of preadsorbed water densities are shown in Fig. 3.5. Also shown are visualizations of these average distributions of molecules in the Na-montmorillonite obtained from GCMC simulation. Density profiles of atoms in the interlayer region with a water content of 0.2 g/cm^3 clearly show separation of CO_2 molecules into two distinct layers, similar to that observed for the water molecules. Further, the sodium cations are solvated within the water layer and mostly remain close to the clay surfaces except a few that are distributed near the midplane of the interlayer region (see Fig. 3.5a). Inspection of Fig. 3.5b reveals that at a water content of 0.4 g/cm^3, water molecules form layered structures (> 3W), and as mentioned above, CO_2 molecules get accumulated in the middle of the interlayer space of the clay as well. Most of the sodium ions are well solvated due to the increase of water content in the interlayer. In contrast to low water content behavior, the sodium ions at this condition are mostly displaced away from the clay surfaces.

To better understand the mechanisms responsible for enhanced adsorption, we calculated the isosteric heats of adsorption of CO_2 from fluctuations of the energy [139] in our GCMC simulations. Fig. 3.6 shows the resulting adsorbate-adsorbate (closed symbols) and the fluid-clay contributions (open symbols) to the isosteric heats of adsorption, at a basal d-spacing of 30 Å, and in the presence of 0.2 and 0.4 g/cm^3 of preadsorbed water densities. The analysis of the isosteric heat of adsorption clearly shows that, initially CO_2 molecules adsorbs at high-affinity adsorption sites on the clay and subsequent adsorption is at less-favored sites. The fluid-clay isosteric heats of adsorption are approximately of the same order of magnitude for both water densities at all loadings. Therefore, the amount of CO_2 intake is determined by the fluid-fluid interaction. The observation that fluid-clay interactions will be of lesser importance as the adsorbed amount or pressure of CO_2 increases is very reasonable and can be

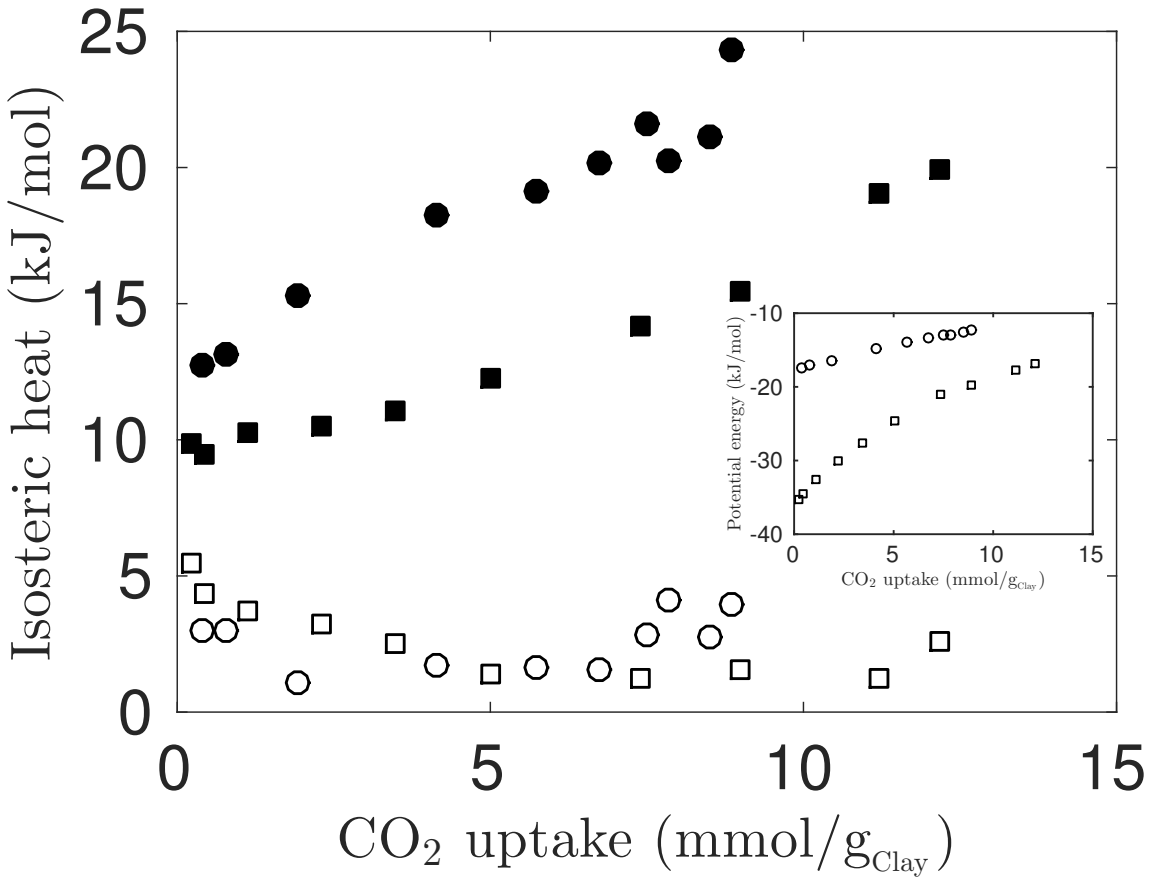

Figure 3.6: Isosteric heats of adsorption (adsorbate-adsorbate) for single-component adsorption of CO_2 molecules on Na-montmorillonite at 298.15 K, computed using GCMC simulations with preadsorbed water contents of 0.2 (filled squares) and 0.4 g/cm^3 (filled circles). The open symbols show the corresponding fluid-clay contribution to the isosteric heats. The inset shows fluid-clay contribution to the potential energy (same symbols as in main figure). The basal d-spacing is 30 Å.

clearly established by plotting the fluid-clay contribution to the potential energy. As shown in the inset to Fig. 3.6, the fluid-clay energetic contribution will weaken upon an increase in the adsorbed amount of CO_2. It is also clear that, the presence of high water content (0.4 g/cm^3) in the clay is energetically less favorable for fluid-clay interaction, in comparison to a water content of 0.2 g/cm^3. On the basis of the above analysis and visual inspection of the snapshots, we arrive at the observation that, this enhanced fluid-fluid interaction due to multilayer adsorption of CO_2 (or H_2S) molecules in the clay increases their uptake with the increase of water content.

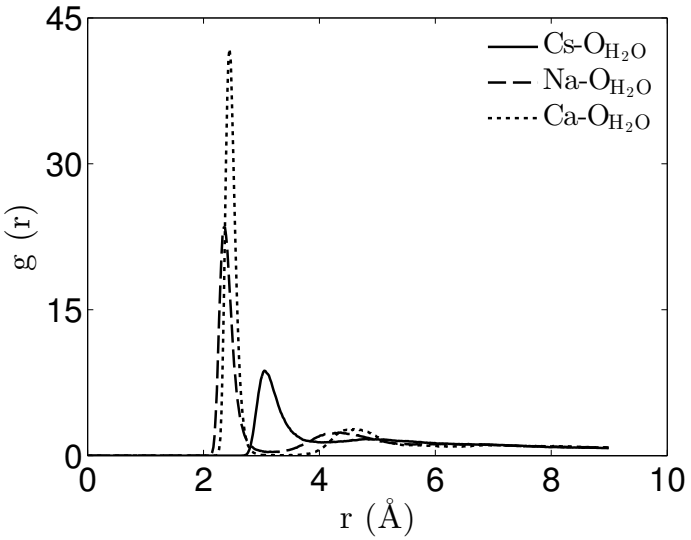

Figure 3.7: The ion-water radial distribution functions for hydrated clays in equilibrium with CH_4 molecules at $T = 298.15$ K and $P = 20$ bar as computed from GCMC simulation. The basal d-spacing and the preadsorbed water content are 15 Å and 0.6 g/cm^3, respectively.

Isotherms calculated for Na-, Cs-, and Ca-montmorillonite clays indicate a strong dependence of adsorption on the interlayer ion identity (Fig. 3.2). Under identical conditions, Ca-montmorillonite clay displays the highest adsorption capacity and Cs-montmorillonite the lowest for all isotherms studied. Note that the extent of water adsorption in clays typically increases with cation-water hydration energy ($Ca^{+2} > Na^+ > Cs^+$) [90, 91, 95]. In dehydrated montmorillonite, the counterions interact strongly with the clay sheets and are trapped within stable potential wells [145]. The radial distribution function between the counterions and tetrahedral oxygen atoms of the dehydrated clay showed strong spatial correlation for Na^+ in comparison to Ca^{+2}. Ions such as Ca^{+2} and Na^+ are able to hydrate in the presence of water, thereby becoming separated from the clay surface [83, 84, 88–90, 92, 94, 145]. In contrast, Cs^+ ions interact strongly with the rigid clay surface than with the mobile water molecules. The occupation of the surface sites may thus render many such sorption sites inaccessible to sorbate reducing the storage capacity of clays. Simulated radial distribution functions describing ion-water oxygen (O_{H_2O}) spatial correlations in the

interlayer of Na-, Cs-, and Ca-montmorillonite at a basal d-spacing of 15 Å, bulk pressure of methane $P = 20$ bar, and in the presence of 0.6 g/cm^3 of preadsorbed water are compared in Fig. 3.7. The ion-water radial distribution functions of Na-montmorillonite clay hydrates containing CO_2 or CH_4 are in fair agreement with previous simulations [100, 106, 113]. According to this distribution, both Ca-O_{H_2O} and Na-O_{H_2O} spatial correlations show sharp peaks in g(r) near ≈ 2.4 Å and ≈ 4.4 Å, evidently because of Ca^{+2} and Na^{+} solvation effects that has been observed also for pure water in montmorillonite [84, 106, 145]. However, this kind of structure in g(r) is relatively weak in the Cs-montmorillonite, suggesting that solvation interactions are important in organizing interlayer water and ions. Typically, sorbate molecules hardly influence [106] the distribution of other atoms.

3.1.2.3 Mixtures adsorption isotherms

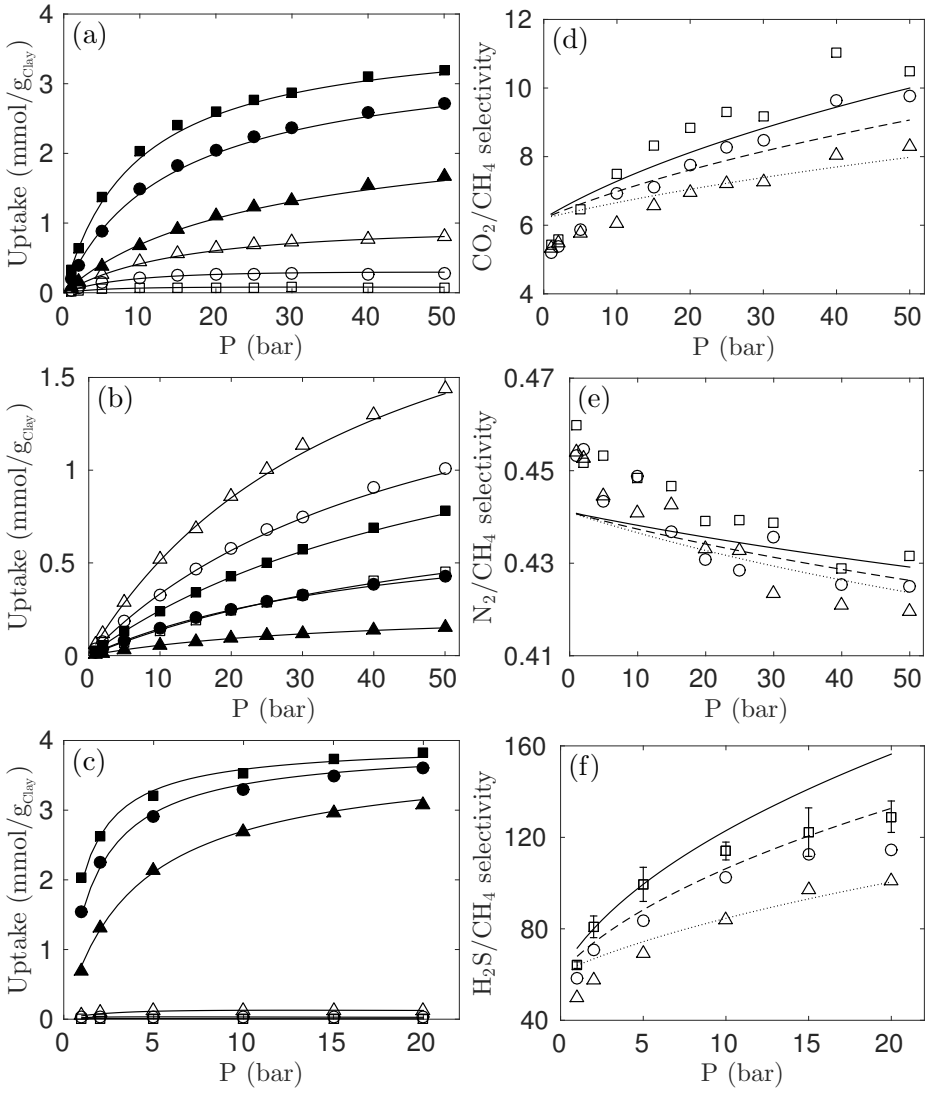

Figure 3.8: Amounts adsorbed for (a) CO_2/CH_4, (b) N_2/CH_4, and (c) H_2S/CH_4 binary mixtures in Na-montmorillonite at 298.15 K, computed using GCMC simulations with CH_4 mole fractions: $y_{CH_4} = 0.2$ (squares), 0.5 (circles), and 0.8 (triangles). The open symbols represent the amount of CH_4 molecules and solid symbols the amount of either CO_2, N_2 or H_2S molecules adsorbed. The solid curves are the predictions of IAST. The basal d-spacing and the preadsorbed water content are 18 Å and 0.4 g/cm^3, respectively. Panels (d)-(f) show selectivities from GCMC simulations (symbols) and IAST (lines) under the same conditions as in panels (a)-(c), respectively: $y_{CH_4} = 0.2$ (squares, solid lines), 0.5 (circles, dashed lines), and 0.8 (triangles, dotted lines).

We also carried out GCMC simulations of adsorption of binary mixtures in Na-montmorillonite clay at 298.15 K. Figs. 3.8a-c show the behavior of CO_2/CH_4, N_2/CH_4, and H_2S/CH_4 binary mixtures, respectively, as computed using GCMC simulations (symbols) at various bulk mole fractions of CH_4, a basal d-spacing of 18 Å, and in the presence of 0.4 g/cm^3 of preadsorbed water. Consistent with the single-component isotherms described in the previous section, Na-montmorillonite selectively adsorbs all sorbate molecules considered, except N_2, relative to CH_4. Our GCMC simulations demonstrate that small amounts of H_2S are able to extremely diminish CH_4 adsorption upon montmorillonite over the studied pressure range (Fig. 3.8c). The smooth lines in Figs. 3.8a-c represent the isotherms predicted by applying IAST to the single-component isotherms described previously. In IAST, there is no constraint on the mathematical formulation of the single-component adsorption isotherm, as the integration can be done analytically or with a numerical procedure (Details can be found in the Supporting Material of our paper [128]).

The qualitative agreement between the GCMC data and the IAST predictions in Fig. 3.8 demonstrates that, the later coupled to the single-component adsorption fits is a good candidate to predict the binary adsorption behavior of gas mixtures in clays across the pressure range studied. The accuracy of IAST here extends to adsorbed mixtures in equilibrium with gas phases that have low or moderate CH_4 partial pressures. The IAST model also predicted the bisolute competitive sorption data favorably in modified montmorillonite [146]. The IAST, however, is known to fail to capture even qualitative features for certain non-ideal mixtures or high pressures [144]. The discrepancy between our calculated adsorption selectivity with IAST and with GCMC simulations is more significant for H_2S/CH_4 selectivity (see below) as the size and electronic character of H_2S and CH_4 are significantly different.

The selectivity is a key and sensitive parameter to quantify the competitive adsorption between two components [74, 120]. For example, ethane selectivity against

methane in montmorillonite reduced with increase in pore size or pressure [74]. The selectivity $S_{A/B}$ in a binary mixture of components A and B is defined as

$$S_{A/B} = \frac{x_A/y_A}{x_B/y_B}, \tag{3.8}$$

where x_A (y_A) and x_B (y_B) are the mole fractions of components A and B in the adsorbed phase (bulk phase), respectively. This quantity can easily be computed directly using results from mixture GCMC simulations as well as using IAST. The resulting GCMC simulation data (symbols) and IAST predictions (lines) are shown in the corresponding Figs. 3.8d-f. All the predicted selectivity values from IAST are in quite good agreement with the simulated values as shown in Figs. 3.8d-f. The IAST predicted H_2S/CH_4 selectivity, displayed the highest deviation among the three cases that we tested here, overestimating the simulation results by factors as high as only ≈ 1.2, at the highest investigated pressure (see Fig. 3.8f). Inspection of Figs. 3.8d and f shows that, CO_2 and H_2S molecules have higher affinity for the montmorillonite surface than methane, a finding consistent with earlier observations. In addition, the simulation and IAST results of CO_2/CH_4 and H_2S/CH_4 selectivities increase as the bulk phase pressure is increased. By contrast, methane has a higher affinity for the montmorillonite surface than nitrogen, and the N_2/CH_4 selectivity decreases as the bulk phase pressure is increased.

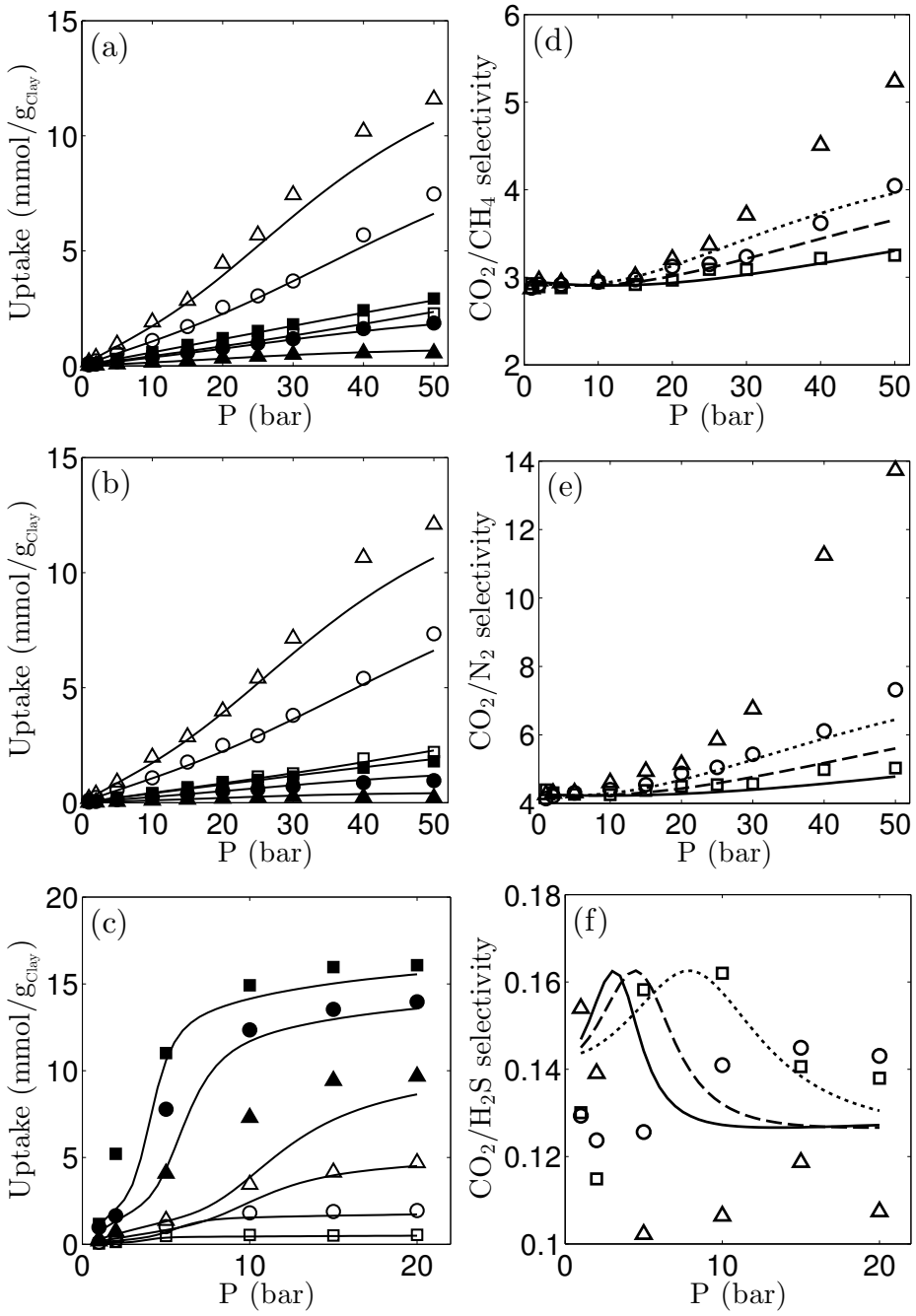

Figure 3.9: Amounts adsorbed for (a) CO_2/CH_4, (b) CO_2/N_2, and (c) CO_2/H_2S binary mixtures in Na-montmorillonite at 298.15 K, computed using GCMC simulations with CO_2 mole fractions: $y_{CO_2} = 0.2$ (squares), 0.5 (circles), and 0.8 (triangles). The open symbols represent the amount of CO_2 molecules and solid symbols the amount of either CH_4, N_2 or H_2S molecules adsorbed. The solid curves are the predictions of IAST. The basal d-spacing and the preadsorbed water content are 30 Å and 0.2 g/cm^3, respectively. Panels (d)-(f) show selectivities from GCMC simulations (symbols) and IAST (lines) under the same conditions as in panels (a)-(c), respectively: $y_{CO_2} = 0.2$ (squares, solid lines), 0.5 (circles, dashed lines), and 0.8 (triangles, dotted lines).

We have seen that at a basal d-spacing of 30 Å, CO_2 and H_2S exhibit multilayer adsorption pattern (see, e.g., Figs. 3.4b and d). In Figs. 3.9a-c, the GCMC simulation results (symbols) of the binary isotherms are compared to IAST (lines) for the adsorption of CO_2/CH_4, CO_2/N_2, and CO_2/H_2S mixtures, respectively, at various bulk mole fractions of CO_2, a basal d-spacing of 30 Å, and in the presence of 0.2 g/cm^3 of preadsorbed water. The selectivity estimates coming from GCMC simulations (symbols) and IAST (lines) are shown in the corresponding Figs. 3.9d-f. The IAST results calculated by neglecting any nonideality, still successfully forecasts the same qualitative behavior in adsorbed amount and selectivity as the GCMC simulations for majority of these binary mixtures. However, the IAST predictions, for example, underestimate the corresponding mixture GCMC results of CO_2/CH_4 and CO_2/N_2 selectivities, as the bulk phase pressure is increased. This is expected, because the performance of IAST has been shown to degenerate for mixtures of molecules of differing sizes and polarities [120, 121]. As a consequence of the significant disparity in the interaction strengths between molecules in a mixture of adsorbates of differing polarities, the average deviations between IAST and GCMC results ranged from 30 to 67% [121]. The selectivity of CO_2/N_2 in the clay system, obtained using IAST, showed the highest deviation among the three cases that we tested, underestimating the simulation results by factors as high as ≈ 2.2, at the highest studied pressure (see Fig. 3.9e). The CO_2/CH_4 selectivity increases as the bulk phase pressure is increased at a basal d-spacing of 30 Å (Fig. 3.9d). This behavior is similar to the above case (Fig. 3.8d), and indeed of potentially great practical importance for CO_2 utilization and storage, such as, in shale gas reservoirs [122]. A comparison between the single-component (Fig. 3.2h) and corresponding mixture isotherm data of CO_2 molecules (Figs. 3.9a-c) indicates that, the multilayer adsorption behavior of CO_2 is almost unaffected by the presence of CH_4, N_2 or H_2S, over the entire range of conditions we considered. Overall, water does not affect the pore-filling mechanisms of various binary mixtures

we studied, by comparison with the corresponding pure ones (compare, e.g., Fig. 3.2 with Figs. 3.8 and 3.9). This is because, as in the case of pure CH_4, CO_2, N_2, and H_2S adsorption in the presence of water (Fig. 3.2), the coadsorption isotherms (Figs. 3.8 and 3.9) are either type I or type IV according to the IUPAC classification.

3.1.3 Conclusions

We have completed extensive grand canonical Monte Carlo simulations to better assess the adsorption behavior of CH_4, CO_2, and their mixtures in Na-, Cs-, and Ca-montmorillonite clays in the presence of water. We have shown that CO_2 uptake and its selectivity over CH_4 in the clays can also be significantly increased with the increase of water content in the interlayer which is attributed to the multilayer adsorption of CO_2 molecules in the clay. This result could open new potentials for tuning the adsorption behavior of clays for CO_2 storage and other applications.

In general, the presence of increasing water molecules in the clay reduced adsorption amounts of pure CH_4 and CO_2, possibly because preadsorbed water render many sorption sites inaccessible to sorbate by occupying the sorption sites or filling interlayer space. The Ca-montmorillonite clay displayed the highest adsorption capacity and the Cs-montmorillonite the lowest, for all single component isotherms studied, consistent with the order of increasing cation hydration energies. The Na-montmorillonite clay preferentially adsorbs CO_2 over CH_4 during both pure component and mixture adsorption, and the CO_2/CH_4 selectivity is observed to increase as the bulk phase pressure is increased. Further, GCMC simulations were performed to estimate the effect of water on the adsorption of N_2/CH_4, H_2S/CH_4, CO_2/N_2, and CO_2/H_2S binary mixtures in Na-montmorillonite clay. We found that water does not seem to affect the pore-filling mechanisms in all investigated binary mixtures, by comparison with the corresponding pure cases.

On the basis of the single-component CH_4, CO_2, N_2, and H_2S isotherms, that

showed the order of adsorption capacities following the order of their polarizabilities, binary mixture adsorption and selectivity were predicted using the IAST. Reasonably good agreement with IAST is obtained for the observed adsorption capacities and selectivities of binary mixtures in the clay.

3.2 MD Study of CH_4, CO_2 and Water Diffusion in Montmorillonite

Molecular dynamics simulations were carried out to study the structural and transport properties of carbon dioxide, methane, and their mixture at 298.15 K in Na-montmorillonite clay in the presence of water. The simulations show that, the self-diffusion coefficients of pure CO_2 and CH_4 molecules in the interlayers of Na-montmorillonite decrease as their loading increases, possibly because of steric hindrance. The diffusion of CO_2 in the interlayers of Na-montmorillonite, at constant loading of CO_2, is not significantly affected by CH_4 for the investigated CO_2/CH_4 mixture compositions. We attribute this to the preferential adsorption of CO_2 over CH_4 in Na-montmorillonite. While the presence of adsorbed CO_2 molecules, at constant loading of CH_4, very significantly reduces the self-diffusion coefficients of CH_4, and relatively larger decrease in those diffusion coefficients are obtained at higher loadings. The preferential adsorption of CO_2 molecules to the clay surface screens those possible attractive surface sites for CH_4. The competition between screening and steric effects leads to a very slight decrease in the diffusion coefficients of CH_4 molecules at low CO_2 loadings. The steric hindrance effect, however, becomes much more significant at higher CO_2 loadings and the diffusion coefficients of methane molecules significantly decrease. Our simulations also indicate that, similar effects of water on both carbon dioxide and methane, increase with increasing water concentration, at constant loadings of CO_2 and CH_4 in the interlayers of Na-montmorillonite.

Our results could be useful, because of the significance of shale gas exploitation and carbon dioxide storage.[2]

3.2.1 Simulation Details

The final configurations obtained from the previous GCMC simulations (details in Section 3.1) were used as the initial configurations in our MD simulations. Equilibration runs of 1 ns were carried out in the NVT ensemble at $T = 298.15$ K, followed by 2 ns production runs in the NVE ensemble. The use of the NVE ensemble ensures that dynamical properties, such as self-diffusion coefficients, are not biased by the extended system algorithms used to produce constant-temperature ensemble [8, 84]. Three independent trajectories each of length 3 ns per simulation were computed to achieve good statistical averages. The equations of motion were integrated using the velocity Verlet algorithm with a time step of 1 fs. Temperature was controlled by a Nośe-Hoover thermostat [8] with a relaxation time of 0.1 ps and a drag value of 1.0. The nonbond terms were handled with a cutoff at 9.5 Å. The extra skin distance for building neighbor lists was set to 2 Å. The long-range van der Waals interactions were included via tail corrections. The long-range part of the electrostatic interactions was treated using the particle-particle particle-mesh (PPPM) method with a precision value of 10^{-5} and a grid order value of 5. The differences of system temperatures from the preset value during NVE production runs were mostly negligible (typically $< 1\%$). To estimate the self-diffusion coefficients of the interlayer species, the Einstein relation was employed:

$$D_{\mathrm{xy}} = \lim_{t\to\infty} \frac{1}{4N_m t} \left\langle \sum_{j=1}^{N_m} [r_j(t) - r_j(0)]^2 \right\rangle, \tag{3.9}$$

where N_m is the number of a selected species, and $r_j(t)$ is the center-of-mass

[2]The following content corresponding to this part appears in our publication in The Journal of Physical Chemistry C [147].

position of the jth species at time t. The self-diffusion coefficients were thus obtained from the linear slope of the mean square displacements (MSDs) as a function of the simulation time. The MSDs of the molecules and ions were evaluated using the 2 ns equilibrated atomic trajectories from the NVE ensemble simulations with 0.8 ps sampling. Different restart points in the analysis were taken to check the convergence. Note that, D_{xy} values in clay interlayer nanopores were insensitive to the size of the simulation cell [148].

3.2.2 Water/CO_2 and Water/CH_4 Binary Mixtures Results

3.2.2.1 Atomic density profiles

In order to explore the distribution of the various species in the interlayer space of the clay, number density profiles were estimated for carbon dioxide and methane in variably hydrated Na-montmorillonite at 298.15 K. The final configurations reported from the GCMC simulation study of adsorption of CO_2 and CH_4 by Na-montmorillonite in the presence of preadsorbed water [128], were used as the initial configurations in our NVT simulations. Fig. 3.10 displays the average density profiles of carbon dioxide (carbon atoms) and methane molecules computed along the z-axis (perpendicular to the Na-montmorillonite surface) at compositions obtained for a bulk pressure of 20 bar each. Figs. 3.10a and c report the profiles of different species adsorbed at a preadsorbed water content of 0.2 g/cm^3, while Figs. 3.10b and d represent a preadsorbed water content of 0.4 g/cm^3. In the Supporting Material of our paper [147], the corresponding distributions of water oxygens and sodium ions in the interlayers of the Na-montmorillonite-water-CO_2 and the Na-montmorillonite-water-CH_4 systems are given.

The profiles of carbon dioxide molecules (Figs. 3.10a and b) exhibit quite similar qualitative agreement to those reported at nonsaturation or fully saturation [103,106, 108]. GCMC and MD simulations by Botan et al. showed that, CO_2 molecules in

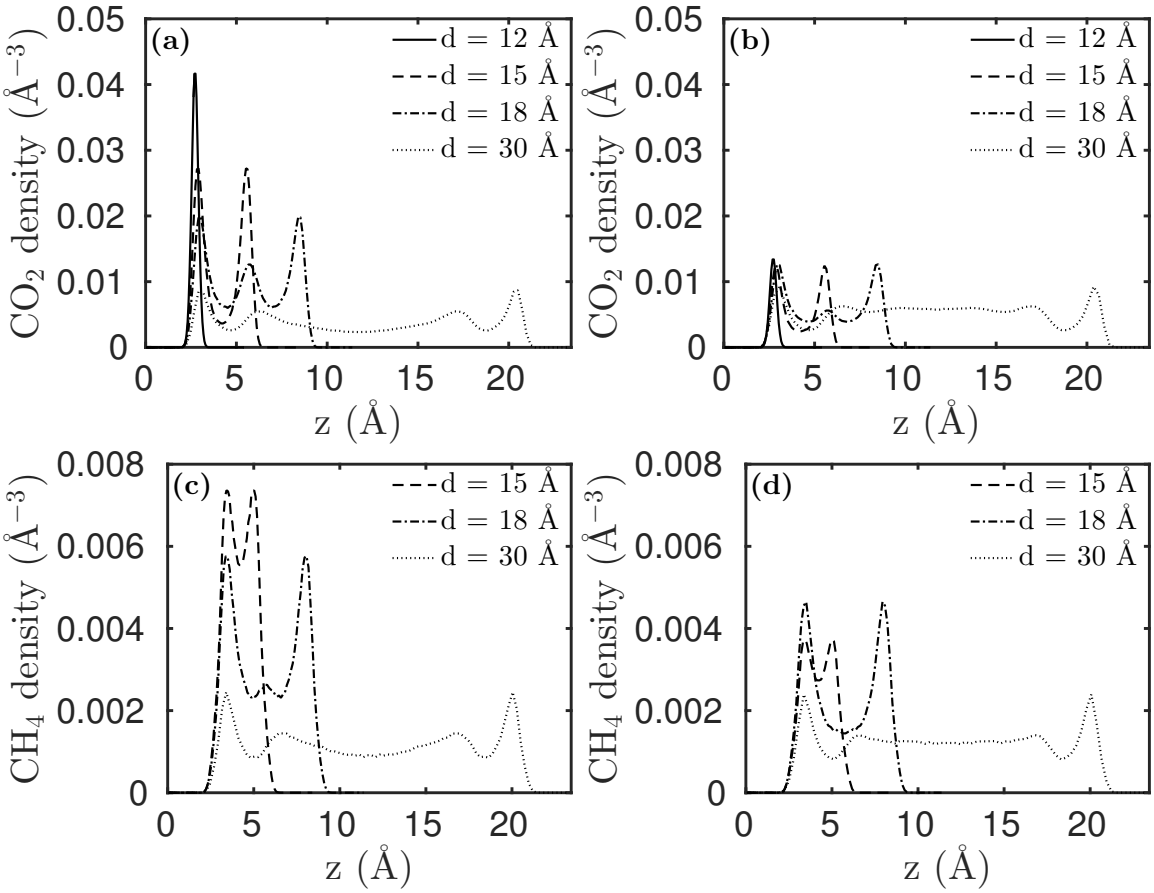

Figure 3.10: Equilibrium distributions of CO_2 (top panels) and CH_4 (bottom panels) molecules in the interlayers of Na-montmorillonite at $T = 298.15$ K and a bulk pressure of 20 bar each. The origin corresponds to the clay surface oxygen. The preadsorbed water contents are 0.2 (left panels) and 0.4 g/cm^3 (right panels).

the interlayers of hydrated Na-montmorillonite hardly influence the distribution of other atoms [106]. This observation is consistent with our simulation results over the investigated basal d-spacings and pressure/loading range. Gibbs ensemble Monte Carlo and NPT simulations, both of which take into account the effects of interlayer volume change, have shown that the incorporation of CO_2 molecules into the interlayer region modified the water and Na^+ profiles, however, mostly due to swelling [108,141]. Fig. 3.10c reports the profiles of methane adsorbed at a preadsorbed water content of 0.2 g/cm^3, which, upon addition of more preadsorbed water molecules, start to form relatively low peaks (Fig. 3.10d). A high peak is obtained for the first-layer adsorption

in almost all cases, which indicates that the strong methane-solid interaction causes the molecules to pack much closer to the clay surface. This behavior is consistent with previous studies on clays under near-surface geological conditions [103, 113] and that observed at basin conditions [101]. Our simulations show that methane molecules in the interlayer region of hydrated Na-montmorillonite hardly affect the distribution of other atoms. It is evident from the simulated profiles that CO_2 molecule essentially by virtue of its, e.g., quadrupole moment, has much more affinity to the hydrophilic montmorillonite framework than the methane molecule. Under identical conditions, therefore, carbon dioxide molecules exhibit a higher adsorbed amount and lie closer to the clay surface than molecules of methane. We find that the adsorption amounts of CH_4 and CO_2 in the clay interlayers, generally increase with an increase in basal spacing or bulk pressure. At low-pressure range (typically, $\lesssim$ 20 bar), however, the adsorption amounts of carbon dioxide and methane in the small pores is higher than that of the corresponding larger ones [125, 128]. This phenomenon is due to the relatively stronger fluid-clay interaction in the small pores, and becomes negligible for high water contents. The presence of increasing water molecules in the clay, in general, reduces adsorption amounts of pure CO_2 (Figs. 3.10a and b) and CH_4 (Figs. 3.10c and d), possibly because preadsorbed water render many sorption sites inaccessible to sorbate by occupying the sorption sites or filling interlayer space. In contrast, with a relatively large basal spacing ($d = 30$ Å), our results (see, e.g., Figs. 3.10a and b) show the favorability of adsorption of CO_2 by montmorillonite at relatively low pressures and intermediate water contents [103, 128]. This enhancement is attributed to the multilayer adsorption of CO_2 molecules in the clay interlayers [103, 128]. In addition, the clay-CO_2 interaction was shown to be more favorable in sub- to single-hydrated montmorillonite systems, when compared to $>$ 2W hydration states [123,141,143]. As an aside, we note that an increase of CO_2 adsorption capacity in the presence of water has been previously reported in adsorbents, such as metal-organic frameworks [149],

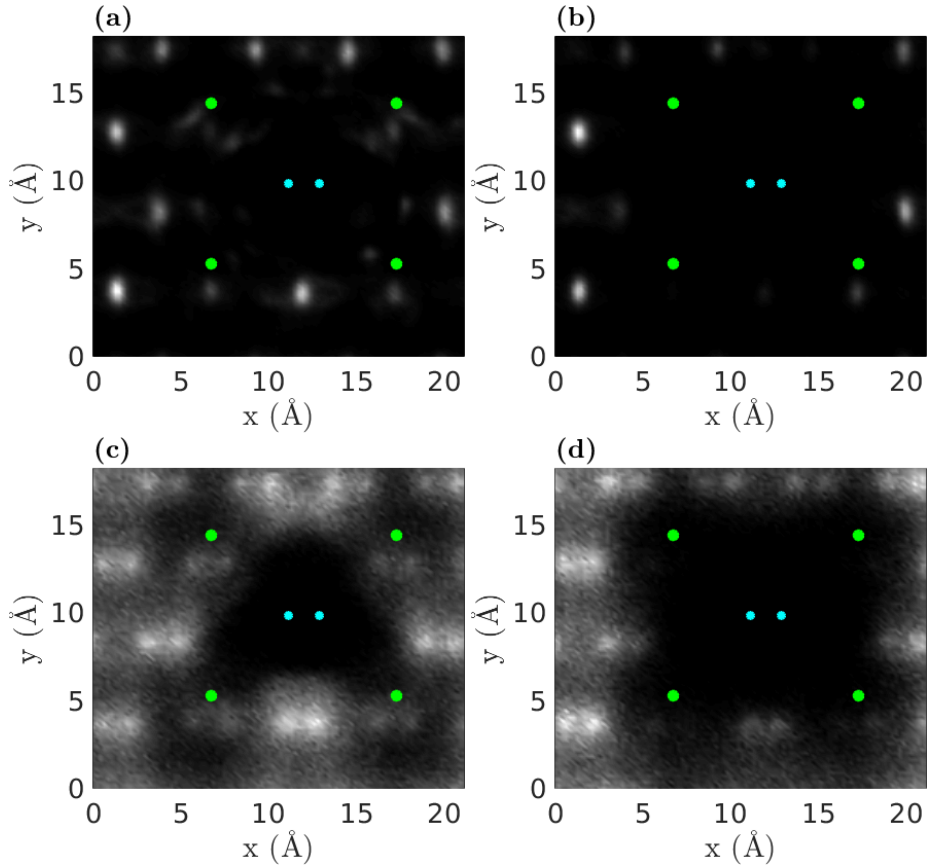

Figure 3.11: Equilibrium distributions (in-plane) of CO_2 (top panels) and CH_4 (bottom panels) molecules at basal $d-$spacings of 12 and 15 Å, respectively, in the interlayers of Na-montmorillonite at $T = 298.15$ K and a bulk pressure of 20 bar each. The preadsorbed water contents are 0.2 (left panels) and 0.4 g/cm^3 (right panels). Bright regions correspond to high density. The positions of Al substitutions in the adjacent tetrahedral layers are indicated by small circles (cyan), while that of Mg in the inner octahedral layer are given by large circles (green).

single-walled carbon nanotubes [150], and mesoporous carbons [151]. At the same time, the increase of water content results in a very slight enhancement of CH_4 intake at low pressures and $d = 30$ Å (Figs. 3.10c and d). The main characteristic of the enhanced intake is an increase in CO_2 or methane density with water, also away in $z-$ direction from the first closest adsorption layer near to the clay surface.

3.2.2.2 Preferential adsorption sites

The main factors affecting the distribution of sorbate molecules in the interlayer include the size and charge of the interlayer cations [152, 153], the positions of the isomorphic substitution in the clay sheets [113, 141, 154], and the ditrigonal ring locations of the basal surfaces [135, 141, 155]. To understand the molecular structure of adsorbed layers of the different species in the interlayers and to identify the preferential adsorption sites on the montmorillonite substrate, we calculated in-plane (xy) density distributions. All these calculations were performed for each molecule found within either the first closest adsorption layer (monolayer) near to the clay surface or away from it (see, e.g., Fig. 3.10). The computed distributions of carbon dioxide (carbon atoms) at a basal $d-$spacing of 12 Å and methane at a basal $d-$spacing of 15 Å, each with compositions obtained for a bulk pressure of 20 bar are shown in Fig. 3.11. In the Supporting Information of our paper [147] we provide the corresponding distributions of water oxygens and sodium ions, and additionally, the distributions of different species in the interlayer of the Na-montmorillonite-water-CO_2 and the Na-montmorillonite-water-CH_4 systems, respectively, at identical conditions and for the other investigated basal spacings.

In good agreement with a most recent simulation study [141], CO_2 molecules clearly tend to locate in the areas away from the charge originating due to the isomorphic tetrahedral substitutions (Figs. 3.11a-b). As the water content increases from 0.2 (Fig. 3.11a) to 0.4 g/cm^3 (Fig. 3.11b), water displaces more CO_2 molecules from near the sites of the substitutions in the clay sheets. The spatial distribution of CO_2 molecules in Na-montmorillonite is also correlated with the positions of the substitutions in the octahedral sheets. However, any correlation of CO_2, water or sodium ions with the octahedral substitutions is expected to disappear [141], e.g., upon reaching the saturated RH. We observe a similar distribution behavior of CO_2 molecules close to the clay surfaces as above at basal $d-$spacings of 15, 18, and 30

Å, while such a behavior persists, albeit to a much lesser extent, also away from the clay surfaces at basal $d-$spacings of 18 and 30 Å. To improve the statistics, the horizontal positions of an atom closer to the clay surface than a fixed cutoff distance are registered and brought back to the unit cell of the sheet (data not shown). Indeed, carbon atom in a CO_2 molecule near the clay surface has a tendency to occupy the ditrigonal cavities. Note that, the specific patterns of sorbate distribution on the clay surface depend on properties such as turbostratic stacking and registry motion of clay sheets [84, 110, 152], which were not considered in our study.

Consistent with previous simulation work [113], methane molecules (Figs. 3.11c-d) and Na^+ ions in the interlayers are positioned in mutually exclusive regions on the clay surfaces, and, as mentioned above, distribution of water coincides with the sodium region. This result can be explained by considering that sodium ion has larger hydration energy than methane, and methane is hydrophobic in nature [113]. As with CO_2, an increase in the water content from 0.2 (Fig. 3.11c) to 0.4 g/cm^3 (Fig. 3.11d), favors displacement of CH_4 molecules by water from near the sites of the substitutions in the clay sheets. We find a similar distribution behavior of CH_4 molecules close to the clay surfaces at basal $d-$spacings of 18 and 30 Å, while such a behavior persists, albeit diminished, also away from the clay surfaces at a basal $d-$spacing of 30 Å. Methane molecule near the clay surface also has a tendency to occupy the ditrigonal cavities. Inspection of the in-plane density maps reveals that while CO_2 molecules occupy more than one of the identified cavity patches of a ditrigonal ring, hydrophobic CH_4 molecules occupy only the larger region.

Monte Carlo simulations by Park and Sposito found methane molecule surmounted by a clathrate-like water structure, while below it was a hexagonal ring of clay surface oxygens [99]. The locations of the isomorphic substitution in the clay sheets, thus tend to inhibit the active involvement of the clay mineral surface in promoting methane clathrate formation (Figs. 3.11c-d). As the tetrahedral negative charge site is closer

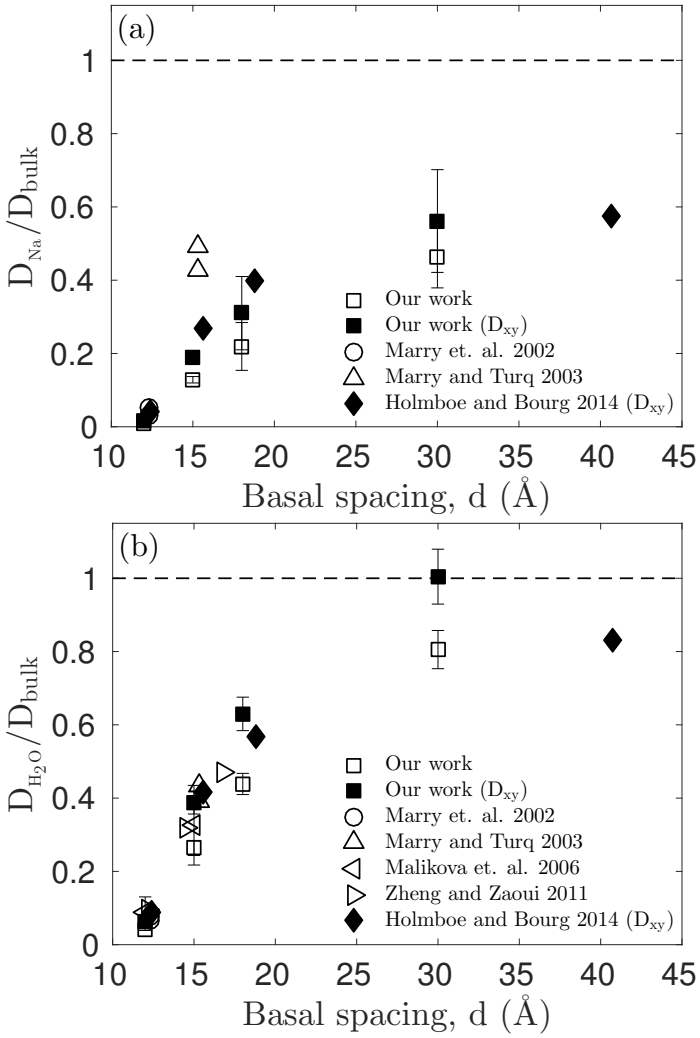

Figure 3.12: Normalized diffusion coefficients of (a) sodium ions and (b) water oxygen atoms in the interlayers of Na-montmorillonite at $T = 298.15$ K and a bulk pressure of 1 bar. Open and filled symbols are for D_{xyz} and D_{xy}, respectively.

to the outer surface and is more effective in confining water, therefore, the final stable state of the mixture hydrate depends on the interplay between those two effects, in addition to the associated swelling [102, 156]. For example, simulation studies reported that the type of clay influence the stability of the smectite-hydrate complexes, being more feasible to form those complexes on octahedrally charged smectites like montmorillonite than in tetrahedrally charged like beidellite [156].

3.2.2.3 Dynamical properties

The diffusion coefficients of ions and water computed at RH of 100% in Na-montmorillonite are provided in Fig. 3.12. In this study, all the reported diffusion coefficients of the different species, unless otherwise stated, are normalized by the corresponding bulk diffusion coefficients at about 298.15 K. The bulk diffusion coefficients of Na^+, H_2O, CO_2, and CH_4, are 1.34, 2.30, 2.00, and 1.49 $\times 10^{-9}$ m^2/s, respectively [99,148,157]. Also, all the diffusion coefficients are compared at constant water content and basal d-spacing, unless otherwise mentioned. The diffusion coefficients of sodium ions are in agreement with those reported using a similar Wyoming type montmorillonite [84,148,155]. The diffusion coefficients of water are also in close agreement with previous simulation results [84, 95, 155, 158] and experimental values obtained using QENS spectroscopy for interlayer water in montmorillonite [158]. These results show that the diffusion coefficient values of water and Na^+ in smectites increase with basal spacing, as expected. The trend is similar to that observed for water diffusion in planar nanopores between silica surfaces [159] and mica surfaces [160]. These simulations showed that the diffusion coefficient values of different species under sub-nanometer confinement decrease by about one to three orders of magnitude under the extreme confinement, as compared to its bulk value.

Fig. 3.13 reports the resulting in-plane diffusion coefficients of CO_2 (carbon atoms) and CH_4 molecules in the interlayers of Na-montmorillonite for the different water/CO_2 and water/CH_4 binary mixture compositions in the pore, respectively, outputted by the GCMC simulations [128]. Tables in the Supporting Information in our paper [147] report the corresponding diffusion coefficients of water oxygens and sodium ions, in addition to the diffusion coefficients of these different adsorbates in each simulation. The diffusion coefficients of sodium ions and water in mixtures are also smaller than their corresponding bulk values typically due to the confinement effect of clay surfaces. Similar to previous studies [106,108,113], at a relatively low constant loading of

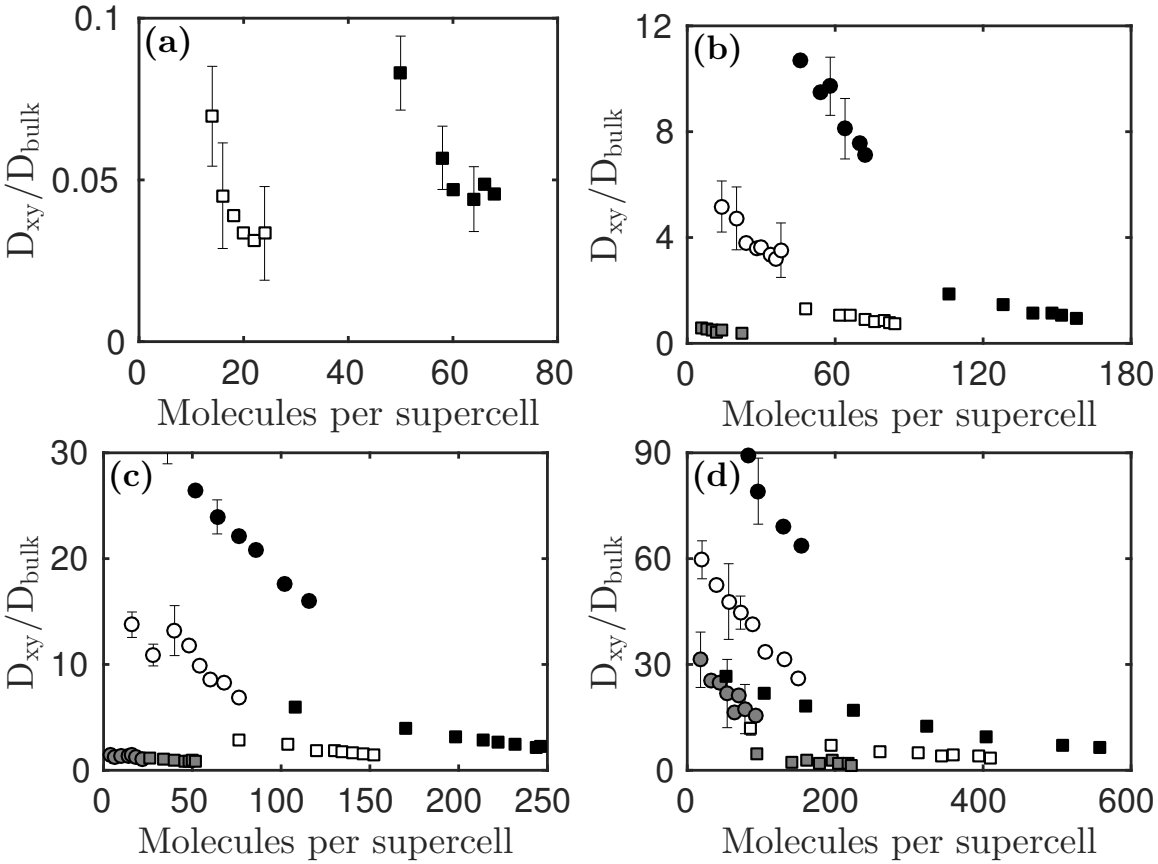

Figure 3.13: Normalized diffusion coefficients of CO_2 (squares) and CH_4 (circles) molecules in the interlayers of Na-montmorillonite at $T = 298.15$ K and for different water/CO_2 and water/CH_4 binary mixture compositions in the pore, respectively. The basal $d-$spacings are (a) 12, (b) 15, (c) 18, and (d) 30 Å, and the preadsorbed water contents are 0.2 (black-filled), 0.4 (open), and 0.6 g/cm^3 (gray-filled).

CO_2 or CH_4 in the clay interlayers, the diffusion coefficients of sodium ions and water decrease due to the less hydrated environment. This is because water molecules effectively screen the surface charges and this effect increases with water. The increase in loading of CO_2 or CH_4 in the clay interlayers, even further reduced those diffusion coefficients. The deviation of the diffusion coefficients of water at $d = 30$ Å from this general behavior is due to, e.g., its multilayer adsorption.

The diffusion coefficients for CO_2 decrease as its loading increases (Fig. 3.13), possibly because of steric hindrance. Similarly, at constant loading of CO_2, the diffusion coefficients for CO_2 decrease as water concentration increases. While, under identical

conditions, the diffusion coefficients for CO_2 increase with the increase in the interlayer space. A similar behavior is observed for methane, in the case of water/CH_4 mixture (Fig. 3.13). The self-diffusion results of CO_2 and CH_4 match the type I behavior as classified by Kärger and Pfeifer [161]. The transport of CO_2 and CH_4 across nanoporous materials showed that steric hindrance causes a decrease in self-diffusion coefficient as loading increases for each substance [161–163]. The self-diffusion coefficient of very strongly adsorbed CO_2 molecules near the surface -OH groups of the solid silica substrate and in zeolite imidazolate frameworks (ZIFs) displayed, however, a maximum at intermediate loadings, while that of pure hydrocarbons in nanopores typically decreased with increasing loading [74, 161–165]. It is seen that in most cases the diffusion coefficients of CO_2 and CH_4 under our employed different conditions are larger than their corresponding bulk values. As in previous simulations [74, 113], we attribute this very high mobility of CO_2 and CH_4 molecules to the less hydrated environment in the clay interlayer. The diffusion coefficients of CO_2 and CH_4 molecules are below their corresponding bulk values due to factors such as the confinement effect of clay surfaces and increase of loading. Previous simulations reported D_{xy}/D_{bulk} of CO_2 from ≈ 0.03 to 0.3, for the different water/CO_2 compositions, basal d-spacings in the range ≈ 12 (1W) to 15 Å (2W), and basin conditions [106, 108]. Incidentally, we also find that D_{xy}/D_{bulk} values of CO_2 are ≈ 0.03 and 0.4 for $d = 12$ and 15 Å, respectively, at their highest considered water contents and loadings (see Figs. 3.13a-b). Additionally, the lateral diffusion coefficient for CO_2 at $d \approx 18$ Å attained a value (4.23×10^{-9} m^2/s) comparable to that measured for diffusion of CO_2 in bulk water [108], which is again consistent to our work (see our results for a water content of 0.6 g/cm^3, in Fig. 3.13c). Note that the diffusion coefficients of both water and CO_2 increased with increasing loading of CO_2, due to the associated expansion of the interlayer space, which is not explicitly included in our simulations [108].

The diffusion coefficients of CH_4 molecules and, to a very lesser extent, CO_2

molecules at constant loadings of CH_4 and CO_2 molecules in the clay interlayers, respectively, are mostly much higher than those of corresponding water molecules, which is in good accordance with the observed fluid-clay interaction energies [125,128] and preferential adsorption to the pore walls (see Fig. 3.10) [103, 106, 108, 113, 125, 128, 141]. Previous simulations reported the ratios of diffusion coefficients of CO_2 to water for states $\leq$ 3W, in the range $\approx$ 0.5 to 4.0 [106, 108], which are close to our corresponding results at the highest considered water contents (e.g., 0.4 and 0.6 g/cm^3). Likewise, high diffusion coefficients of methane in the clay interlayer are in agreement with those reported by Rao and Leng [113]. For example, they obtained the ratios of diffusion coefficients of CH_4 to water for states from $\approx$ 2W to 3W, in the range $\approx$ 0.5 to 16.0, consistent with our results, e.g., at the highest considered water contents (0.4 and 0.6 g/cm^3). Such high values of the diffusion coefficient of methane in dry clay samples have also been reported [66,74]. Cha et al. reported the dissociation of methane hydrate at about ambient temperature and lower pressure ($<$ 50 bar) in the presence of bentonite, which is mainly Na-montmorillonite, than observed for the same process in water alone [104]. The diffusion coefficients of methane we obtained are lower than its bulk value and that of water, e.g., at $d = 18$ Å and the highest studied water content (0.6 g/cm^3). This suggests that only that system contains a stable methane clathrate [99, 104]. At the same conditions, the diffusion coefficients of CO_2 are, however, of the same order of magnitude as its bulk value and that of water, making the montmorillonite-CO_2 hydrate a relatively less stable system. Furthermore, MSDs and adsorption energies supported that smectite-methane hydrate complexes are more stable than smectite-CO_2 complexes [156].

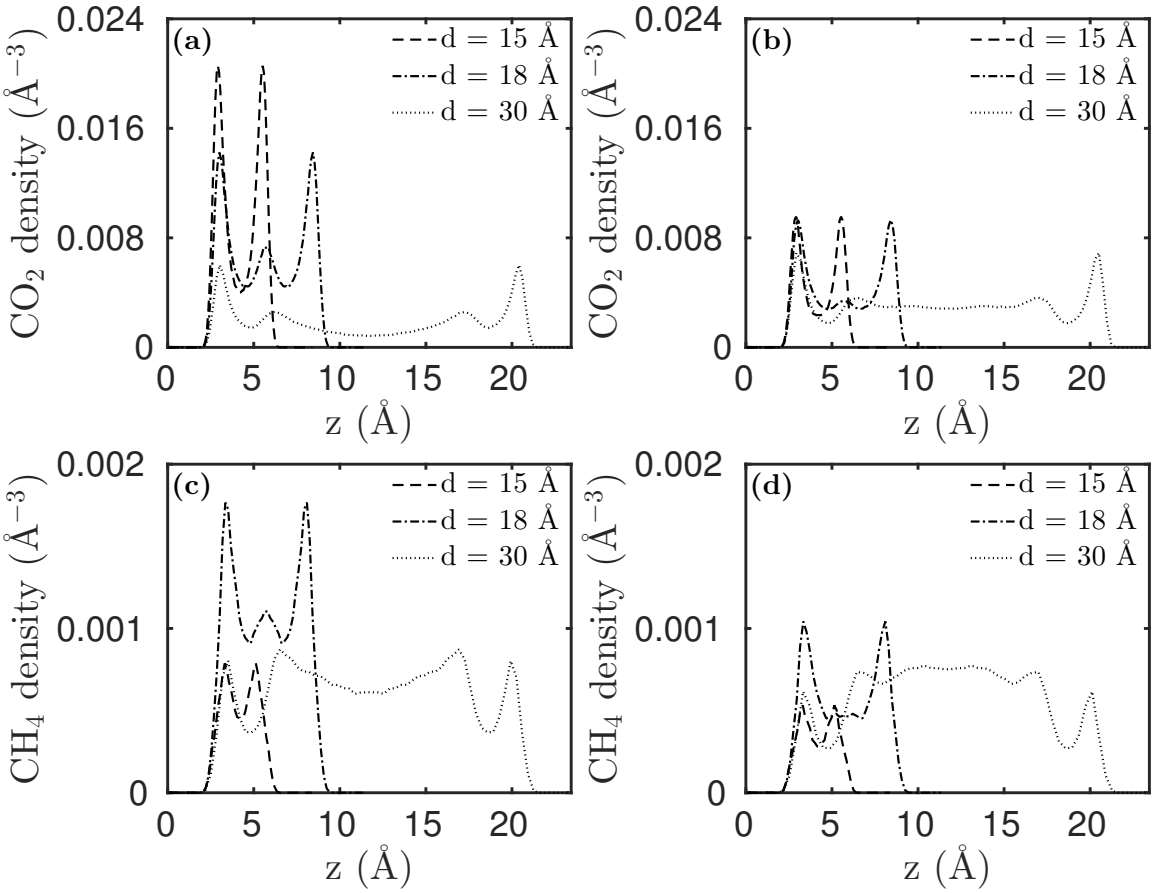

Figure 3.14: Equilibrium distributions of CO_2 (top panels) and CH_4 (bottom panels) molecules in the interlayers of Na-montmorillonite at $T = 298.15$ K and in contact with equimolar CO_2/CH_4 mixture at a bulk pressure of 20 bar. The origin corresponds to the clay surface oxygen. The preadsorbed water contents are 0.2 (left panels) and 0.4 g/cm^3 (right panels).

3.2.3 Water/CO_2/CH_4 Ternary Mixture Results

3.2.3.1 Atomic density profiles

We also carried out MD simulations of ternary water/CO_2/CH_4 mixture in Na-montmorillonite clay at 298.15 K. The final configurations obtained from the GCMC simulation study of adsorption of, e.g., equimolar CO_2/CH_4 binary mixture by Na-montmorillonite in the presence of preadsorbed water [128], were used as the initial configurations in our NVT simulations. In order to examine the distributions of the different species in the interlayer space of the clay, number density profiles were

calculated for carbon dioxide and methane in variably hydrated Na-montmorillonite. Fig. 3.14 shows the average density profiles of carbon dioxide (carbon atoms) and methane molecules in the ternary mixture computed along the z-axis at compositions obtained for a bulk pressure of 20 bar (equimolar CO_2/CH_4). Figs. 3.14a and c represent the profiles of the various species adsorbed at a preadsorbed water content of 0.2 g/cm^3, while Figs. 3.14b and d report a preadsorbed water content of 0.4 g/cm^3. The distributions confirm that carbon dioxide and methane molecules here form well-defined layered structures, similar to the behavior of pure hydration states or binary mixture (water/CO_2 and water/CH_4) profiles described in the previous section.

The density profiles of carbon dioxide and methane molecules in the ternary mixture demonstrate that the clay material has high adsorption selectivity for carbon dioxide over methane (Fig. 3.14). This observation is consistent with our simulation results over the studied basal d-spacings and pressure/loading range. A recent study also reported that, CO_2 molecules with enhanced adsorption strength are able to competitively replace CH_4 molecules within the clay samples in their dehydrated states [125]. The features of the profiles of the binary mixtures are mostly conserved for the ternary mixture, over the studied conditions. For example, the presence of increasing water molecules in the clay, in general, reduces adsorption amounts of both CO_2 (see Figs. 3.14a and b) and CH_4 (see Figs. 3.14c and d) molecules. The density profiles show that the favorability of adsorption of CO_2 and, to a very lesser extent, CH_4 by montmorillonite observed in the binary mixture case (see Fig. 3.10), at relatively low pressures, intermediate water contents, and large basal d-spacings is retained during the ternary mixture adsorption. Similarly, the density peak of CO_2 is closer to the clay surface, as compared with CH_4. A notable exception is that, methane molecules now have lower densities near to the clay surface relative to the bulk at $d = 30$ Å. This result is expected due to the stronger affinity of water and CO_2 towards the surface than methane [125].

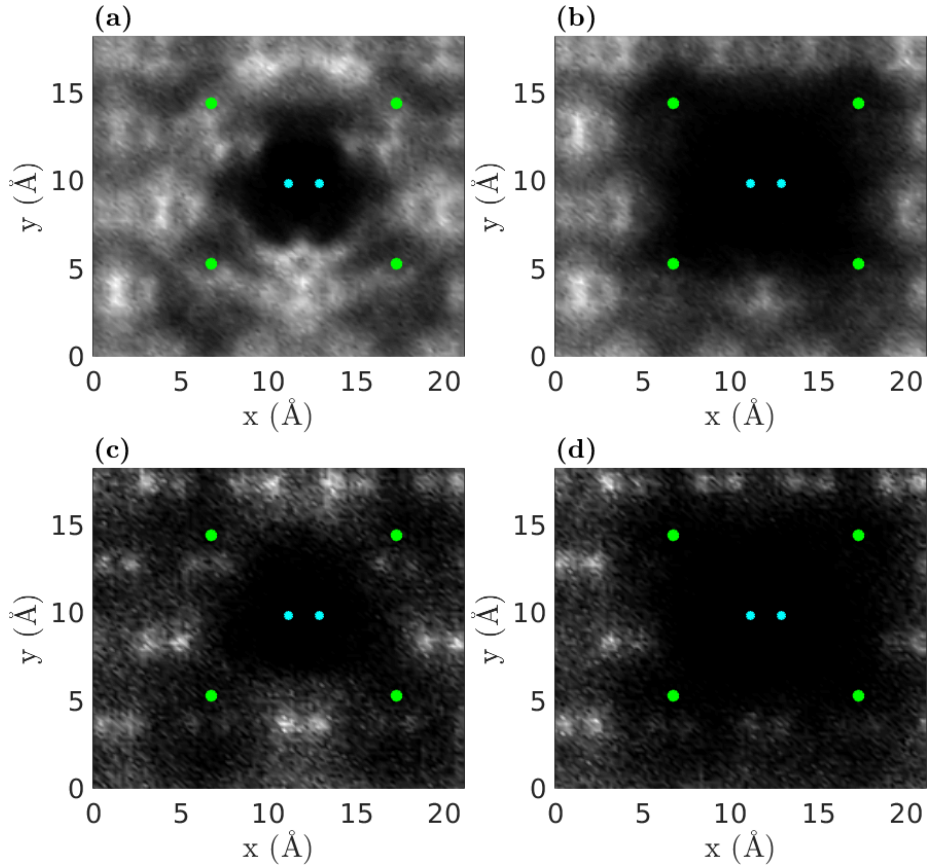

Figure 3.15: Equilibrium distributions (in-plane) of CO_2 (top panels) and CH_4 (bottom panels) molecules at a basal $d-$spacing of 15 Å in the interlayers of Na-montmorillonite at $T = 298.15$ K and in contact with equimolar CO_2/CH_4 mixture at a bulk pressure of 20 bar. The preadsorbed water contents are 0.2 (left panels) and 0.4 g/cm^3 (right panels). Bright regions correspond to high density. The positions of Al substitutions in the adjacent tetrahedral layers are indicated by small circles (cyan), while that of Mg in the inner octahedral layer are given by large circles (green).

3.2.3.2 Preferential adsorption sites

To identify the preferential adsorption sites on the montmorillonite substrate of the different species of the ternary mixture, we calculated in-plane (xy) density distributions. All these calculations were performed again for each molecule found within either the first closest adsorption layer (monolayer) near to the clay surface or away from it (see, e.g., Fig. 3.14). The computed distributions of carbon dioxide (carbon atoms) and methane molecules in the ternary mixture at a basal $d-$spacing of 15 Å

and compositions obtained for a bulk pressure of 20 bar (equimolar CO_2/CH_4) are shown in Fig. 3.15. Again the Supporting Information in our paper [147] provides the corresponding distributions of water oxygens and sodium ions, and additionally, the distributions of different species in the interlayer of the Na-montmorillonite-water-CO_2-CH_4 system at identical conditions and for the other investigated basal spacings.

The in-plane distributions of water and sodium ions in the ternary mixture are hardly different from the corresponding binary mixture cases (see previous section). The distribution of methane coincides with the CO_2 region showing that these molecules can coexist near the clay plane. Besides that, methane and carbon atom in a CO_2 molecule near the clay surface again have a tendency to occupy the ditrigonal cavities. An increase in the water content from 0.2 (Figs. 3.15a and c) to 0.4 g/cm^3 (Figs. 3.15b and d), favors displacement of both CO_2 and CH_4 molecules by water from near the sites of the substitutions in the clay sheets. We observe a similar distribution behavior of CO_2 and CH_4 molecules close to the clay surfaces at basal $d-$spacings of 18 and 30 Å, while such a behavior persists, albeit diminished, also away from the clay surfaces at a basal $d-$spacing of 30 Å for CH_4, and at basal $d-$spacings of 18 and 30 Å for CO_2 molecules.

3.2.3.3 Dynamical properties

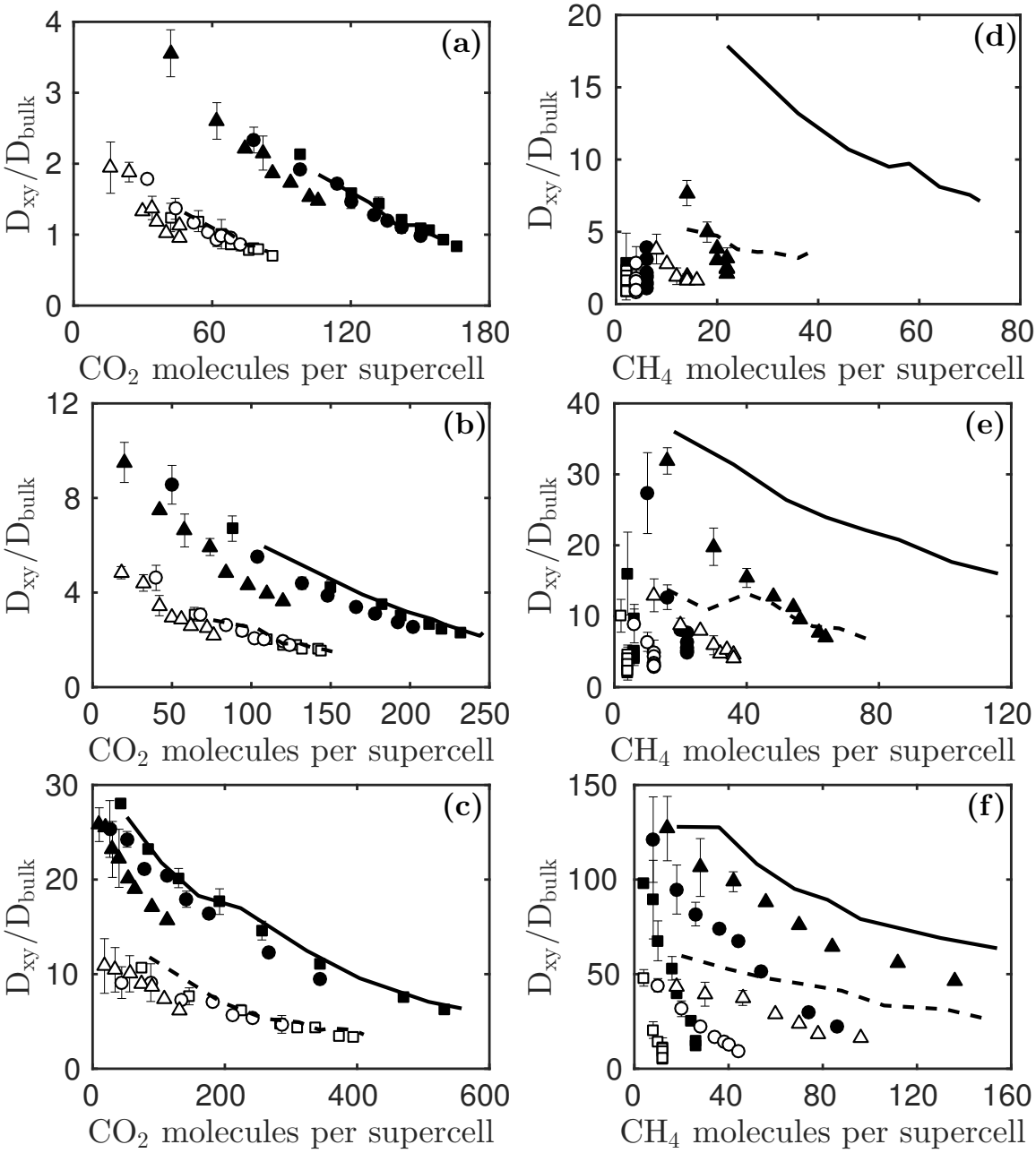

Figure 3.16: Normalized diffusion coefficients of CO_2 (left panels) and CH_4 (right panels) molecules in the interlayers of Na-montmorillonite at $T = 298.15$ K and for different water/CO_2/CH_4 ternary mixture compositions in the pore. The mole fractions of methane in bulk phase are 0.2 (squares), 0.5 (circles) and 0.8 (triangles). The basal $d-$spacings are 15 (top panels), 18 (middle panels), and 30 Å (bottom panels), and the preadsorbed water contents are 0.2 (filled) and 0.4 g/cm^3 (open). The solid (water content of 0.2 g/cm^3) and dashed (water content of 0.4 g/cm^3) lines represent the diffusion coefficients of CO_2 and CH_4 molecules in the interlayers of Na-montmorillonite for the corresponding water/CO_2 and water/CH_4 binary mixtures, respectively (see Fig. 3.13).

The main factors affecting the self diffusion coefficients of sorbate molecules in a confined mixture include such as steric hindrance of the motion of a tagged particle by neighboring particles as the loading increases, momentum transfer correlations, wherein the faster diffusing species diffuses slower in the mixture relative to the pure component and the slower diffusing species diffuses faster in the mixture than in the pure state, and preferential adsorption [162, 164–166]. Fig. 3.16 reports the in-plane diffusion coefficients of CO_2 and CH_4 molecules in the interlayers of Na-montmorillonite for the different water/CO_2/CH_4 ternary mixture compositions in the pore outputted by the GCMC simulations [128]. The most important observation from our results is that the diffusion of CO_2, at constant loading of CO_2, is not much affected by CH_4 for the investigated mixture compositions (Figs. 3.16a-c). The response of CH_4 to CO_2, at constant loading of CH_4 is, however, quite different from the effect of CH_4 on CO_2, because the presence of adsorbed CO_2 reduces very significantly the diffusion coefficients of CH_4 (Figs. 3.16d-f), and relatively larger decrease in those diffusion coefficients are seen at higher loadings.

In montmorillonite, CO_2 is very strongly preferred over CH_4 at all loadings, as is evident from the above described density profiles. Therefore, the adsorbed mixture is very CO_2 dominant relative to CH_4 and the distribution of CO_2 in the mixture is not dramatically perturbed from that of the pure fluid due to loading of methane. For this reason, the diffusion coefficients of CO_2 in the pore is largely independent of the fraction of CH_4 in the bulk mixture, at all investigated conditions (see Figs. 3.16a-c). As demonstrated above, CO_2 molecules introduced in the interlayer space of the Na-montmorillonite preferentially adsorb to the surface and in this way screen these possible attractive sites for CH_4. Such a microscopic behavior should lead to an enhancement of the diffusion coefficient of methane molecules. However, CO_2 molecules simultaneously occupy and/or crowd the interlayer space and consequently reduce the effective diffusing space for CH_4 molecules. At low loadings of CO_2, the combination

of both effects leads to a very slight decrease in the self-diffusion coefficients of CH_4 molecules, while at higher loadings of CO_2, the steric hindrance effect becomes much more important and the self-diffusion coefficients of CH_4 molecules significantly decrease (see Figs. 3.16d-f). Water preferentially binds to the clay surface and therefore, similar effects of water on both carbon dioxide and methane, increase with increasing water content at constant loadings of CO_2 and CH_4. Because of this, the diffusion coefficients of CO_2 and CH_4 decrease with an increase in water content (see Fig. 3.16). A similar observation was reported for the self-diffusion properties of CH_4 in CO_2/CH_4 binary mixtures within NaY zeolite [167]. In contrast, ZIFs showed that the diffusivity of CH_4 is essentially independent of the loading of CO_2 in the CO_2/CH_4 mixture, while CO_2 diffusivity significantly decreased with an increase in loading of CH_4 due to differences in adsorption site preferences [165]. Also, carbon dioxide enhanced the self-diffusion coefficients of hydrocarbons such as methane and butane possibly by decreasing their diffusion activation energies due to the competitive adsorption of carbon dioxide on the pore surfaces [164, 168].

3.2.4 Conclusions

We have completed extensive molecular dynamics simulations to better assess the diffusion behavior of CO_2, CH_4, and their mixture in Na-montmorillonite clay in the presence of water at 298.15 K. The simulations show that the self-diffusion data of CO_2 and CH_4 match the type I behavior as classified by Kärger and Pfeifer [161]. The diffusion coefficients of CO_2 in the interlayers of Na-montmorillonite decrease as its loading increases, possibly because of steric hindrance. Similarly, at constant loading of CO_2, the diffusion coefficients for CO_2 decrease as water content increases. While, under same conditions, the diffusion coefficients for CO_2 increase with increasing basal *d*-spacing. Similar behavior is seen with the diffusion behavior of methane in Na-montmorillonite in the presence of water. The self-diffusion coefficients of water

and sodium ions, at a relatively low constant loading of CO_2 or CH_4 in the clay interlayers, decrease due to the less hydrated environment where water molecules less effectively screen the surface charges. The increase in loading of CO_2 or CH_4 in the clay interlayers, even further reduced those diffusion coefficients. The diffusion coefficients of CH_4 molecules and, to a very lesser extent, CO_2 molecules are mostly much higher than those of water molecules, which is in good accordance with the observed fluid-clay interaction energies [125, 128] and preferential adsorption close to the pore walls [103, 106, 108, 113, 125, 128, 141].

An important finding is that the diffusion coefficients of CO_2 in the interlayers of Na-montmorillonite, at constant loading of CO_2, are not much affected by CH_4 for the investigated CO_2/CH_4 mixture compositions. Through careful analysis of the atomic density profiles of the different species in the interlayers, we attribute this to the preferential adsorption of CO_2 over CH_4 in Na-montmorillonite. While the presence of adsorbed CO_2 molecules, at constant loading of CH_4, very significantly reduces the self-diffusion coefficients of methane, and relatively larger decrease in those diffusion coefficients are observed at higher loadings. The preferential adsorption of CO_2 molecules to the clay surface screens those possible attractive surface sites for CH_4 which may lead to an enhancement of the diffusion coefficient of methane molecules. However, CO_2 molecules simultaneously occupy the interlayer region and, therefore, decrease the effective diffusing space for CH_4 molecules. The interplay of both effects leads to a very slight decrease in the self-diffusion coefficients of CH_4 molecules at low loadings of CO_2. The steric hindrance effect becomes much more significant at higher loadings of CO_2 and the self-diffusion coefficients of methane molecules significantly decrease. The simulations show that similar effects of water on both carbon dioxide and methane, increase with increasing water content, at constant loadings of CO_2 and CH_4 in the interlayers of Na-montmorillonite.

Chapter 4

Concluding Remarks

4.1 Summary

To sum up, the main theme of the conducted research was to benefit from the advancements in molecular simulation tools to build efficient and representative subsurface reservoirs models. The dissertation started with an overview about the principles of Monte Carlo (MC) and Molecular Dynamics (MD) methods in addition to the motivations behind this work.

The first part of the conducted research deals with the fundamentals of MC molecular simulation by looking for smart strategies in order to overcome the time scale challenge in coupling with flow simulators. For the quest towards coupling, a set of early rejection schemes capable of accelerating MC simulations to speeds that can reach the double was successfully demonstrated. Moreover, a novel reconstruction and reweighting extrapolation scheme was introduced which allows the usage of offline Markov chains table to efficiently obtain relevant thermophysical properties in a way that a single simulation can replace multiple ones in few seconds. Furthermore, polynomial chaos expansions were used to replace long MC simulations and helped in optimizing single-site Lennard-Jones (LJ) parameters for many molecules of interest.

In the second part, adsorption isotherms for pure methane, carbon dioxide, and their mixtures for bulk pressures up to 50 bar were reported in the presence of various

preadsorbed water quantities for different basal sizes at room temperature. Moreover, the self-diffusion coefficients of the various molecules in these systems were determined via MD simulations using the final configurations from the equilibrium studies.

4.2 Future Directions

In terms of accelerating MC molecular simulations, despite the effort invested in developing the techniques presented in the previous parts a vast room of improvement is still ahead. For instance, all the proposed methodologies are so far only designed to simulate systems of structureless LJ particles of the same type. Therefore, it is indispensable to extend these methods to be able to account for more complicated scenarios such as: structured molecules, molecules with partial charges, and systems with multi-components. For example, we are currently working on these methods to simulate LJ model plus quadrupolar term [169, 170]. This model is capable of accurately representing more complex molecules such as carbon dioxide, water, and many others. In addition, there exist other strategies that can enhance the performance of the offline table technique that were not discussed in this dissertation, such as the sparse table method [171, 172].

As for shale gas studies, the possibilities and areas of investigation are tremendous. For instance, the work is in progress to study the effect of the solid charge on the adsorption quantities and dynamics of these adsorbed molecules [173]. Furthermore, in this work the inorganic constituent of shale is only studied while looking into the contribution of the missing organic part (kerogen) [174–176] is equally important.

REFERENCES

[1] A. Firoozabadi. *Thermodynamics of Hydrocarbon Reservoirs*. McGraw Hill Professional, USA, 1999.

[2] A. Firoozabadi. *Thermodynamics and Applications of Hydrocarbons Energy Production*. McGraw Hill Professional, USA, 2015.

[3] P. Ungerer, V. Lachet, and B. Tavitian. Applications of molecular simulation in oil and gas production and processing. *Oil Gas Sci. Technol.*, 61(3):387–403, 2006.

[4] P. Ungerer, C. Nieto-Draghi, V. Lachet, A. Wender, A. di Lella, A. Boutin, B. Rousseau, and A. H. Fuchs. Molecular simulation applied to fluid properties in the oil and gas industry. *Mol. Simulat.*, 33(4–5):287–304, 2007.

[5] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, USA, 1989.

[6] R. L. Rowley. *Statistical Mechanics for Thermophysical Property Calculations*. Prentice Hall, USA, 1994.

[7] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, California, USA, 2000.

[8] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, USA, 2001.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[10] A. Lyubartsev, A. Martsinovski, S. Shevkunov, and P. Vorontsov-Velyaminov. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, 96(2):1776–1785, 1992.

[11] E. Marinari and G. Parisi. Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.*, 19:451–458, 1992.

[12] C. J. Geyer and E. A. Thompson. Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Assoc.*, 90(431):909–920, 1995.

[13] F. A. Escobedo and J. J. de Pablo. Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers. *J. Chem. Phys.*, 105(10):4391–4394, 1996.

[14] D. Frenkel. Speed-up of Monte Carlo simulations by sampling of rejected states. *P. Natl. Acad. Sci.*, 101:17571–17575, 2004.

[15] J. S. Wang and R. Swendsen. Non-universal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[16] A. Kadoura, A. Salama, and S. Sun. A conservative and a hybrid early rejection schemes for accelerating Monte Carlo molecular simulation. *Mol. Phys.*, 112(19):2575–2586, 2014.

[17] A. Kadoura, A. Salama, and S. Sun. Speeding up Monte Carlo molecular simulation by a non-conservative early rejection scheme. *Mol. Simulat.*, 42(3):241–241, 2015.

[18] A. Z. Panagiotopoulos. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.*, 61(4):813–826, 1987.

[19] A. Panagiotopoulos, N. Quirke, M. Stapleton, and D. Tildesley. Phase equilibria by simulation in the Gibbs ensemble: Alternative derivation, generalization and application to mixture and membrane equilibria. *Mol. Phys.*, 63(4):527–545, 1988.

[20] B. Smit, P. Smedt, and D. Frenkel. Computer simulations in the Gibbs ensemble. *Mol. Phys.*, 68(4):931–950, 1989.

[21] A. Z. Panagiotopoulos. Direct determination of fluid phase equilibria by simulation in the Gibbs ensemble: A review. *Mol. Simulat.*, 9(1):1–23, 1992.

[22] B. Efron. Bootstrap methods: Another look at the Jackknife. *Ann. Statist.*, 7:1–26, 1979.

[23] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, USA, 1999.

[24] R. Ghanem and J. Red-Horse. Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Fluid Dyn. Res.*, 133:137–144, 1999.

[25] O. M. Knio and O. P. Maitre. Uncertainty propagation in CFD using polynomial chaos decomposition. *Physica D*, 38:616–640, 2006.

[26] I. Sraj, M. Iskandarani, A. Srinivasan, W. C. Thacker, J. Winokur, A. Alexanderian, C. Lee, S. S. Chen, and O. M. Knio. Bayesian inference of drag parameters using AXBT data from Typhoon Fanapi. *Mon. Weather Rev.*, 141(7):2347–2367, 2013.

[27] I. Sraj, K. T. Mandli, O. M. Knio, C. N. Dawson, and I. Hoteit. Uncertainty quantification and inference of Manning's friction coefficients using DART buoy data during the Tōhoku tsunami. *Ocean Model.*, 83:82–97, 2014.

[28] A. Alexanderian, J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knio. Global sensitivity analysis in an ocean general circulation model: A sparse spectral projection approach. *Comput. Geosci.*, 16(3):757–778, 2012.

[29] R. G. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer, New York, 1991.

[30] O. P. Le Maitre and O. M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, New York, 2010.

[31] O. P. Le Maitre, O. Knio, H Najm, and R. G. Ghanem. A stochastic projection method for fluid flow. I. Basic formulation. *J. Comput. Phys.*, 173:481–511, 2001.

[32] O. P. Le Maitre, H. N. Najm, P. P. Pebay, R. G. Ghanem, and O. M. Knio. Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.*, 29(2):864–889, 2007.

[33] H. N. Najm, B. Debusschere, Y. Marzouk, S. Widmer, and O. P. Le Maitre. Uncertainty quantification in chemical systems. *Int. J. Numer. Meth. Eng.*, 80(6):789–814, 2009.

[34] B. D. Phenix, J. L. Dinaro, M. A. Tatang, J. W. Tester, J. B. Howard, and G. J. McRae. Incorporation of parametric uncertainty into complex kinetic mechanisms: Application to hydrogen oxidation in supercritical water. *Combust. Flame*, 112:132–146, 1998.

[35] M. T. Reagan, H. N. Najm, P. P. Pbay, O. M. Knio, and R. G. Ghanem. Quantifying uncertainty in chemical systems modeling. *Int. J. Chem. Kinet.*, 37(6):368–382, 2005.

[36] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio. Uncertainty quantification in MD simulations. Part I: Forward propagation. *Multiscale Model. Simul.*, 10(4):1428–1459, 2012.

[37] F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio. Uncertainty quantification in MD simulations. Part II: Bayesian inference of force-field parameters. *Multiscale Model. Simul.*, 10(4):1460–1492, 2012.

[38] F. Rizzi, R. Jones, B. J. Debusschere, and O. M. Knio. Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. Part I: Sensitivity to physical parameters of the pore. *J. Chem. Phys.*, 138:194104, 2013.

[39] F. Rizzi, R. Jones, B. J. Debusschere, and O. M. Knio. Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. Part II: Uncertain potential parameters. *J. Chem. Phys.*, 138:194105, 2013.

[40] P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. Data driven predictive molecular dynamics for nanoscale flow simulations under uncertainty. *J. Phys. Chem. B*, 117(47):14808–14816, 2013.

[41] M. Salloum, K. Sargsyan, R. Jones, H. N. Najm, and B. Debusschere. Quantifying sampling noise and parametric uncertainty in atomistic-to-continuum simulations using surrogate models. *Multiscale Model. Simul.*, 13(3):953–976, 2015.

[42] A. Kadoura, A. Siripatana, S. Sun, O. M. Knio, and I. Hoteit. Single-site Lennard-Jones models via polynomial chaos surrogates of Monte Carlo molecular simulation. *J. Chem. Phys.*, 144:214301, 2016.

[43] N. Wiener. The homogeneous chaos. *Amer. J. Math.*, 60(4):897–936, 1938.

[44] E. Lemmon, M. McLinden, and D. Friend. NIST Chemistry Webbook, NIST Standard Reference Database 69, 2005. http://webbook.nist.gov.

[45] S. I. Sandler. *An Introduction to Applied Statistical Thermodynamics*. Wiley, USA, 2010.

[46] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61(23):2635–2638, 1988.

[47] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 63:1658, 1989.

[48] J. R. Errington and A. Z. Panagiotopoulos. Phase equilibria of the modified Buckingham exponential-6 potential from Hamiltonian scaling grand canonical Monte Carlo. *J. Chem. Phys.*, 109(3):1093–1100, 1998.

[49] G. M. Torrie and J. P. Valleau. Monte Carlo study of a phase-separating liquid mixture by umbrella sampling. *J. Chem. Phys.*, 66(4):1402–1408, 1977.

[50] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.

[51] J. P. Valleau. The Coulombic phase transition: Density-scaling Monte Carlo. *J. Chem. Phys.*, 95(1):584–589, 1991.

[52] J. P. Valleau. Density-scaling: A new Monte Carlo technique in statistical mechanics. *J. Comput. Phys.*, 96(1):193–216, 1991.

[53] J. P. Valleau. Density-scaling Monte Carlo study of subcritical Lennard-Jonesium. *J. Chem. Phys.*, 99(6):4718–4728, 1993.

[54] J. P. Valleau. Temperature-and-density-scaling Monte Carlo: Methodology and the canonical thermodynamics of Lennard- Jonesium. *Mol. Simulat.*, 31(4):223–253, 2005.

[55] J. P. Valleau. Temperature-and-density-scaling Monte Carlo: Isothermal-isobaric thermodynamics of Lennard-Jonesium. *Mol. Simulat.*, 31(4):255–275, 2005.

[56] K. Kiyohara. Thermodynamic scaling Gibbs ensemble Monte Carlo: A new method for determination of phase coexistence properties of fluids. *Mol. Phys.*, 89(4):965–974, 1996.

[57] A. Kadoura, S. Sun, and A. Salama. Accelerating Monte Carlo molecular simulations by reweighting and reconstructing Markov chains: Extrapolation of canonical ensemble averages and second derivatives to different temperature and density conditions. *J. Comput. Phys.*, 270(4):70–85, 2014.

[58] S. Sun, A. Kadoura, and A. Salama. An efficient method of reweighting and reconstructing Monte Carlo molecular simulation data for extrapolation to different temperature and density conditions. *Procedia Computer Science*, 18:2147–2156, 2013.

[59] A. Kadoura, A. Salama, and S. Sun. Switching between the NVT and NpT ensembles using the reweighting and reconstruction scheme. *Procedia Computer Science*, 51:1259–1268, 2015.

[60] A. Saad, A. Kadoura, and S. Sun. Multi-scale coupling between Monte Carlo molecular simulation and Darcy-scale flow in porous media. *Procedia Computer Science*, 80:1354–1363, 2016.

[61] D. W. Peaceman. Interpretation of well-block pressures in numerical reservoir simulation. *SPE Journal*, 18(3):183–194, 1978.

[62] A. L. Lee, M. H. Gonzalez, and B. E. Eakin. The viscosity of natural gases. *J. Petrol. Technol.*, 18(8):997–1000, 1966.

[63] C. E. Neuzil. How permeable are clays and shales? *Water Resour. Res.*, 30(2):145–150, 1994.

[64] D. J. K. Ross and R. M. Bustin. The importance of shale composition and pore structure upon gas storage potential of shale gas reservoirs. *Mar. Pet. Geol.*, 26(6):916–927, 2009.

[65] D. Liu, P. Yuan, H. Liu, T. Li, D. Tan, W. Yuan, and H. He. High-pressure adsorption of methane on montmorillonite, kaolinite and illite. *Appl. Clay Sci.*, 85:25–30, 2013.

[66] Z. Zhai, X. Wang, X. Jin, L. Sun, J. Li, and D. Cao. Adsorption and diffusion of shale gas reservoirs in modeled clay minerals at different geological depths. *Energ. Fuel.*, 28(12):7467–7473, 2014.

[67] A. Busch, S. Alles, Y. Gensterblum, D. Prinz, D. N. Dewhurst, M. D. Raven, H. Stanjek, and B. M. Krooss. Carbon dioxide storage potential of shales. *Int. J. Greenh. Gas Con.*, 2:297–308, 2008.

[68] R. T. Cygan, V. N. Romanov, Myshakin, and E. M. Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field. *J. Phys. Chem. C*, 116(24):13079–13091, 2012.

[69] V. N. Romanov. Evidence of irreversible CO_2 intercalation in montmorillonite. *Int. J. Greenh. Gas Con.*, 81:220–226, 2013.

[70] P. R. Jeon, J. Choi, T. S. Yun, and C. H. Lee. Sorption equilibrium and kinetics of CO_2 on clay minerals from subcritical to supercritical conditions: CO_2 sequestration at nanoscale interfaces. *Chem. Eng. J.*, 225:705–715, 2014.

[71] B. Sawhiney. Selective sorption and fixation of cations by clay minerals: A review. *Clay. Clay Miner.*, 20:93–100, 1972.

[72] C. Volzone, J. G. Thompson, A. Melnitchenko, J. Ortiga, and S. R. Palethorpe. Selective gas adsorption by amorphous clay-mineral derivatives. *Clay. Clay Miner.*, 47(5):647–657, 1999.

[73] A. L. Cheng and W. L. Huang. Selective adsorption of hydrocarbon gases on clays and organic matter. *Org. Geochem.*, 35(4):413–423, 2004.

[74] A. Sharma, S. Namsani, and J. K. Singh. Molecular simulation of shale gas adsorption and diffusion in inorganic nanopores. *Mol. Simulat.*, 41(5):414–422, 2015.

[75] M. Josh, L. Esteban, C. Delle Piane, J. Sarout, D. N. Dewhurst, and M. B. Clennell. Laboratory characterisation of shale properties. *J. Pet. Sci. Eng.*, 88:107–124, 2012.

[76] W. Yuan, Z. Pan, X. Li, Y. Yang, C. Zhao, L. D. Connell, S. Li, and J. He. Experimental study and modeling of methane adsorption and diffusion in shale. *Fuel*, 117:509–519, 2014.

[77] F. Javadpour. Nanopores and apparent permeability of gas flow in mudrocks (shales and siltstone). *J. Can. Pet. Technol.*, 48(8):16–21, 2009.

[78] U. S. Energy Information Administration. *Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formations in 41 Countries Outside the United States*. U. S. Energy Information Administration, Washington, DC, USA, 2013.

[79] Q. Rao, Y. Xiang, and Y. Leng. Molecular simulations on the structure and dynamics of water-methane fluids between Na-montmorillonite clay surfaces at elevated temperature and pressure. *J. Phys. Chem. C*, 117(27):14061–14069, 2013.

[80] M. Firouzi, K. Alnoaimi, A. Kovscek, and J. Wilcox. Klinkenberg effect on predicting and measuring helium permeability in gas shales. *Int. J. Coal Geol.*, 123:62–68, 2014.

[81] N. T. Skipper, F. R. C. Chang, and G. Sposito. Monte Carlo simulation of interlayer molecular structure in swelling clay minerals. I: Methodology. *Clay. Clay Miner.*, 43(3):285–293, 1995.

[82] B. J. Teppen, K. Rasmussen, P. M. Bertsch, D. M. Miller, and L. Schaefer. Molecular dynamics modeling of clay minerals. 1. Gibbsite, kaolinite, pyrophyllite, and beidellite. *J. Phys. Chem. B*, 101(9):1579–1587, 1997.

[83] M. Chávez-Páez, K. Van Workum, L. de Pablo, and J. J. de Pablo. Monte Carlo simulations of Wyoming sodium montmorillonite hydrates. *J. Chem. Phys.*, 114(3):1405–1413, 2001.

[84] V. Marry, P. Turq, T. Cartailler, and D. Levesque. Microscopic simulation of structure and dynamics of water and counterions in a monohydrated montmorillonite. *J. Chem. Phys.*, 117(7):3454–3463, 2002.

[85] C. Volzone and J. Ortiga. Influence of the exchangeable cations of montmorillonite on gas adsorptions. *Process Saf. Environ.*, 82(2):170–174, 2004.

[86] P. Giesting, S. Guggenheim, A. F. Koster van Groos, and A. Busch. Interaction of carbon dioxide with Na-exchanged montmorillonite at pressures to 640 bar: Implications for CO_2 sequestration. *Int. J. Greenh. Gas Con.*, 8:73–81, 2012.

[87] M. Hu, Z. Cheng, M. Zhang, M. Liu, L. Song, Y. Zhang, and J. Li. Effect of calcite, kaolinite, gypsum, and montmorillonite on Huadian oil shale kerogen pyrolysis. *Energ. Fuel.*, 28(3):1860–1867, 2014.

[88] E. S. Boek, P. V. Coveney, and N. T. Skipper. Monte Carlo molecular modeling studies of hydrated Li-, Na-, and K-smectites: Understanding the role of potassium as a clay swelling inhibitor. *JACS*, 117(50):12608–12617, 1995.

[89] S. H. Park and G. Sposito. Monte Carlo simulation of total radial distribution functions for interlayer water in Li-, Na-, and K-montmorillonite hydrates. *J. Phys. Chem. B*, 104(19):4642–4648, 2004.

[90] M. Chávez-Páez, L. de Pablo, and J. J. de Pablo. Monte Carlo simulations of Ca-montmorillonite hydrates. *J. Chem. Phys.*, 114(24):10948–10953, 2001.

[91] E. J. M. Hensen, T. J. Tambach, A. Bliek, and B. Smit. Adsorption isotherms of water in Li-, Na-, and K-montmorillonite by molecular simulation. *J. Chem. Phys.*, 115(7):3322– 3329, 2001.

[92] H. D. Whitley and D. E. Smith. Free energy, energy, and entropy of swelling in Cs, Na, and Sr montmorillonite clays. *J. Chem. Phys.*, 120(11):5387–5395, 2004.

[93] R. T. Cygan, J. J. Liang, and A. G. Kalinichev. Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *J. Phys. Chem. B*, 108(4):1255–1266, 2004.

[94] T. J. Tambach, P. G. Bolhuis, E. J. Hensen, and B. Smit. Hysteresis in clay swelling induced by hydrogen bonding: Accurate prediction of swelling states. *Langmuir*, 22(3):1223–1234, 2006.

[95] Y. Zheng, A. Zaoui, and I. Shahrour. A theoretical study of swelling and shrinking of hydrated Wyoming montmorillonite. *App. Clay Sci.*, 51(1):177–181, 2011.

[96] A. Botan, V. Marry, B. Rotenberg, P. Turq, and B. Noetinger. How electrostatics influences hydrodynamic boundary conditions: Poiseuille and electroosmostic flows in clay nanopores. *J. Phys. Chem. C*, 117(2):978–985, 2013.

[97] B. F. Ngouana W and A. G. Kalinichev. Structural arrangements of isomorphic substitutions in smectites: Molecular simulation of the swelling properties, interlayer structure, and dynamics of hydrated Cs-montmorillonite revisited with new clay models. *J. Phys. Chem. C*, 118(24):12758–12773, 2014.

[98] M. H. Fu, Z. Z. Zhang, and P. F. Low. Changes in the properties of a montmorillonite-water system during the adsorption and desorption of water: Hysteresis. *Clay. Clay Miner.*, 38:485–492, 1990.

[99] S. H. Park and G. Sposito. Do montmorillonite surfaces promote methane hydrate formation? Monte Carlo and molecular dynamics simulations. *J. Phys. Chem. B*, 107(10):2281–2290, 2003.

[100] R. T. Cygan, S. Guggenheim, and A. F. Koster van Groos. Molecular models for the intercalation of methane hydrate complexes in montmorillonite clay. *J. Phys. Chem. B*, 108(39):15141–15149, 2004.

[101] J. O. Titiloye and N. T. Skipper. Monte Carlo and molecular dynamics simulations of methane in potassium montmorillonite clay hydrates at elevated pressures and temperatures. *J. Colloid Interf. Sci.*, 282(2):422–427, 2005.

[102] Q. Zhou, X. C. Lu, X. D. Liu, L. H. Zhang, H. P. He, J. X. Zhu, and P. Yuan. Hydration of methane intercalated in Na-smectites with distinct layer charge: Insights from molecular simulations. *J. Colloid Interf. Sci.*, 355(1):237–242, 2011.

[103] Z. Jin and A. Firoozabadi. Effect of water on methane and carbon dioxide sorption in clay minerals by Monte Carlo simulations. *Fluid Phase Equilib.*, 382:10–20, 2014.

[104] S. B. Cha, H. Ouar, T. R. Wildeman, and E. D. Sloan. A third-surface effect on hydrate formation. *J. Phys. Chem.*, 92:6492–6494, 1988.

[105] Z. Bacsik, R. Atluri, A. E. Garcia-Bennett, and N. Hedin. Temperature-induced uptake of CO_2 and formation of carbamates in mesocaged silica modified with n-propylamines. *Langmuir*, 26(12):10013–10024, 2010.

[106] A. Botan, B. Rotenberg, V. Marry, P. Turq, and B. Noetinger. Carbon dioxide in montmorillonite clay hydrates: Thermodynamics, structure, and transport from molecular simulation. *J. Phys. Chem. C*, 114(35):14962–14969, 2010.

[107] H. T. Schaef, E. S. Ilton, O. Qafoku, P. F. Martin, A. R. Felmy, and K. M. Rosso. In situ XRD study of Ca^{2+} saturated montmorillonite (STX-1) exposed to anhydrous and wet supercritical carbon dioxide. *Int. J. Greenh. Gas Con.*, 6:220–229, 2012.

[108] E. M. Myshakin, W. A. Saidi, V. N. Romanov, R. T. Cygan, and K. D. Jordan. Molecular dynamics simulations of carbon dioxide intercalation in hydrated Na-montmorillonite. *J. Phys. Chem. C*, 117(21):11028–11039, 2013.

[109] Z. Jin and A. Firoozabadi. Methane and carbon dioxide adsorption in clay-like slit pores by Monte Carlo simulations. *Fluid Phase Equilib.*, 360:456–465, 2013.

[110] E. M. Myshakin, M. Makaremi, V. N. Romanov, K. D. Jordan, and G. D. Guthrie. Molecular dynamics simulations of turbostratic dry and hydrated montmorillonite with intercalated carbon dioxide. *J. Phys. Chem. A*, 118(35):7454–7468, 2014.

[111] Y. Yu and X. Yang. Molecular simulation of swelling and interlayer structure for organoclay in supercritical CO_2. *Phys. Chem. Chem. Phys.*, 13:282–290, 2011.

[112] M. Krishnan, M. Saharay, and R. J. Kirkpatrick. Molecular dynamics modeling of CO_2 and poly (ethylene glycol) in montmorillonite: The structure of claypolymer composites and the incorporation of CO_2. *J. Phys. Chem. C*, 117(40):20592–20609, 2013.

[113] Q. Rao and Y. Leng. Methane aqueous fluids in montmorillonite clay interlayer under near-surface geological conditions: A grand canonical Monte Carlo and molecular dynamics simulation study. *J. Phys. Chem. B*, 118(37):10956–10965, 2014.

[114] K. A. Bullin and P. E. Krouskop. Compositional variety complicates processing plans for US shale gas. *Oil Gas J.*, 107(10):50–55, 2009.

[115] S. Clauzier, L. N. Ho, M. Pera-Titus, B. Coasne, and D. Farrusseng. Enhanced H_2 uptake in solvents confined in mesoporous metal-organic framework. *JACS*, 134(42):17369–17371, 2012.

[116] B. M. Krooss, F. Van Bergen, Y. Gensterblum, N. Siemons, H. J. M. Pagnier, and P. David. High-pressure methane and carbon dioxide adsorption on dry and moisture-equilibrated Pennsylvanian coals. *Int. J. Coal Geol.*, 51(2):69–92, 2002.

[117] L. Michels, J. O. Fossum, Z. Rozynek, H. Hemmen, K. Rustenberg, P. A. Sobas, G. N. Kalantzopoulos, K. D. Knudsen, M. Janek, T. S. Plivelic, and G. J.

da Silva. Intercalation and retention of carbon dioxide in a smectite clay promoted by interlayer cations. *Sci. Rep.*, 5, 2015.

[118] A. L. Myers and J. M. Prausnitz. Thermodynamics of mixed-gas adsorption. *AIChE J.*, 11:121–127, 1965.

[119] D. D. Do. *Adsorption Analysis: Equilibria and Kinetics Vol. 2.* Imperial College Press, London, UK, 1998.

[120] A. Goj, D. S. Sholl, E. D. Akten, and D. Kohen. Atomistic simulations of CO_2 and N_2 adsorption in silica zeolites: The impact of pore size and shape. *J. Phys. Chem. B*, 106(33):8367–8375, 2002.

[121] N. F. Cessford, N. A. Seaton, and T. Dren. Evaluation of ideal adsorbed solution theory as a tool for the design of metal-organic framework materials. *Ind. Eng. Chem. Res.*, 51:4911–4921, 2012.

[122] H. T. Schaef, C. L. Davidson, A. T. Owen, Q. R. Miller, J. S. Loring, C. J. Thompson, D. H. Bacon, V. A. Glezakou, and B. P. McGrail. CO_2 utilization and storage in shale gas reservoirs. *Energy Procedia*, 63:7844–7851, 2014.

[123] M. S. Lee, B. P. McGrail, and V. A. Glezakou. Microstructural response of variably hydrated Ca-rich montmorillonite to supercritical CO_2. *Environ. Sci. Technol.*, 48(15):8612–8619, 2014.

[124] P. Billemont, B. Coasne, and G. de Weireld. Adsorption of carbon dioxide, methane, and their mixtures in porous carbons: Effect of surface chemistry, water content, and pore disorder. *Langmuir*, 29(10):3328–3338, 2013.

[125] N. Yang, S. Liu, and X. Yang. Molecular simulation of preferential adsorption of CO_2 over CH_4 in Na-montmorillonite clay material. *Appl. Surf. Sci.*, 356:1262–1271, 2015.

[126] V. Romanov, Y. Soong, C. Carney, G. E. Rush, B. Nielsen, and W. O'Connor. Mineralization of carbon dioxide: A literature review. *ChemBioEng Rev.*, 2(4):231–256, 2015.

[127] A. Busch, P. Bertier, Y. Gensterblum, G. Rother, C. J. Spiers, M. Zhang, and H. M. Wentinck. On sorption and swelling of CO_2 in clays. *Geomech. Geophys. Geo-energ.*, 2:111–130, 2016.

[128] A. Kadoura, A. K. N. Nair, and S. Sun. Adsorption of carbon dioxide, methane and their mixture by montmorillonite in the presence of water. *Micropor. Mesopor. Mat.*, 225:331–341, 2016.

[129] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[130] S. Koneshan, J. C. Rasaiah, R. M. LyndenBell, and S. H. Lee. Solvent structure, dynamics, and ion mobility in aqueous solutions at 25 °C. *J. Phys. Chem. B*, 102(21):4193–4204, 1998.

[131] M. G. Martin and J. I. Siepmann. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B*, 102(14):2569–2577, 1998.

[132] J. Pérez-Pellitero, E. Bourasseau, I. Demachy, J. Ridard, P. Ungerer, and A. D. Mackie. Anisotropic united-atoms (AUA) potential for alcohols. *J. Phys. Chem. B*, 112(32):9853–9863, 2008.

[133] Kristóf T. and Liszi J. Effective intermolecular potential for fluid hydrogen sulfide. *J. Phys. Chem. B*, 101(28):5480–5483, 1997.

[134] J. G. Harris and K. H. Yung. Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model. *J. Phys. Chem.*, 99(31):12021–12024, 1995.

[135] B. Rotenberg, V. Marry, J. F. Dufrche, N. Malikova, E. Giffaut, and P. Turq. Modeling water and ion diffusion in clays: A multiscale approach. *C. R. Chim.*, 10(10):1108–1116, 2007.

[136] H. Maghsoudi and M. Soltanieh. Simultaneous separation of H_2S and CO_2 from CH_4 by a high silica CHA-type zeolite membrane. *J. Membr. Sci.*, 470:159–165, 2014.

[137] D. Nguyen-Thanh, K. Block, and T. J. Bandosz. Adsorption of hydrogen sulfide on montmorillonites modified with iron. *Chemosphere*, 59(3):343–353, 2005.

[138] G. Sethia, R. S. Somani, and H. C. Bajaj. Sorption of methane and nitrogen on cesium exchanged zeolite-X: Structure, cation position and adsorption relationship. *Ind. Eng. Chem. Res.*, 53(16):6807–6814, 2014.

[139] T. Vuong and P. A. Monson. Monte Carlo simulation studies of heats of adsorption in heterogeneous solids. *Langmuir*, 12(22):5425–5432, 1996.

[140] J. W. Lee, H. C. Kang, W. G. Shim, C. Kim, and H. Moon. Methane adsorption on multi-walled carbon nanotube at (303.15, 313.15, and 323.15) K. *J. Chem. Eng. Data*, 51(3):963–967, 2006.

[141] M. Makaremi, K. D. Jordan, G. D. Guthrie, and E. M. Myshakin. Multiphase Monte Carlo and molecular dynamics simulations of water and CO_2 intercalation in montmorillonite and beidellite. *J. Phys. Chem. C*, 119(27):15112–15124, 2015.

[142] D. Kim, D. W. Kim, H. K. Lim, J. Jeon, H. Kim, H. T. Jung, and H. Lee. Intercalation of gas molecules in graphene oxide interlayer: The role of water. *J. Phys. Chem. C*, 118(20):11142–11148, 2014.

[143] J. S. Loring, E. S. Ilton, J. Chen, C. J. Thompson, P. F. Martin, P. Bénézeth, K. M. Rosso, A. R. Felmy, and H. T. Schaef. In situ study of CO_2 and H_2O partitioning between Na-montmorillonite and variably wet supercritical carbon dioxide. *Langmuir*, 30(21):6120–6128, 2014.

[144] N. S. Suraweera, C. E. Barnes, and D. J. Keffer. The adsorption properties of amorphous, metal-decorated microporous silsesquioxanes for mixtures of carbon dioxide, methane and hydrogen. *J. Phys. Chem. C*, 118(20):13008–13017, 2014.

[145] P. Boulet, P. V. Coveney, and S. Stackhouse. Simulation of hydrated Li^+-, Na^+- and K^+-montmorillonite/polymer nanocomposites using large-scale molecular dynamics. *Chem. Phy. Let.*, 389(4):261–267, 2004.

[146] J. H. Bae, D. I. Song, and Y. W. Jeon. Adsorption of anionic dye and surfactant from water onto organomontmorillonite. *Separ. Sci. Technol.*, 35(3):353–365, 2000.

[147] A. Kadoura, A. K. N. Nair, and S. Sun. Molecular dynamics simulations of carbon dioxide, methane, and their mixture in montmorillonite clay hydrates. *J. Phys. Chem. C*, 120(23):12517–12529, 2016.

[148] M. Holmboe and I. C. Bourg. Molecular dynamics simulations of water and sodium diffusion in smectite interlayer nanopores as a function of pore size and temperature. *J. Phys. Chem. C*, 118(2):1001–1013, 2014.

[149] A. O. Yazaydin, A. I. Benin, S. A. Faheem, P. Jakubczak, J. J. Low, R. R. Willis, and R. Q. Snurr. Enhanced CO_2 adsorption in metal-organic frameworks via occupation of open-metal sites by coordinated water molecules. *Chem. Mater.*, 21(8):1425–1430, 2009.

[150] L. Liu and S. K. Bhatia. Molecular simulation of CO_2 adsorption in the presence of water in single-walled carbon nanotubes. *J. Phys. Chem. C*, 117(26):13479–13491, 2013.

[151] A. A. Sizova, V. V. Sizov, and E. N. Brodskaya. Adsorption of CO_2/CH_4 and CO_2/N_2 mixtures in SBA-15 and CMK-5 in the presence of water: A computer simulation study. *Colloid. Surface. A*, 474:76–84, 2015.

[152] D. A. Young and D. E. Smith. Simulations of clay mineral swelling and hydration: Dependence upon interlayer ion size and charge. *J. Phys. Chem. B*, 104(39):9163–9170, 2000.

[153] L. Zhang, X. Lu, X. Liu, J. Zhou, and H. Zhou. Hydration and mobility of interlayer ions of (Na_x, Ca_y)-montmorillonite: A molecular dynamics study. *J. Phys. Chem. C*, 118(51):29811–29821, 2014.

[154] Q. Rao and Y. Leng. Molecular understanding of CO_2 and H_2O in montmorillonite clay interlayer under CO_2 geological sequestration conditions. *J. Phys. Chem. C*, 120(5):2642–2654, 2016.

[155] V. Marry and P. Turq. Microscopic simulations of interlayer structure and dynamics in bihydrated heteroionic montmorillonites. *J. Phys. Chem. B*, 107(8):1832–1839, 2003.

[156] R. Martos-Villa, M. P. Mata, and C. I. Sainz-Daz. Characterization of CO_2 and mixed methane/CO_2 hydrates intercalated in smectites by means of atomistic calculations. *J. Mol. Graph. Model.*, 49:80–90, 2014.

[157] G. F. Versteeg and W. van Swaalj. Solubility and diffusivity of acid gases (CO_2, N_2O) in aqueous alkanolamine solutions. *J. Chem. Eng. Data*, 33(1):29–34, 1988.

[158] N. Malikova, A. Cadene, V. Marry, E. Dubois, and P. Turq. Diffusion of water in clays on the microscopic scale: Modeling and experiment. *J. Phys. Chem. B*, 110(7):3206–3214, 2006.

[159] S. Romero-Vargas Castrillón, N. Giovambattista, I. A. Aksay, and P. G. Debenedetti. Evolution from surface-influenced to bulk-like dynamics in nanoscopically confined water. *J. Phys. Chem. B*, 113(23):7973–7976, 2009.

[160] Leng Y. Hydration force between mica surfaces in aqueous KCl electrolyte solution. *Langmuir*, 28(12):5339–5349, 2012.

[161] J. Kärger and H. Pfeifer. N.m.r self-diffusion studies in zeolite science and technology. *Zeolites*, 7(2):90–107, 1987.

[162] R. Babarao and J. Jiang. Diffusion and separation of CO_2 and CH_4 in silicalite, C_{168} schwarzite, and IRMOF-1: A comparative study from molecular dynamics simulation. *Langmuir*, 24(10):5474–5484, 2008.

[163] E. Atci, I. Erucar, and S. Keskin. Adsorption and transport of CH_4, CO_2, H_2 mixtures in a bio-MOF material from molecular simulations. *J. Phys. Chem. C*, 115(14):6833–6840, 2011.

[164] T. Le, A. Striolo, and D. R. Cole. CO_2-C_4H_{10} mixtures simulated in silica slit pores: Relation between structure and dynamics. *J. Phys. Chem. C*, 119(27):15274–15284, 2015.

[165] J. Liu, S. Keskin, D. S. Sholl, and J. K. Johnson. Molecular simulations and theoretical predictions for adsorption and diffusion of CH_4/H_2 and CO_2/CH_4 mixtures in ZIFs. *J. Phys. Chem. C*, 115(25):12560–12566, 2011.

[166] S. E. Jee and D. S. Sholl. Carbon dioxide and methane transport in DDR zeolite: Insights from molecular simulations into carbon dioxide separations in small pore zeolites. *JACS*, 131(22):7896–7904., 2009.

[167] I. Deroche, G. Maurin, B. J. Borah, S. Yashonath, and H. Jobic. Diffusion of pure CH_4 and its binary mixture with CO_2 in faujasite NaY: A combination of neutron scattering experiments and molecular dynamics simulations. *J. Phys. Chem. C*, 114(11):5027–5034, 2010.

[168] F. Salles, H. Jobic, T. Devic, V. Guillerm, C. Serre, M. M. Koza, G. Ferey, and G. Maurin. Diffusion of binary CO_2/CH_4 mixtures in the MIL-47 (V) and MIL-53 (Cr) metal-organic framework type solids: A combination of neutron scattering measurements and molecular dynamics simulations. *J. Phys. Chem. C*, 117(21):11275–11284, 2013.

[169] B. M. Mognetti, L. Yelash, P. Virnau, W. Paul, K. Binder, M. Müller, and L. G. MacDowell. Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide. *J. Chem. Phys.*, 128:104501, 2008.

[170] B. M. Mognetti, P. Virnau, L. Yelash, W. Paul, K. Binder, M. Müller, and L. G. MacDowell. Coarse-grained models for fluids and their mixtures: Comparison of Monte Carlo studies of their phase behavior with perturbation theory and experiment. *J. Chem. Phys.*, 130:044101, 2009.

[171] Y. Wu, C. Kowitz, S. Sun, and A. Salama. Speeding up the flash calculations in two-phase compositional flow simulations–The application of sparse grids. *J. Comput. Phys.*, 285:88–99, 2015.

[172] Y. Wu. *Parallel Reservoir Simulations with Sparse Grid Techniques and Applications to Wormhole Propagation*. King Abdullah University of Science and Technology, KSA, 2015.

[173] B. Dazas, B. Lanson, A. Delville, J. L. Robert, S. Komarneni, L. J. Michot, and E. Ferrage. Influence of tetrahedral layer charge on the organization of interlayer water and ions in synthetic Na-saturated smectites. *J. Phys. Chem. C*, 119(8):4158–4172, 2015.

[174] M. Yiannourakou, P. Ungerer, B. Leblanc, X. Rozanska, P. Saxe, S. Vidal-Gilbert, F. Gouth, and F. Montel. Molecular simulation of adsorption in microporous materials. *Oil Gas Sci. Technol.*, 68(3):977–994, 2013.

[175] P. Ungerer, J. Collell, and M. Yiannourakou. Molecular modelling of the volumetric and thermodynamic properties of kerogen: Influence of organic type and maturity. *Energ. Fuel.*, 29(1):91–105, 2015.

[176] C. Bousige, C. M. Ghimbeu, C. Vix-Guterl, A. E. Pomerantz, A. Suleimenova, G. Vaughan, G. Garbarino, M. Feygenson, C. Wildgruber, F.-J. Ulm, R. J.-M. Pellenq, and B. Coasne. Realistic molecular model of kerogen's nanostructure. *Nat. Mater.*, 15:576–582, 2016.

[177] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts, UK, 2010.

APPENDIX A

Configurational Quantities Derivations

Derivations in reference [177] were used as a starting point to obtain all the quantities derived below:

System's Total Energy (E)

Starting from the definition of the partitioning function (Q) of N identical gas particles

$$Q = \frac{Z}{N!\lambda^{3N}}, \tag{A.1}$$

where $Z(N, V, T)$ is configurational partitioning function and λ is the plank's constant, the system's total energy can be expressed as:

$$E = -\left(\frac{\partial \ln Q}{\partial \beta}\right)_{N,V} = \frac{3N}{2\beta} - \frac{1}{Z}\left(\frac{\partial Z}{\partial \beta}\right)_{N,V}. \tag{A.2}$$

Noting that,

$$\frac{\partial \lambda}{\partial \beta} = \frac{\lambda}{2\beta}, \tag{A.3}$$

and

$$Z(N, V, T) = \int_{D(V)} \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 ... \mathrm{d}\mathbf{r}_N \exp\left[-\beta u(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)\right]. \tag{A.4}$$

In the equation above, $D(V)$ stands for the spatial domain defined by the physical container, while $\mathbf{r}_i$ and u represent the particle i coordinates vector and the system's

potential energy, respectively. Proceeding with derivation of E:

$$\left(\frac{\partial Z}{\partial \beta}\right)_{N,V} = -\int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 ... \mathrm{d}\mathbf{r}_N u \exp\left(-\beta u\right). \tag{A.5}$$

By definition, the canonical average of a quantity X is:

$$\langle X \rangle_{\mathrm{can}} = \frac{1}{Z} \int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 ... \mathrm{d}\mathbf{r}_N X \exp\left(-\beta u\right). \tag{A.6}$$

Therefore, the total energy of the system can be expressed as:

$$E = \frac{3N}{2\beta} + \langle u \rangle_{\mathrm{can}}, \tag{A.7}$$

where

$$\langle u \rangle_{\mathrm{can}} = -\frac{1}{Z}\left(\frac{\partial Z}{\partial \beta}\right)_{N,V}. \tag{A.8}$$

Heat Capacity (C_{v})

By definition:

$$C_{\mathrm{v}} = \left(\frac{\partial E}{\partial T}\right)_{N,V} = -k_{\mathrm{B}}\beta^2 \left(\frac{\partial E}{\partial \beta}\right)_{N,V}. \tag{A.9}$$

By substituting E expression in the equation above we get:

$$C_{\mathrm{v}} = \frac{3Nk_{\mathrm{B}}}{2} - k_{\mathrm{B}}\beta^2 \left[\frac{1}{Z^2}\left(\frac{\partial Z}{\partial \beta}\right)^2_{N,V} - \frac{1}{Z}\left(\frac{\partial^2 Z}{\partial \beta^2}\right)_{N,V}\right]. \tag{A.10}$$

where

$$\left(\frac{\partial^2 Z}{\partial \beta^2}\right)_{N,V} = \int \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 ... \mathrm{d}\mathbf{r}_N u^2 \exp\left(-\beta u\right). \tag{A.11}$$

Therefore the heat capacity becomes,

$$C_{\mathrm{v}} = \frac{3Nk_{\mathrm{B}}}{2} + k_{\mathrm{B}}\beta^2 \left(\left\langle u^2 \right\rangle_{\mathrm{can}} - \langle u \rangle^2_{\mathrm{can}}\right), \tag{A.12}$$

such that

$$\left\langle u^2 \right\rangle_{\text{can}} = \frac{1}{Z}\left(\frac{\partial^2 Z}{\partial \beta^2}\right)_{N,V} \text{ and } \langle u \rangle^2_{\text{can}} = \frac{1}{Z^2}\left(\frac{\partial Z}{\partial \beta}\right)^2_{N,V}. \tag{A.13}$$

Pressure (p)

Based on classical thermodynamics p can be related to the Helmholtz free energy (A):

$$p = -\left(\frac{\partial A}{\partial V}\right)_{N,T} = \frac{1}{Z\beta}\left(\frac{\partial Z}{\partial V}\right)_{N,T}, \tag{A.14}$$

where

$$A = -\frac{1}{\beta}\ln Q. \tag{A.15}$$

In order to be able to differentiate the p equation above, the dependence of the integration limits in (A.4) on V needs to be moved into the integrand. In fact, such transformation is possible using the scalability property of the particles coordinates with respect to the simulation box length (L). Assuming that the simulation box is cubic, the following new set of coordinates can be introduced:

$$\mathbf{s}_i = \frac{1}{L}\mathbf{r}_i = \frac{1}{V^{1/3}}\mathbf{r}_i, \quad i = 1, 2, ..., N. \tag{A.16}$$

As a result, Z can be rewritten as,

$$Z(N,V,T) = V^N \int \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 ... \mathrm{d}\mathbf{s}_N \exp\left[-\beta u\left(V^{1/3}\mathbf{s}_1, V^{1/3}\mathbf{s}_2, ..., V^{1/3}\mathbf{s}_N\right)\right], \tag{A.17}$$

while its derivative with respect to volume is:

$$\begin{aligned}\left(\frac{\partial Z}{\partial V}\right)_{N,T} &= \frac{NZ}{V} \\ &- \beta V^N \int \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 ... \mathrm{d}\mathbf{s}_N \left(\frac{\partial u}{\partial V}\right)_{N,T} \exp\left[-\beta u\left(V^{1/3}\mathbf{s}_1, V^{1/3}\mathbf{s}_2, ..., V^{1/3}\mathbf{s}_N\right)\right].\end{aligned} \tag{A.18}$$

Now let us evaluate the $(\partial u/\partial V)_{N,T}$ term, starting from the total differential of u:

$$\mathrm{d}u = \left(\frac{\partial u}{\partial \left(V^{1/3}\mathbf{s}_1\right)}\right).\mathrm{d}(V^{1/3}\mathbf{s}_1) + ... + \left(\frac{\partial u}{\partial \left(V^{1/3}\mathbf{s}_N\right)}\right).\mathrm{d}(V^{1/3}\mathbf{s}_N), \tag{A.19}$$

differentiating with respect to V:

$$\left(\frac{\partial u}{\partial V}\right)_{N,T} = \left(\frac{\partial u}{\partial \left(V^{1/3}\mathbf{s}_1\right)}\right).\left(\frac{\partial \left(V^{1/3}\mathbf{s}_1\right)}{\partial V}\right)_{N,T} + ...+ \left(\frac{\partial u}{\partial \left(V^{1/3}\mathbf{s}_N\right)}\right).\left(\frac{\partial \left(V^{1/3}\mathbf{s}_N\right)}{\partial V}\right)_{N,T}, \tag{A.20}$$

with further rearrangements:

$$\begin{aligned}\left(\frac{\partial u}{\partial V}\right)_{N,T} &= \frac{\mathrm{d}V^{1/3}}{\mathrm{d}V}\frac{\partial u}{\partial \mathbf{r}_1}.\ \mathbf{s}_1 + ... + \frac{\mathrm{d}V^{1/3}}{\mathrm{d}V}\frac{\partial u}{\partial \mathbf{r}_N}.\ \mathbf{s}_N \\ &= \frac{1}{3V^{2/3}}\sum_{i=1}^{N}\mathbf{s}_i\ .\ \frac{\partial u}{\partial \mathbf{r}_i} = \frac{1}{3V}\sum_{i=1}^{N}\mathbf{r}_i\ .\ \frac{\partial u}{\partial \mathbf{r}_i}.\end{aligned} \tag{A.21}$$

Plugging it back in equation (A.18), while writing it in terms of $\mathbf{r}$

$$\left(\frac{\partial Z}{\partial V}\right)_{N,T} = \frac{NZ}{V} - \frac{\beta}{3V}\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2...\mathrm{d}\mathbf{r}_N \left(\sum_{i=1}^{N}\mathbf{r}_i\ .\ \frac{\partial u}{\partial \mathbf{r}_i}\right)\exp\left[-\beta u\left(\mathbf{r}_1,\mathbf{r}_2,...,\mathbf{r}_N\right)\right]. \tag{A.22}$$

So the pressure equation becomes:

$$p = \frac{\rho}{\beta} + \frac{1}{3V}\left\langle\sum_{i=1}^{N}\mathbf{r}_i\ .\ \mathbf{F}_i\right\rangle, \tag{A.23}$$

such that $\mathbf{F}_i = -\partial u/\partial \mathbf{r}_i$ is the sum of forces on particle i, and $\langle X\rangle$ stands for the ensemble average of quantity X. For a given pair-wise additive potential (u_{ij}), it is convenient to define the vector $\mathbf{f}_{ij}$, which is the force on particle i due to particle j,

such that:

$$\mathbf{F}_i = \sum_{i=1, j\neq i}^{N} \mathbf{f}_{ij}, \text{ and } \mathbf{f}_{ij} = -\mathbf{f}_{ji}. \tag{A.24}$$

Consequently, the virial term can be rewritten as:

$$\sum_{i=1}^{N} \mathbf{r}_i \,.\, \mathbf{F}_i = \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_i \,.\, \mathbf{f}_{ij}. \tag{A.25}$$

By interchanging the summation index it is possible to reformulate the virial term as:

$$\sum_{i=1}^{N} \mathbf{r}_i \,.\, \mathbf{F}_i = \frac{1}{2} \left(\sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_i \,.\, \mathbf{f}_{ij} + \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_j \,.\, \mathbf{f}_{ji} \right), \tag{A.26}$$

but $\mathbf{f}_{ij}$ = - $\mathbf{f}_{ji}$, then:

$$\begin{aligned} \sum_{i=1}^{N} \mathbf{r}_i \,.\, \mathbf{F}_i &= \frac{1}{2} \left(\sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_i \,.\, \mathbf{f}_{ij} - \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_j \,.\, \mathbf{f}_{ij} \right) \\ = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} (\mathbf{r}_i - \mathbf{r}_j) \,.\, \mathbf{f}_{ij} &\equiv \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, i\neq j}^{N} \mathbf{r}_{ij} \,.\, \mathbf{f}_{ij} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbf{r}_{ij} \,.\, \mathbf{f}_{ij}. \end{aligned} \tag{A.27}$$

The final equation to compute the ensemble average pressure (p) can be reached by substituting the above relation in equation (A.23)

$$p = \frac{\rho}{\beta} + \frac{1}{3V} \left\langle \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbf{r}_{ij} \,.\, \mathbf{f}_{ij} \right\rangle_{\text{can}}. \tag{A.28}$$

In fact, it is possible, by introducing a configurational quantity v, to show that:

$$p = \frac{\rho}{\beta} + \frac{1}{V} \langle v \rangle_{\text{can}}, \tag{A.29}$$

such that:

$$\begin{aligned} v = -V\left(\frac{\partial u}{\partial V}\right)_{N,T} &= -\frac{1}{3}\sum_{i=1}^{N}\mathbf{r}_i \,.\, \left(\frac{\partial u}{\partial \mathbf{r}_i}\right)_{N,T} = \frac{1}{3}\sum_{i=1}^{N}\mathbf{r}_i \,.\, \mathbf{F}_i \\ &= \frac{1}{3}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbf{r}_{ij} \,.\, \mathbf{f}_{ij} = -\frac{1}{3}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbf{r}_{ij} \,.\, \left(\frac{\partial u_{ij}}{\partial \mathbf{r}_{ij}}\right)_{N,T}. \end{aligned} \tag{A.30}$$

Inverse of Isothermal Compressibility (c_{T}^{-1})

By definition:

$$c_{\mathrm{T}}^{-1} = -V\left(\frac{\partial p}{\partial V}\right)_{N,T}. \tag{A.31}$$

By substituting p expression in the equation above we get:

$$c_{\mathrm{T}}^{-1} = -\frac{V}{\beta}\left[-\frac{1}{Z^2}\left(\frac{\partial Z}{\partial V}\right)_{N,T}^{2} + \frac{1}{Z}\left(\frac{\partial^2 Z}{\partial V^2}\right)_{N,T}\right]. \tag{A.32}$$

Starting with the first term using Equ. (A.18)

$$\begin{aligned} \left(\frac{\partial Z}{\partial V}\right)_{N,T}^{2} = \rho^2 Z^2 + \beta^2 &\left[V^N \int \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 ... \mathrm{d}\mathbf{s}_N \left(\frac{\partial u}{\partial V}\right)_{N,T} \exp\left(-\beta u\right)\right]^2 \\ &- 2\rho\beta Z V^N \int \mathrm{d}\mathbf{s}_1 \mathrm{d}\mathbf{s}_2 ... \mathrm{d}\mathbf{s}_N \left(\frac{\partial u}{\partial V}\right)_{N,T} \exp\left(-\beta u\right), \end{aligned} \tag{A.33}$$

the complete first term becomes:

$$-\frac{1}{Z^2}\left(\frac{\partial Z}{\partial V}\right)_{N,T}^{2} = -\rho^2 - \frac{\beta^2}{V^2}\left\langle v\right\rangle_{\mathrm{can}}^{2} - \frac{2\rho\beta}{V}\left\langle v\right\rangle_{\mathrm{can}}. \tag{A.34}$$

On the other hand, the second term can be evaluated as follows:

$$\left(\frac{\partial^2 Z}{\partial V^2}\right)_{N,T} = N\left[-\frac{Z}{V^2}+\frac{1}{V}\left(\frac{\partial Z}{\partial V}\right)_{N,T}\right]-\beta\rho V^N\int \mathrm{d}\mathbf{s}_1\mathrm{d}\mathbf{s}_2...\mathrm{d}\mathbf{s}_N\left(\frac{\partial u}{\partial V}\right)_{N,T}\exp\left(-\beta u\right)$$
$$-\beta V^N\int \mathrm{d}\mathbf{s}_1\mathrm{d}\mathbf{s}_2...\mathrm{d}\mathbf{s}_N\left[\left(\frac{\partial^2 u}{\partial V^2}\right)_{N,T}-\beta\left(\frac{\partial u}{\partial V}\right)^2_{N,T}\right]\exp\left(-\beta u\right), \quad \text{(A.35)}$$

leading to:

$$\left(\frac{\partial^2 Z}{\partial V^2}\right)_{N,T} = -\frac{\rho Z}{V}+\rho^2 Z-2\beta\rho V^N\int \mathrm{d}\mathbf{s}_1\mathrm{d}\mathbf{s}_2...\mathrm{d}\mathbf{s}_N\left(\frac{\partial u}{\partial V}\right)_{N,T}\exp\left(-\beta u\right)$$
$$-\beta V^N\int \mathrm{d}\mathbf{s}_1\mathrm{d}\mathbf{s}_2...\mathrm{d}\mathbf{s}_N\left(\frac{\partial^2 u}{\partial V^2}\right)_{N,T}\exp\left(-\beta u\right)$$
$$+\beta^2 V^N\int \mathrm{d}\mathbf{s}_1\mathrm{d}\mathbf{s}_2...\mathrm{d}\mathbf{s}_N\left(\frac{\partial u}{\partial V}\right)^2_{N,T}\exp\left(-\beta u\right), \quad \text{(A.36)}$$

and then the final form of the second term becomes:

$$\frac{1}{Z}\left(\frac{\partial^2 Z}{\partial V^2}\right)_{N,T} = -\frac{\rho}{V}+\rho^2+\frac{2\beta\rho}{V}\left\langle v\right\rangle_{\mathrm{can}}-\frac{\beta}{V^2}\left\langle w\right\rangle_{\mathrm{can}}+\frac{\beta^2}{V^2}\left\langle v^2\right\rangle_{\mathrm{can}}, \quad \text{(A.37)}$$

such that:

$$w = V^2\left(\frac{\partial^2 u}{\partial V^2}\right)_{N,T} = -V^2\frac{\partial}{\partial V}\left(\frac{1}{V}v\right). \quad \text{(A.38)}$$

The final relation to compute inverse of isothermal compressibility is obtained by summing the two terms above:

$$c_{\mathrm{T}}^{-1} = \frac{\rho}{\beta}+\frac{1}{V}\left\langle w\right\rangle_{\mathrm{can}}-\frac{\beta}{V}\left(\left\langle v^2\right\rangle_{\mathrm{can}}-\left\langle v\right\rangle^2_{\mathrm{can}}\right). \quad \text{(A.39)}$$

Lennard-Jones Potential Model

For the Lennard-Jones model the quantities discussed above take the following forms:

$$u = 4\varepsilon \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{A.40}$$

$$v = -V \left(\frac{\partial u}{\partial V} \right)_{N,T} = 8\varepsilon \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[2 \left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{A.41}$$

$$w = -V^2 \frac{\partial}{\partial V} \left(\frac{1}{V} v \right)_{N,T} = 8\varepsilon \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[10 \left(\frac{\sigma}{r_{ij}} \right)^{12} - 3 \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{A.42}$$

$$E_{tail} = \frac{1}{2} N \int_{r_c}^{\infty} 4\pi\rho r^2 u dr = \frac{8}{3} \varepsilon \sigma^3 N \pi \rho \left[\frac{1}{3} \left(\frac{\sigma}{r_c} \right)^{9} - \left(\frac{\sigma}{r_c} \right)^{3} \right]. \tag{A.43}$$

$$p_{tail} = \frac{1}{2} N \int_{r_c}^{\infty} 4\pi\rho r^2 \left(\frac{1}{V} v \right) dr = \frac{16}{3} \varepsilon \sigma^3 \pi \rho^2 \left[\frac{2}{3} \left(\frac{\sigma}{r_c} \right)^{9} - \left(\frac{\sigma}{r_c} \right)^{3} \right]. \tag{A.44}$$

$$c_{tail} = \frac{1}{2} N \int_{r_c}^{\infty} 4\pi\rho r^2 \left(\frac{1}{V} w \right) dr = \frac{16}{3} \varepsilon \sigma^3 \pi \rho^2 \left[\frac{10}{3} \left(\frac{\sigma}{r_c} \right)^{9} - 3 \left(\frac{\sigma}{r_c} \right)^{3} \right]. \tag{A.45}$$

APPENDIX B

Table B.1: Examples of commonly used statistical ensembles.

Statistical ensemble	Imposed variables	Application
Canonical ensemble	N,V,T	Phase properties
Isothermal-Isobaric ensemble	N,p,T	Phase properties
Grand Canonical ensemble	μ,V,T	Adsorption
Gibbs ensemble	N_{tot},V_{tot},T	Phase equilibrium

Table B.2: Comparing the canonical average normalized pressure (p^*) out of the bond formation and non-conservative early rejection methods against the pressures obtained by the conventional algorithm. Results from the conservative scheme are not reported as it produces exactly the same values as the conventional one. Statistical uncertainties were computed using bootstrap method (Efron, 1979; Newman and Barkema, 1999).

		p^*		
β^*	ρ^*	Conventional	Bond formation	Non-conservative
0.10	0.10	1.1102 ± 0.0002	1.1099 ± 0.0002	1.1102 ± 0.0002
	0.30	4.3109 ± 0.0009	4.3109 ± 0.0009	4.3109 ± 0.0010
	0.50	9.9329 ± 0.0021	9.9305 ± 0.0023	9.9301 ± 0.0021
	0.70	20.269 ± 0.0035	20.279 ± 0.0038	20.254 ± 0.0040
	0.90	39.419 ± 0.0064	39.401 ± 0.0058	39.331 ± 0.0058
0.40	0.10	0.2381 ± 0.0001	0.2382 ± 0.0001	0.2381 ± 0.0001
	0.30	0.7451 ± 0.0003	0.7451 ± 0.0003	0.7447 ± 0.0003
	0.50	1.7079 ± 0.0008	1.7124 ± 0.0007	1.7109 ± 0.0008
	0.70	4.3793 ± 0.0016	4.3528 ± 0.0012	4.3635 ± 0.0012
	0.90	11.529 ± 0.0022	11.531 ± 0.0017	11.588 ± 0.0020
0.70	0.10	0.1072 ± 0.0001	0.1072 ± 0.0001	0.1073 ± 0.0001
	0.30	0.1876 ± 0.0002	0.1868 ± 0.0003	0.1880 ± 0.0003
	0.50	0.3177 ± 0.0006	0.3169 ± 0.0006	0.3145 ± 0.0005
	0.70	1.3563 ± 0.0009	1.3707 ± 0.0009	1.3741 ± 0.0008
	0.90	5.9638 ± 0.0013	5.9960 ± 0.0027	5.9149 ± 0.0013

Table B.3: Comparing the computed normalized liquid and vapor number densities (ρ_l^* and ρ_v^*) out of the hybrid and non-conservative early rejection schemes against the conventional algorithm. Statistical uncertainties were computed using bootstrap method (Efron, 1979; Newman and Barkema, 1999).

	Conventional		Hybrid		Non-conservative	
T^*	ρ_l^*	ρ_v^*	ρ_l^*	ρ_v^*	ρ_l^*	ρ_v^*
0.80	$0.7975 \pm 3e^{-5}$	$0.0062 \pm 1e^{-5}$	$0.8008 \pm 3e^{-5}$	$0.0067 \pm 1e^{-5}$	$0.7968 \pm 2e^{-5}$	$0.0057 \pm 1e^{-5}$
0.85	$0.7771 \pm 3e^{-5}$	$0.0095 \pm 1e^{-5}$	$0.7781 \pm 3e^{-5}$	$0.0093 \pm 1e^{-5}$	$0.7764 \pm 3e^{-5}$	$0.0093 \pm 1e^{-5}$
0.90	$0.7547 \pm 3e^{-5}$	$0.0150 \pm 2e^{-5}$	$0.7547 \pm 3e^{-5}$	$0.0144 \pm 1e^{-5}$	$0.7538 \pm 3e^{-5}$	$0.0146 \pm 2e^{-5}$
0.95	$0.7274 \pm 3e^{-5}$	$0.0211 \pm 2e^{-5}$	$0.7270 \pm 3e^{-5}$	$0.0212 \pm 2e^{-5}$	$0.7250 \pm 4e^{-5}$	$0.0201 \pm 2e^{-5}$
1.00	$0.7035 \pm 4e^{-5}$	$0.0308 \pm 2e^{-5}$	$0.7030 \pm 3e^{-5}$	$0.0303 \pm 2e^{-5}$	$0.7019 \pm 4e^{-5}$	$0.0296 \pm 2e^{-5}$
1.05	$0.6709 \pm 4e^{-5}$	$0.0397 \pm 2e^{-5}$	$0.6727 \pm 4e^{-5}$	$0.0406 \pm 2e^{-5}$	$0.6728 \pm 4e^{-5}$	$0.0406 \pm 2e^{-5}$
1.10	$0.6435 \pm 5e^{-5}$	$0.0558 \pm 3e^{-5}$	$0.6408 \pm 5e^{-5}$	$0.0553 \pm 3e^{-5}$	$0.6425 \pm 5e^{-5}$	$0.0558 \pm 3e^{-5}$
1.15	$0.6061 \pm 5e^{-5}$	$0.0738 \pm 4e^{-5}$	$0.6063 \pm 5e^{-5}$	$0.0740 \pm 4e^{-5}$	$0.6061 \pm 5e^{-5}$	$0.0737 \pm 4e^{-5}$
1.20	$0.5626 \pm 6e^{-5}$	$0.0984 \pm 5e^{-5}$	$0.5622 \pm 6e^{-5}$	$0.0973 \pm 5e^{-5}$	$0.5652 \pm 6e^{-5}$	$0.1006 \pm 5e^{-5}$

Table B.4: Theoretical (Sandler, 2010) and recommended single-site LJ model parameters.

Molecules	$\varepsilon_0/k_\mathrm{B}$ (K)	$\varepsilon_\mathrm{rec}/k_\mathrm{B}$ (K)	σ_0 (Å)	σ_rec (Å)
Ar	111.6200	116.4030	3.5128	3.3911
Kr	155.1704	162.2714	3.7696	3.6264
Xe	214.6170	224.6165	4.1089	3.9484
CH_4	141.1585	147.7146	3.8555	3.7309
CO	98.4148	91.6006	3.7696	3.5876
N_2	93.4756	98.3203	3.7317	3.6272

Table B.5: Charges q and LJ parameters σ and ε of water, clay, sorbates and ions.

Force or charge site	σ (Å)	ε/k_B (K)	q (e)
Water (SPC/E model) [129]			
O	3.1656	78.20	−0.8476
H	0.0	0.0	0.4238
Montmorillonite (CLAYFF) [93]			
Hydroxyl H	0.0	0.0	0.4250
Hydroxyl O	3.1656	78.20	−0.9500
Hydroxyl O with substitution	3.1656	78.20	−1.0808
Bridging O	3.1656	78.20	−1.0500
Bridging O with octahedral substitution	3.1656	78.20	−1.1808
Bridging O with tetrahedral substitution	3.1656	78.20	−1.1688
Tetrahedral Si	3.3020	9.2618×10^{-4}	2.1000
Octahedral Al	4.2712	6.6918×10^{-4}	1.5750
Tetrahedral Al	3.3020	9.2618×10^{-4}	1.5750
Octahedral Mg	5.2643	4.5440×10^{-4}	1.3600
Methane (TraPPE model) [131]			
C	3.73	148	0.0
Carbon dioxide (flexible EPM2 model) [134]			
C	2.757	28.129	0.6512
O	3.033	80.507	−0.3256
Nitrogen (three-site model) [132]			
N	3.3	36	−0.5075
q_1	0.0	0.0	1.0150
Hydrogen sulfide (four-site model) [133]			
S	3.73	250	0.4000
H	0.0	0.0	0.2500
q_1	0.0	0.0	−0.9000
Ions [92, 130]			
Na	2.350	65.47	1.0000
Cs	3.831	50.32	1.0000
Ca	2.872	50.32	2.0000