UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PRODUCTION OF HYDROCARBONS FROM ORGANIC NANOPORES: A COMPREHENSIVE STUDY USING MOLECULAR SIMULATIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

FELIPE PEREZ
Norman, Oklahoma
2020

PRODUCTION OF HYDROCARBONS FROM ORGANIC NANOPORES: A COMPREHENSIVE STUDY USING MOLECULAR SIMULATIONS

A DISSERTATION APPROVED FOR THE
MEWBOURNE SCHOOL OF PETROLEUM AND GEOLOGICAL ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Deepak Devegowda, Chair

Dr. Carl Sondergeld

Dr. Chandra Rai

Dr. Sepideh Razavi

To my wife Andrea, my parents and brothers, who have always believed in me even in my weakest moments.

"To feed applied science by starving basic science is like economizing on the foundations of a building so that it may be built higher. It is only a matter of time before the whole edifice crumbles."

George Porter, Nobel Prize in Chemistry in 1967

ACKNOWLEDGEMENTS

I do not think there is a simple way to keep it short and concise. First and foremost, I praise and thank God for guiding my way to the point where I am today. *"While from behind, a voice shall sound in your ears: "This is the way; walk in it," when you would turn to the right or to the left." (Is 30:21).* There is no other place in space and time in which I wish I were.

I have never been alone in this journey, He has placed so many people along the road... I want to thank my parents for their unceasing support and example. They have always believed in me even in the moments when I was about to give up. They are a living testimony of faith and resiliency. Juan, my brother, thanks for opening the doors for us to leave the nest and fly away. Jose, I will be forever thankful for the support you provide and the peace you bring to our lives. Andrea, my beloved wife, my eternal support and companion in this adventure, I would not be here today without you. No words can express my debt to you. To my in-laws, this journey would have not started if you had not encouraged me to pursue my dreams. Thank you.

I also want to thank my adviser, Dr. Deepak Devegowda for the continuous challenges he threw at me during these years. Instead of requesting things from me, his approach was to inquire into something he wished to see from a molecular standpoint. It was on me to come up with something, and I personally consider that this is a very effective learning path.

I would like to thank one by one the members of my committee. To Dr. Carl

Sondergeld I owe countless laughs, meaningful discussions, valuable advice and moral support. He his a great person to talk about anything. His jokes always made the difference between a good day and a terrific day. I acknowledge Dr. Chandra Rai for believing in my work and seeking the connections between simulations and experiments, but above all for stepping up for me in moments that so demanded. He makes me feel my work is appreciated. To Dr. Sepideh Razavi, thank you for kind disposition to help.

I cannot carry on without mentioning Dr. Kieran Mullen, who revived in me the passion for teaching and reminded me of why I chose physics in the first place. No lecture was ever boring, he has the very unique ability to make you feel physics is alive and mathematics is all fun. I wish I will instill the same enthusiasm to my students when I teach again.

The Mewbourne School of Petroleum and Geological Engineering outstanding staff cannot go unnoticed. My special thanks to Danika Hines-Barnett, who takes great care of us, graduate students, and makes our lives infinitely much easier. If I had had to deal with all the additional things that popped up down the road all by myself, I would not have made it. My gratitude also to Sonya Grant, Haley Pearson, Phoebe Morales and Francey Freeman, who cheered me up with a smile every time we saw each other.

My work was tremendously enriched by the technical talks with Richard Newhart, Tom Smagala, Chris Bement, and John Mansoori, formerly at Encana Corporation/Ovintiv and who were always open to discuss and ask questions from an engineering perspective. They helped shape my work and see it as part of a bigger picture. My especial thanks also go to Sidi Mamoudou who conducted the in-house experiments in the Integrated Core Characterization Center (IC^3) and spent time with me connecting the dots between experiments and simulations. I truly enjoyed our brainstorms on my whiteboard.

All the simulations in this work were possible thanks to the OU Supercomputing Center for Education and Research (OSCER). It granted me hundreds of hours of simulation. Moreover, I want to acknowledge the great human team at OSCER; their opportune support with any technical issues made my work much more efficient. They are the perfect example of a good service.

Believe it or not, there is also a life out there during graduate studies. I have met amazing friends along the road. I have to say that my time at St. Thomas More University Parish have boosted my personal and spiritual growth. God used every moment here to show me his face disguised in so many different ways. I thank Fr. Jim Goins for his leadership, for being the shepherd of this flock. The work done by Nolan Reilly and the choir always makes the liturgy so rich, sacred and beautiful that it literally uplifts the soul, balancing my earthly and spiritual worlds. To Ed and Pat Chojnicki, Jay and Marie Johnson, Dan and Lori Haken, and Vickie Newton, thank you for sharing with us our faith in Jesus Christ. Thank you for all the nights we spent together talking about the continuous action of God in our lives. To Dr. Jean-Claude Roegiers and his wife Jeanne, all my gratitude for opening the doors of your house to us and making us part of your family. To Natasha Roegiers and Joseph Barrett, the joy of having Evelina in our lives is something indescribable. I will be forever thankful. To Pierce and Vickie Newton, thank you both for the love you so bountifully express to us. It is well reciprocated.

In part, this experience has been successful thanks to the Colombian crowd at the University of Oklahoma. To Yoana Walshchap, thanks for having that heart. To Michael Mendez, Estefania Mejia, Yuliana Zapata, Carolina Mejia, Dr. Andres Gonzalez, and Colombian emeritus Any Ordoñez and Saurabh Sinha, you guys are the Colombian flavor in Oklahoma.

Norman is much more than just a college town. It has presented me with many more things than I ever imagined. My especial gratitude to Mike and Beth Wilson for supporting my better half to pursue her dreams. Her realization is my own realization.

I am unpurposely leaving out so many names that should be mentioned here, but it is already long as it is. Know that you all have your share in this achievement. Thank you.

Boomer Sooner.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Molecular models of organic pores are created using molecular dynamics simulations to investigate fluid behavior of confined reservoir fluids and fluid transport through nanoporous shales, aiming to reveal the underlying physics laws that govern the phenomenology that occurs at molecular scale. This is done through a study of the spatial distribution of fluids within organic pores, followed by the exploration of the fluid transport mechanisms during primary depletion from a molecular point of view. Finally, huff-n-puff enhanced oil recovery simulations are designed to examine the mechanisms that govern hydrocarbon recovery in shale rocks. The general observations are experimentally confirmed, providing a unique understanding of shales and the means by which oil and natural gas are produced.

PREFACE

"We have arranged a global civilization in which most crucial elements profoundly depend on science and technology. We have also arranged things so that almost no one understands science and technology. This is a prescription for disaster. We might get away with it for a while, but sooner or later this combustible mixture of ignorance and power is going to blow up in our faces."
Carl Sagan, The Demon-Haunted World: Science as a Candle in the Dark

One thing that struck me a few years ago while attending a class with Dr. Carl Sondergeld at the University of Oklahoma is that most of the oil and natural gas produced today in the U.S. reside in pores that are one thousand times smaller than the thickness of an average human hair. Such pores are located in a type of rock called shale, which is the most common sedimentary rock made up of an inorganic matrix in which organic matter (the so-called kerogen) is dispersed. Kerogen is the naturally occurring solid organic matter in shales where petroleum was generated, and is non-extractable using organic solvents but can yield oil upon heating (Durand, 1980; Tissot and Welte, 1984). When I became a researcher member for the Unconventional Shale Gas Consortium, I knew I would have to deal with these things every single day.

The interest in shale reservoirs bloomed in the early 2000s due to technological breakthroughs, mainly in horizontal drilling and hydraulic fracturing, that made viable

the development of projects to recover oil and gas resources from the very source rocks where they were formed over geologic time and in which there are still huge amounts of these resources. However, these projects came along with new fundamental challenges; shale reservoirs are quite different from the kind of reservoirs from which petroleum had been produced for over 150 years with very unique characteristics that break with the conventional understanding of oil reservoirs, for instance the presence of oil above gas. Therefore they are referred to as unconventional reservoirs.

The learning curve has grown tremendously over the past two decades. The primordial approach has been to investigate the properties of shales using samples retrieved from unconventional reservoirs in a case-by-case basis, figuring the particular characteristics of every specific formation. Surprisingly, being the most common sedimentary rock, shale formations are very different when compared with each other. They share some general properties, for example a grain size below 1/256 mm, very low porosity, ultra-low permeability (in the range of nanodarcies[1], six orders of magnitude below permeability in conventional reservoirs), and fissility (the rock tends to split along flat planes of weakness). Nevertheless, the particular differences outnumber their similarities, for instance the spatial arrangement and maturity of the organic matter dispersed in a mineral matrix, the amount of organic matter in a rock, the distribution of pore sizes and shapes, the location of the pores, the pore throats that connect pores, just to mention a few. Moreover, shales are anything but homogeneous and isotropic rocks.

Understanding the rock is important, so is understanding the fluids trapped in the tiny pores in those rocks. With pore sizes in the range of nanometers, it is just expected that the fluids do not behave like their bulk counterparts. Why would they? Furthermore, the fluids are located now in pores with two different chemical environments, one hydrophobic (the organic matter), the other hydrophilic (the mineral matrix), and this fact has consequences as to what type of fluids are in either kind of pore and how they will flow through the pore network.

All in all, unconventional reservoirs represent a huge intellectual challenge. Oil

[1] 1 darcy $\sim 1 \times 10^{-13}$ m^2

companies spend considerable amounts of money trying to understand the rock that makes up their reservoir, but it is very frustrating that what seems to work in one does not work as well in another. This signals the need for a more fundamental comprehension of the phenomenology that occurs in the system shale/reservoir fluids. The goal is to reveal the physics laws that govern fluid behavior and fluid flow in shales. The physical systems are different based on the characteristics of the shale and the type of fluid in its pore space, but there must be some fundamental underlying phenomenological ground common to them all. And here we are, given the inaccessibility of the pore space to routine (and even novel) laboratory measurements, the use of molecular simulations for studying nanoporous organic rich shale and confined fluid behavior is very handy.

The motivation of this work is then to gain some insights from molecular simulations. The advent of more powerful computers on the one hand, and the development of more accurate sets of parameters to describe the atoms (the so-called force fields) that make up a system like kerogen/organic fluids on the other, have paved the way to take on this endeavor.

The materials contained in this work present what I hope is a comprehensive study from a molecular standpoint of fluid behavior and transport in organic nanopores in shales. My ultimate intention is to communicate the findings in plain English, trying to stay away from technicalities when possible, so that this work be accessible to and engage as many interested readers as possible.

Felipe Perez
Norman, OK
June, 2020

Chapter 1

INTRODUCTION

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

Max Planck, Nobel Prize in Physics in 1918

This works attempts to provide the reader with a comprehensive study of fluid behavior and transport in shale organic nanopores from a molecular standpoint. The interest in this topic comes from the increasing exploitation of oil and natural gas from shale reservoirs since the early 2000s, brought about by technological breakthroughs such as horizontal drilling and hydraulic fracturing.

Shales are the most common sedimentary rock, with a grain size below 1/256 mm and characterized by a very low porosity, ultra-low permeability (in the range of nanodarcies[1], six orders of magnitude below permeability in conventional reservoirs), and fissility (the rock tends to split along flat planes of weakness) (Tarbuck et al., 2006). Dispersed across the mineral matrix of shales there are accumulations of organic matter (the so-called kerogen) that were deposited and subsequently buried at high conditions of pressure and temperature yielding oil and natural gas over geologic time. A fraction of those hydrocarbons migrated from the source rocks (i.e. from shales) and, under the right set of conditions, accumulated in porous rocks to form oil reservoirs called *conven-*

[1] 1 darcy $\sim 1 \times 10^{-13}$ m^2

tional reservoirs that have been produced for over 150 years. However, there are still very large amounts of hydrocarbons stored in the source rocks. According to Aminzadeh and Dasgupta (2013), while the conventional proved reserves in 2013 were approximately 1.5 trillion barrels, the world's estimated oil reserves from unconventional resources such as heavy oils and tar sands alone were about 3 trillion barrels. Moreover, the U.S Energy Information Administration (EIA) reports that as of December 31, 2018, the U.S. proved oil reserves from shale plays is 22.9 billion barrels, and the proved reserves of U.S. natural gas from shale is 342.1 trillion cubic feet (EIA, 2019).

The huge amount of proved reserves and low permeabilities challenge oil companies to design better production strategies. The primary recovery factors in shale reservoirs are normally between 5 and 10% (Clark, 2009; King, 2014; Hoffman and Evans, 2016), which underscores the necessity to implement enhanced oil recovery in shale plays but it comes with its own set of challenges. With pore sizes in the range of nanometers in shales, the behavior of the stored fluids does not correspond to their unconfined counterparts (i.e. in bulk conditions) (Devegowda et al., 2012; Teklu et al., 2014). Additionally, the strong adsorption of fluid molecules breaks down the continuum description of fluids, which fails to predict transport through shale nanopores (Falk et al., 2015).

The limited understanding of fluid storage and transport in shale reservoirs has opened the door to new research enterprises from three fronts: computational modeling/theoretical, experimental, and field tests. This work belongs to the first group.

Among the vast list of unknowns, this work focuses on four main questions, summarized in Figure 1.1:

1. How are the fluids organized within organic pores?
2. How do fluids move in the absence of external gradients?
3. How do fluids work their way out of the pores?
4. How to enhance oil recovery?

The pores in shales can be located in kerogen, the inorganic matrix or both, depending on the geologic origin of the shale. It is currently accepted that the hydrocarbons

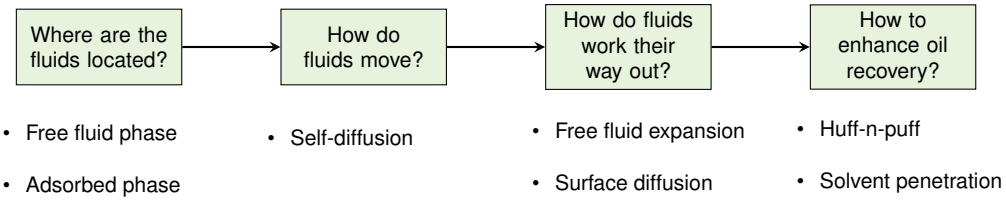


Figure 1.1: Summary flowchart of the course of this study.

in shale reservoirs are largely stored in the nanopores within organic matter (Rexer et al., 2014). The small size of the pores results in very large internal surface area relative to the pore volume, which in conjunction with the close proximity of the fluid molecules to the walls enhance their interactions and brings about the existence of an adsorbed phase that increases the storage capacity of shale reservoirs (Ambrose et al., 2012).

Molecular simulations have been increasingly used to to investigate the interactions between reservoir fluids and kerogen-like frameworks. In the early works, Müller et al. (2000) use grand canonical Monte Carlo simulation to study adsorption for a molecular model of water and water vapor/methane mixtures on activated carbon pores. Wongkoblap and Do (2008) also use grand canonical Monte Carlo simulations to study adsorption in a carbon slit pore of finite length. Methane is used as a model for the nonpolar molecule while water is the polar molecule. While Monte Carlo simulations sample configurations of a system, the drawback is that the time evolution of the system is not known. Ambrose et al. (2012) formulate a new gas-in-place equation that accounts for the pore space occupied by the adsorbed phase and use molecular dynamics simulations of methane in small carbon slit pores of varying size and temperature to predict density profiles across the pores. Seeking to provide insights of relevance in the oil industry, Mosher et al. (2013) use grand canonical Monte Carlo simulations to estimate the adsorption isotherms of methane across a range of pore sizes at various temperature and pressure conditions characteristic of subsurface conditions. The disadvantage of these works, however, is the use of simplistic surrogate models for kerogen such as graphene sheets (graphene is a honeycomb arrangement of carbon atoms). There have been also attempts to incorporate more details into the models. Cristancho et al. (2016) study supercritical methane storage influenced by nanoscale pore confinement and pore

wall surface heterogeneities such as vacancies and surface substitutions. In a more recent work, Herdes et al. (2018) use an experimental, theoretical, and molecular simulation consolidated approach to study adsorption and fluid phase behavior of multi-component hydrocarbon mixtures within tight shales represented by graphite walls.

Only a few years ago, Ungerer et al. (2015) created and made available kerogen macromolecules suitable for the construction of realistic models of kerogen. In a different approach Bousige et al. (2016) use a hybrid experimental-simulation method to build realistic molecular models of mature and immature kerogen that provide a detailed picture of the structure of kerogen. All these models have paved the way for studies in which more kerogen features and more complex fluid mixtures are considered. For instance, Collell et al. (2014) use molecular dynamics simulations to generate microporous structures of kerogen and grand canonical Monte Carlo simulations to produce pure methane and methane/ethane adsorption isotherms on kerogen. Wang et al. (2018) investigate competitive adsorption of methane/carbon dioxide in kerogen using grand canonical Monte Carlo simulations for CO_2 sequestration and enhanced oil recovery. Ho et al. (2018) use hybrid Monte Carlo/molecular dynamics simulation to study kerogen swelling associated with helium, methane and carbon dioxide. Thus the work done to investigate the adsorbed and free fluid phases in kerogen has largely focused on pure methane, methane/ethane and methane/carbon dioxide mixtures. Reservoir fluids, on the other hand, comprise more complex hydrocarbon mixtures.

To investigate confined fluid motion and the potential implications in oil recovery, Collell et al. (2015) use nonequilibrium molecular dynamics simulations to study the transport of hydrocarbon mixtures through a model of oil-prone type II kerogen and conclude that the transport mechanism through nanoporous kerogen is purely diffusive, and develop a surface diffusion theory in an attempt to create a general framework for mixture transport in kerogen. Falk et al. (2015) use the theory of statistical mechanics and nonequilibrium molecular dynamics simulations to propose that fluid transport in kerogen arises from a combination of a slip-like friction of hydrocarbon molecules with the pore walls and a free volume term. Later, Obliger et al. (2016) extend that theory

to hydrocarbon mixtures. Ho et al. (2016) study methane release in a kerogen matrix and determine that it is characterized by a fast release of pressurized free gas driven by a pressure gradient, followed by a slow release of adsorbed gas as the pore pressure decreases controlled by gas desorption and diffusion. Lee et al. (2016) use statistical models and molecular dynamics simulations to propose an activated interfacial transport model in which methane desorption is an activated process, i.e. methane molecules must overcome an energy barrier before desorption occurs.

The study of enhanced oil recovery in unconventional reservoirs using molecular simulations is in an incipient stage. Most of the work done uses carbon dioxide as the injection gas. Pathak et al. (2016) use molecular dynamics simulations to estimate self-diffusion coefficients of methane and carbon dioxide in kerogen as a proxy for gas retention in organic pores. Takbiri-Borujeni et al. (2019) use molecular dynamics to study soaking of a slit pore filled with dodecane with carbon dioxide and determine that there is an optimal soaking time after which production of dodecane is not affected significantly by soaking. Kazemi et al. (2019) expand this work and design a production stage after soaking with carbon dioxide.

This work intends to provide a systematic study of fluid behavior and transport in organic nanopores in shales. Chapter 2 describes the spatial organization of fluids inside organic pores and provides an estimate of the thickness of the adsorbed phase, which is used to quantify the fraction of molecules in the free and adsorbed phases. Chapter 3 discusses the random motion of the fluids in the pores in the absence of external gradients. The self-diffusion coefficients of the hydrocarbon species give a measure of their mobility in organic pores. Chapter 4 explores the migration of fluid molecules from within kerogen to a microfracture placed next to the kerogen model and provides insights into the composition of the fluids that migrate to the microfracture and those that remain in the pores. Chapter 5 implements a strategy to simulate primary depletion of kerogen, compares the composition of the produced and unproduced fluids, quantifies the recovery and explores the fluid transport mechanisms in nanoporous kerogen. Chapter 6 presents two cycles of a huff-and-puff enhanced oil recovery process using an organic solvent. The

concept of residence time is evaluated and the results are utilized to suggest an optimum design of a huff-and-puff strategy. Chapter 7 presents the molecular picture of the huff-and-puff process studied in Chapter 6. The penetration of the solvent into kerogen is investigated to assess where the oil-solvent mixing occurs and the transport diffusion coefficient of the solvent is estimated. Finally, Chapter 8 summarizes the lessons about confined fluid behavior and transport in kerogen learned from molecular simulations.

Chapter 2

SPATIAL DISTRIBUTION OF RESERVOIR FLUIDS

2.1 Introduction

The work presented in this chapter is a follow up of the article *Spatial distribution of reservoir fluids in mature kerogen using molecular simulations* (Perez and Devegowda, 2019). It focuses on the distribution of volatile oil and other reservoir fluids in kerogen pertaining to the gas condensate/volatile oil window. The main differences between the two works are listed below:

- More appropriate sets of force fields are utilized here to describe the atoms that belong to different molecular species. All atoms in the system in the previous work were described with one set of parameters that yielded accurate kerogen densities and pore size distributions but underestimated the (bulk or unconfined) density of reservoir fluids.

- Longer relaxation times are simulated to guarantee the achievement of equilibrium configurations.

- The reservoir conditions chosen as the final target in the simulations are a temperature $T = 355$ K (180°F) and a pressure $P = 300$ atm (4400 psi) to be consistent with the subsequent studies; in the previous work, the reservoir conditions were $T = 338$ K (149°F) and $P = 100$ atm (1470 psi).

- A more quantitative analysis using distribution functions is presented at the end of this chapter.

The models of organic matter created here are representative of kerogen type II-C, mature kerogen from the middle to the end of the oil formation window and contain pores in the range of micropores (pore diameter < 2 nm). Methane is the molecular species that accounts for most of the molecules that fill the pores while longer alkane chains such as ethane, propane, n-butane and n-octane transition from filling the pore to adsorbing on the pore surface. The heavier components n-tetradecane, toluene, dimethylnaphthalene, and the fraction of asphaltenes and resins are observed to form an adsorbed phase and potentially clog the pore throats. As a general trend, the longer the alkane chain, the more preferentially the molecule is accommodated closer to the pore surface. Additionally, the simulations correctly capture the hydrophobic environment in which water molecules cluster together to form droplets that touch the pore surface but do not wet it. Most molecules of carbon dioxide are part of the adsorbed phase while others diffuse into the kerogen structures and a small number are found in the free fluid phase. These observations underscore the significance of knowing fluid distributions to model oil recovery.

2.2 Organic matter components

Collell et al. (2014) built a molecular model of organic matter that contains a lump version of a volatile oil. Following the same rationale, larger models containing four times as many molecules were constructed, allowing the formation of larger pores in this work. Table 2.1 presents the number of molecules utilized to construct the molecular systems. The relative amount of each species (and therefore the mole fractions) are kept the same as in Collell et al. (2014) to preserve the atomic balance in mature kerogen that matches the data published by Tissot and Welte (1984, p. 176) and is reproduced in Table 2.2.

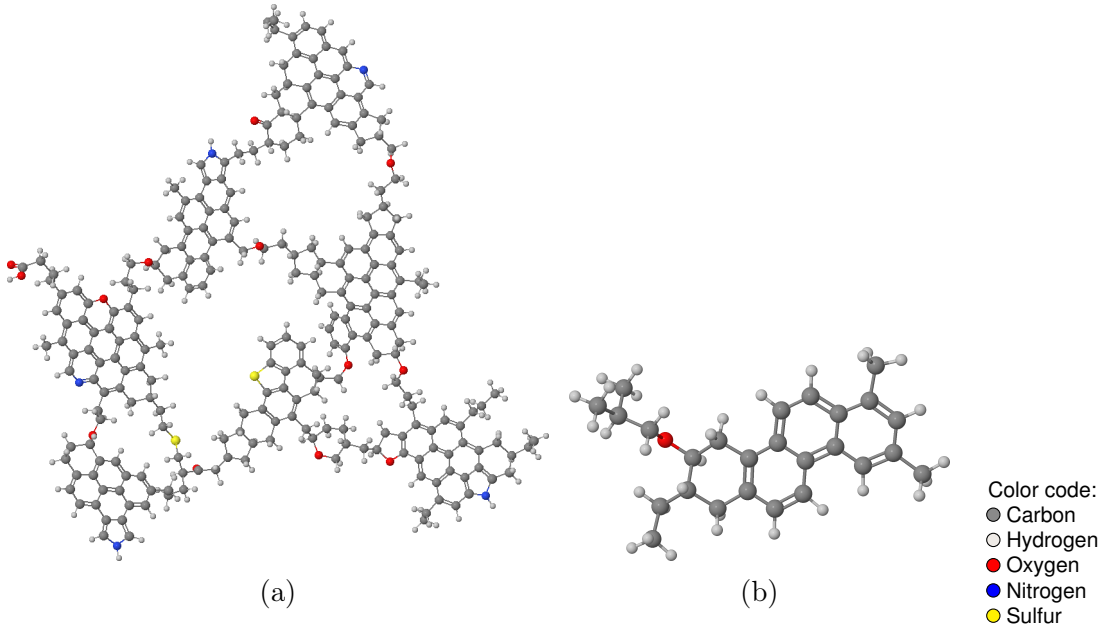
Table 2.1: Number of molecules per species used to create the models of organic matter.

Compound	Number of molecules
kerogen	16
methane	160
ethane	20
propane	12
n-butane	8
n-octane	16
n-tetradecane	12
toluene	8
dimethylnaphthalene	8
asphaltene/resin*	8
carbon dioxide	60
water	108

*See section 2.2.2.

Table 2.2: Atomic balance of the major compounds that make up organic matter present in mature kerogen (in the middle of the oil formation window). Modified from Tissot and Welte (1984, p. 176).

Compound/phase	C	H	O	N	S
kerogen	806	876	40	18	5
hydrocarbons	106	108			
resins	33	41	1.6	0.4	0.1
asphaltenes	22	23	1.6	0.4	0.1
carbon dioxide	5		10		
water		102	51		



(a) (b)

Figure 2.1: Molecular models used to construct organic matter models. (a) Kerogen type II-C monomer with chemical formula $C_{242}H_{219}O_{13}N_5S_2$ (Ungerer et al., 2015); (b) generic model used to represent the asphaltene/resin fraction, with chemical formula $C_{26}OH_{32}$ (Collell et al., 2014). *The figures are not at the same scale.*

2.2.1 Kerogen

Kerogen is defined as the naturally occurring solid organic matter located in source rocks (shale) that is non-extractable using organic solvents and can yield oil upon heating (Durand, 1980; Tissot and Welte, 1984). Ungerer et al. (2015) designed kerogen monomers that can be used to construct realistic kerogen models that reproduce the elemental and functional analysis data reported by Kelemen et al. (2007). The monomer that represents kerogen type II-C with chemical formula $C_{242}H_{219}O_{13}N_5S_2$ corresponds to oil-prone kerogen found in organic-rich marine shales from the middle to the end of the oil formation window and is presented in Figure 2.1(a). The attributes of this monomer model are compared to experimental characterization work in Ungerer et al. (2015).

To capture most of the structural features of kerogen, the models of organic matter must have a minimum number of 1300 atoms (Yiannourakou et al., 2013; Collell et al., 2014; Ungerer et al., 2015), thus four kerogen monomers would suffice to reach that number but 16 monomers are used in this work to create bigger pores.

2.2.2 Asphaltenes and resins

Asphaltenes and resins are aromatic macromolecules with overall composition similar to that of kerogen but with a lower molecular weight (Tissot and Welte, 1984). Resins are more polar than aliphatic hydrocarbons and soluble in organic solvents like n-alkanes while asphaltenes are insoluble in n-alkanes but soluble in light aromatics such as benzene and toluene (Mitchell and Speight, 1973; Tissot and Welte, 1984). Since the fraction of atoms in asphaltenes and resins is relatively small compared to the fraction of atoms in mature kerogen and the associated hydrocarbons (Tissot and Welte, 1984), and due to the similarities in composition between asphaltenes and resins, Collell et al. (2014) devised a single generic molecule with chemical formula $C_{26}OH_{32}$ to represent the asphaltene and resin fraction in simulations and is illustrated in Figure 2.1(b). Nitrogen and sulfur are not included in that model due to their low concentration in these compounds (see Table 2.2).

2.2.3 Water and carbon dioxide

Assuming a closed system, Tissot and Welte (1984) proposed a mass balance between the immature kerogen and the mature organic matter along with the generated hydrocarbons under the presumption that the components generated during the maturation process are not expelled from the system. They report the atomic balance of the total organic matter in the evolution from one thermal maturity level to another. Ungerer et al. (2015) and Collell et al. (2014, 2015) used the same argument and proposed that the elemental analysis of the mature organic matter plus the generated hydrocarbons has to match the atomic balance of immature kerogen detailed in Tissot and Welte (1984, p. 176) (see also Table 2.2). By that rationale, Collell et al. (2014) justify the inclusion of carbon dioxide and water in the system, and therefore their inclusion in the present work as well.

Table 2.3: Molecular composition of the hydrocarbon fraction.

Species	Mole fraction (%)	Weight fraction (%)
methane	63.5	19.4
ethane	7.9	4.5
propane	4.8	4.0
n-butane	3.2	3.5
n-octane	6.3	13.8
n-tetradecane	4.8	18.0
toluene	3.2	5.6
dimethylnaphtalene	3.2	9.4
asphaltene/resin	3.2	21.8
Total molecular weight	52.55 g/mole	

2.2.4 Hydrocarbons

Lagache et al. (2004) created a method to model natural gas compositions with a limited number of linear, branched and cyclic hydrocarbon molecules. Ungerer et al. (2015) used lump fractions to represent the composition of a volatile oil with a molecular weight between 40 and 70 g/mole and a C_{7+} mole fraction greater than 10% (McCain, 1990). The use of lump fractions to model oil mixtures is justified because the use of well-defined components would require a detailed analysis of all the heavy hydrocarbon isomers and the problem would become intractable using molecular simulations. The composition of the hydrocarbon mixture used in this work is given in Table 2.3.

2.3 Simulation details

As summarized in Table 2.1, 16 kerogen type II-C monomers, 8 asphaltene/resin molecules, 160 methane molecules, 20 ethane molecules, 12 propane molecules, 8 n-butane molecules, 16 n-octane molecules, 12 n-tetradecane molecules, 8 toluene molecules, 8 dimethylnaphtalene molecules, 60 carbon dioxide molecules, and 108 water molecules are placed randomly in an empty cubic box with side 100 Å using PACKMOL (Martínez et al., 2009). The atoms in kerogen are described by the consistent-valence force field (cvff) (Dauber-Osguthorpe et al., 1988) that, although optimized for proteins, provides

results for density and pore size distribution of kerogen comparable with those of Ungerer et al. (2015) using the pcff+ force field (Ho et al., 2016; Michalec and Lísal, 2017; Pathak et al., 2017; Perez and Devegowda, 2017, 2019). The parameters to describe the atoms in hydrocarbon molecules are taken from the optimized potentials for liquid simulations all-atom force field (OPLS-AA) (Jorgensen et al., 1996) with the refinement to the torsional parameters described in Price et al. (2001). Atoms in water molecules are simulated using the rigid SPC/E model (Berendsen et al., 1987) and atoms in molecules of carbon dioxide are described with the CLAYFF force field (Cygan et al., 2012). All the force field parameters are summarized in Appendix B. The system undergoes the long simulated annealing process (Kirkpatrick et al., 1983) carried out in LAMMPS (Plimpton, 1995) described in Table 2.4. Periodic boundary conditions are used in all directions. A cutoff distance of 15 Å is selected in evaluating the non-bonded interactions Lennard-Jones and Coulombic, long distance corrections are used to estimate Lennard-Jones energy beyond the cutoff (Sun, 1998) and the particle-particle particle-mesh method is used for electrostatic interactions (Hockney and Eastwood, 1988). Temperature and pressure are controlled by the Nosé–Hoover thermostat (Nosé, 1984a,b, 1991) and the Parrinello–Rahman barostat (Parrinello and Rahman, 1981), respectively. The SHAKE algorithm is used to keep water molecules rigid (Ryckaert et al., 1977). The campaign described in Table 2.4 is carried out three times, starting with different random initial configurations to obtain three different models of organic matter.

2.4 Organic matter models

The three models of organic matter obtained at the reservoir conditions $T = 355$ K (180°F) and $P = 300$ atm (4400 psi) are shown in Figure 2.2. The figures on the left present both kerogen and fluid atoms, while in the figures on the right the kerogen atoms are removed to visualize the pore geometry. The models enclose a large pore connected to other pores through smaller pore throats, which are the constrictions towards the faces of the simulation cells. This is a consequence of the periodic boundary conditions in all

Table 2.4: Simulated annealing process to create organic matter models. NVT means constant number of particles N, volume V, and temperature T, NPT means constant number of particles N, pressure P, and temperature T (see Appendix A), Δt is the time step used in each simulation and t is the total duration of every simulation.

Ensemble	T (K)	P (atm)	Δt (fs)	t (ns)
NVT	1000		1.0	0.5
NPT	1000	400	1.0	0.5
NPT	$1000 \rightarrow 900$	400	1.0	0.5
NPT	$900 \rightarrow 800$	400	1.0	0.5
NPT	$800 \rightarrow 700$	400	1.0	0.5
NPT	$700 \rightarrow 600$	400	1.0	0.5
NPT	$600 \rightarrow 500$	400	1.0	0.5
NPT	$500 \rightarrow 400$	400	1.0	0.5
NPT	$400 \rightarrow 338$	400	1.0	0.5
NPT	338	100	1.0	2.0
NVT	$338 \rightarrow 2000$		0.5	0.025
NVT	$2000 \rightarrow 355$		0.5	1.0
NPT	355	300	0.5	4.0

spatial directions.

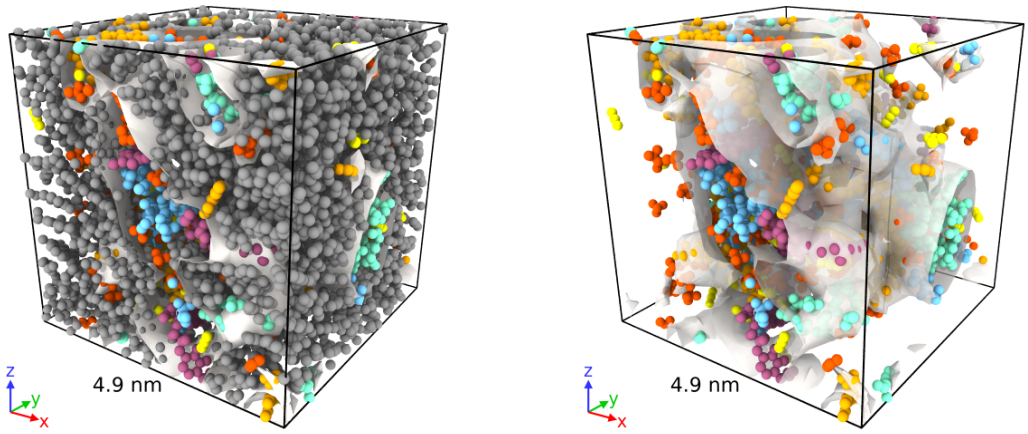
Once the models are created, some physical properties of the kerogen models such as pore size distribution, internal surface and density are computed and compared with experimentally determined values to estimate the goodness of the models.

2.4.1 Pore surface, porosity and density

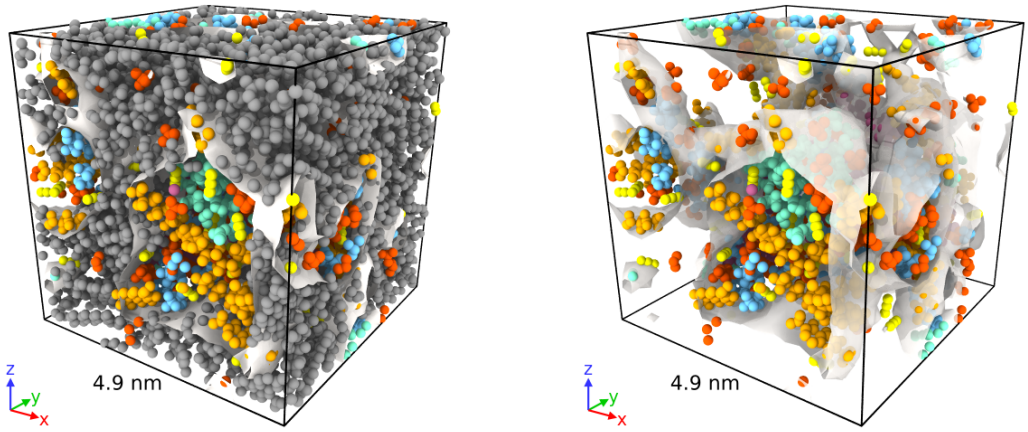
Using the Python interpreter OVITOS that comes with OVITO distributions (Stukowski, 2010), the average pore surface and solid volume fractions are computed by constructing a polyhedral mesh around the kerogen atoms, generating a geometric representation of the pore surface (Stukowski, 2014), as depicted in Figure 2.2. The radius of the probe sphere is chosen so that it represents a helium probe, and the last 2000 simulation frames are utilized to compute the average values, which corresponds to averages over 1 ns when the systems have reached equilibrium.

The porosity ϕ of the models is

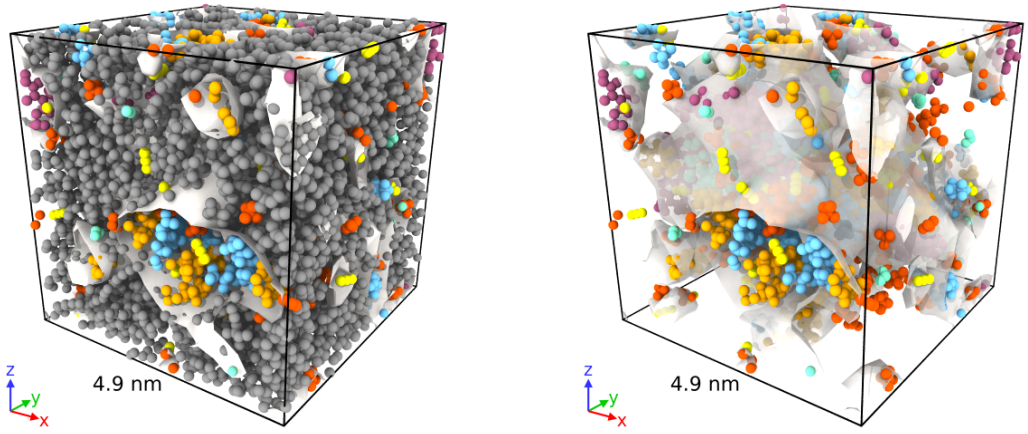
$$\phi = 1 - V_f \tag{2.1}$$



(a) Model 1



(b) Model 2



(c) Model 3

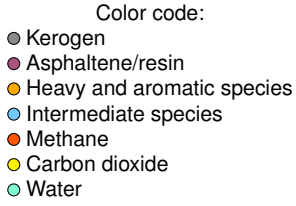


Figure 2.2: Molecular models of organic matter. In the figures on the left, both kerogen and fluids are present; in the figures on the right, kerogen atoms are removed to visualize the pore geometry (transparent gray surface) and the fluids inside the pores.

Table 2.5: Physical properties of the models of organic matter.

Model	Surface area (m^2/g)	Solid volume fraction	Porosity (%)	Density (g/cm^3)
1	1411.15	0.59	41.52	1.33
2	1367.93	0.58	41.70	1.33
3	1435.61	0.60	40.18	1.31

where V_f is the solid volume fraction, and the kerogen density is computed using the equation

$$\rho_k = \frac{m_k}{V_k} = \frac{m_k}{V(1-\phi)} \tag{2.2}$$

where m_k is the mass of kerogen, V_k is the volume occupied by kerogen (i.e. it excludes the pore volume), and V is the volume of the simulation box. The physical properties of the models are summarized in Table 2.5. On average, the kerogen density is 1.32 $\pm$ 0.02 g/cm^3 at 355 K and 300 atm. In comparison, Dang et al. (2016) experimentally determined that the density of kerogen type II ranges from 1.4 to 1.6 g/cm^3, Okiongbo et al. (2005) reported a density range between 1.18 and 1.38 g/cm^3 for isolated kerogen type II with similar maturity levels, and Collell et al. (2014) obtained a kerogen density of 1.42 ± 0.01 g/cm^3 at 300 K and 20 MPa (197 atm) using molecular dynamics simulations.

2.4.2 Pore size distribution

The pore size distribution of each model is computed using a helium probe with the PSDsolv-1.01 package (Bhattacharya and Gubbins, 2006) and Figure 2.3 presents them collectively. The peak around 5 Å[1] corresponds to the distance between the kerogen monomers that make up the organic matter; the rightmost peaks represent the maximum pore size; and the peaks in the range 7 to 15 Å correspond to smaller pores and pore throats. In Sections 2.5.1 and 2.5.2 it is discussed the potential effect that adsorption of large molecules may have on the connectivity of organic pores by comparing pore size distribution curves of kerogen with and without an adsorbed phase.

[1] 1 nm = 10 Å.

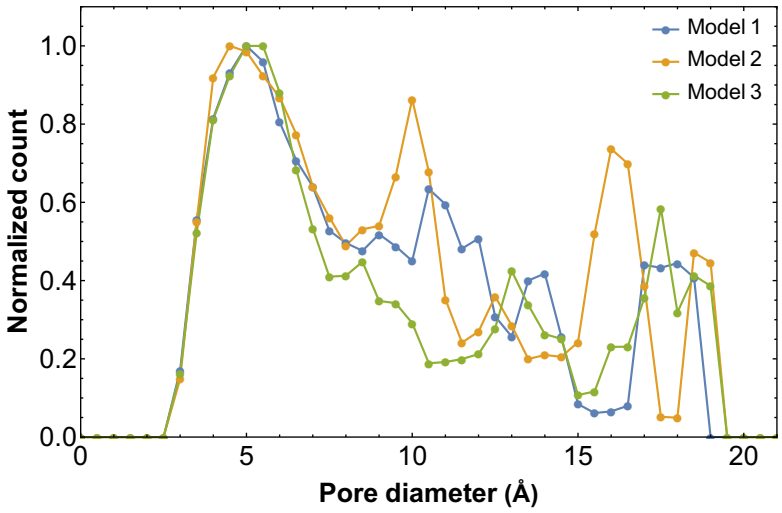


Figure 2.3: Pore size distribution of the models of organic matter.

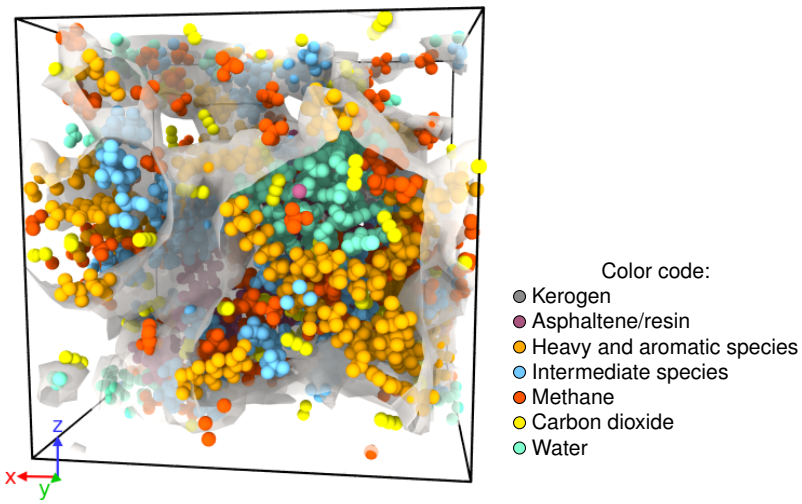


Figure 2.4: Reservoir fluids filling an organic pore. The pore render is the transparent gray surface. Kerogen atoms are not shown. The length of each box edge is 4.9 nm.

2.5 Spatial distribution of fluids

The fluid phase comprises light and intermediate linear alkanes, small and medium size aromatic molecules, an asphaltene/resin fraction, carbon dioxide, and water. These compounds are expected to occupy the pore space in the models with a few molecules of the smallest species (water, carbon dioxide, and methane) diffusing into the kerogen structures, as illustrated in Figure 2.4 where all the fluid species are shown. The perspective of the image is the same as in the figures in the subsequent sections to allow direct comparisons. What follows next is a qualitative description of the spatial distribution of the different fluid species.

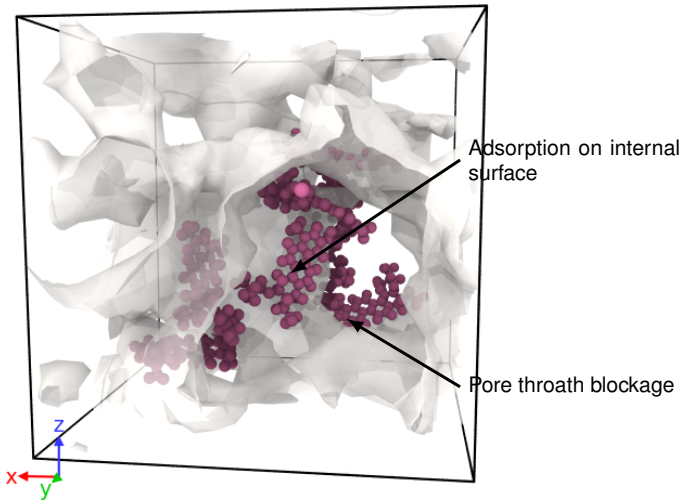


Figure 2.5: Adsorption of asphaltenes and resins on the pore surface. Asphaltene/resin atoms are purple. The length of each box edge is 4.9 nm.

2.5.1 Asphaltene/resin fraction

Due to the large surface area of asphaltenes and resins, this fraction adsorbs on the pore surface, either in the interior of the large pore or in the pore throats. Adsorption of heavy hydrocarbon species in organic matter at reservoir temperature and pressure has been reported previously by Herdes et al. (2018) and Falk et al. (2015). When asphaltenes and resins adsorption occurs in the large pore, it forms a new surface on which other species might adsorb; when adsorption occurs in the pore throats, these large molecules potentially block them preventing the flow of smaller fluid molecules between pores. Figure 2.5 illustrates these two cases.

2.5.2 Heavy components

As in the case of asphaltenes and resins, the relatively large surface area of the aromatic species dimethylnaphthalene and toluene promotes their adsorption on the pore surface. Similarly, the long length of n-tetradecane and its ability to bend also promotes adsorption of this molecular species. Figure 2.6 gives an example of their spatial distribution.

The adsorption of the larger species on the surface of kerogen has a severe effect on pore connectivity, especially when adsorption occurs in the pore throats. Figure 2.7

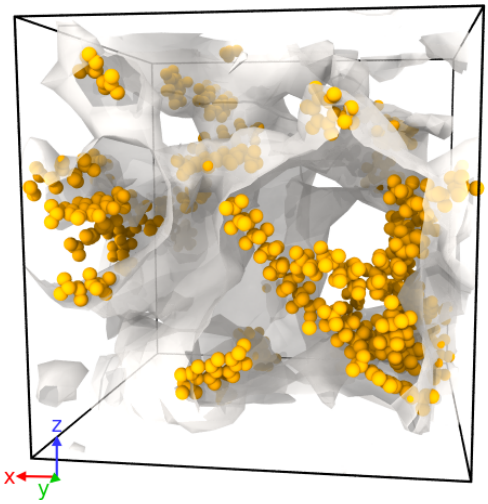
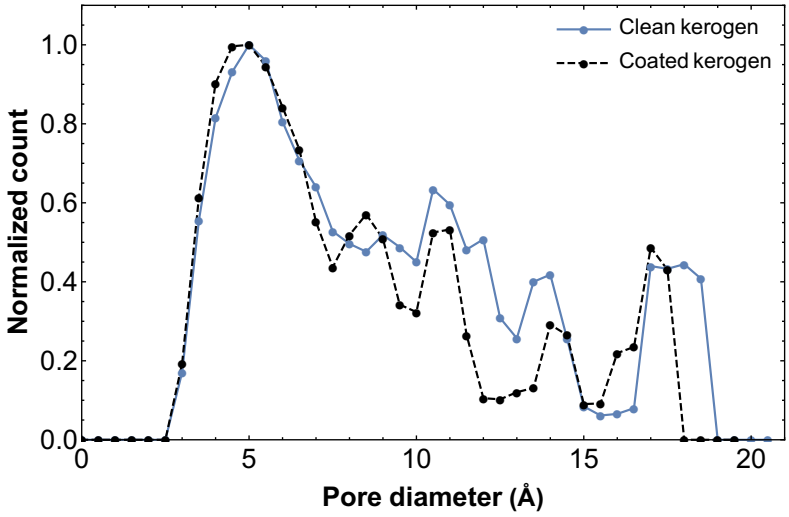


Figure 2.6: Adsorption of the heavy molecular species dimethylnaphthalene, toluene and n-tetradecane on the pore surface. Heavy atoms are dark yellow. The length of each box edge is 4.9 nm.

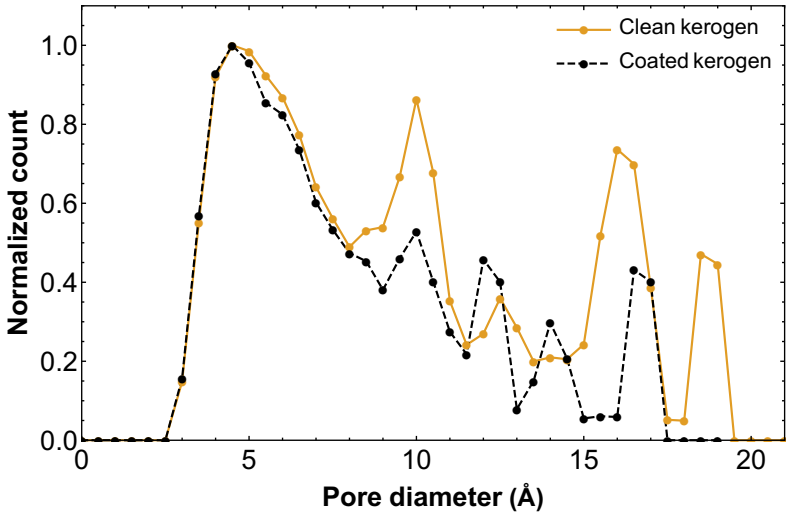
presents a comparison between kerogen without any fluids (referred to as *clean* kerogen) and kerogen with the aromatic molecules asphaltene/resin and dimethylnaphthalene (referred to as *coated* kerogen), whose mobility under confinement is very low, as discussed in Section 3.6. The shrinking and shifting the peaks to the left in the range 7 to 15 Å demonstrate that pore throats are reduced both in size and amount due to the presence of these adsorbed species. If other heavy molecules were included in coated kerogen, the effects on the pore size distributions would be even more dramatic. Figure 2.7 also suggests that removing the heavier components from kerogen using some solvent can ultimately result in opening pore throats that were blocked and an increase in porosity. In fact, Sinha et al. (2017) experimentally observed between 24 and 40% increase in porosity in kerogen samples whose maturity spans the oil and condensate windows after extraction of soluble bitumen and residual hydrocarbons using solvent extraction. This observation has huge implications for the design of enhanced oil recovery projects, as will be discussed in Chapter 6.

2.5.3 Intermediate components

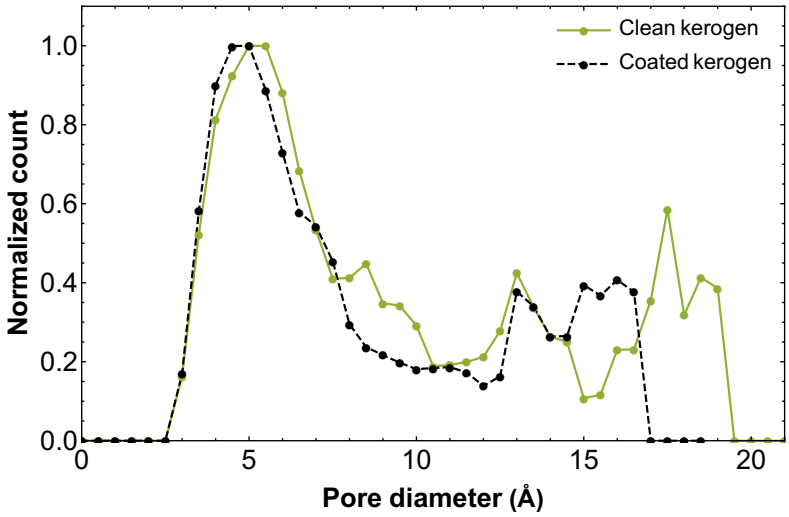
The intermediate components n-octane, n-butane, propane, and ethane transition between adsorbing on the pore surface and filling the pore. As a general observation, the



(a) Model 1



(b) Model 2



(c) Model 3

Figure 2.7: Comparison between pore size distributions of the clean and coated kerogen models.

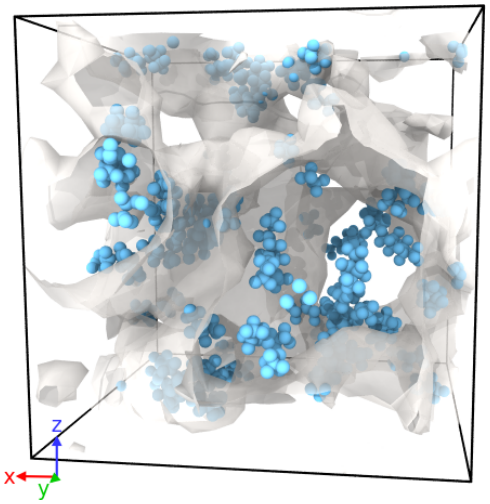


Figure 2.8: Distribution of the middle sized alkanes n-octane, n-butane, propane, and ethane in organic pores. Intermediate atoms are blue. The length of each box edge is 4.9 nm.

longer the alkane chain, the more likely it will be located closer to the pore surface. Thus, n-octane molecules preferentially adsorb whereas a larger fraction of ethane molecules are found in the free fluid phase contributing to fill the pore. Figure 2.8 presents their locations.

2.5.4 Methane

Methane being the lightest component with the smallest surface area accounts for most of the free phase filling a large volume of the pores. However, a large fraction of methane molecules are found adjacent to the pore surface while a few others diffuse into the kerogen body, as will be shown in Section 2.6.2. Adsorption of methane in organic pores has been previously reported (Gasparik et al., 2012; Zhang et al., 2012; Rexer et al., 2013; Collell et al., 2014; Gasparik et al., 2014; Heller and Zoback, 2014; Rexer et al., 2014; Tinni et al., 2018; Perez and Devegowda, 2017). In the presence of competitive adsorption between different hydrocarbon species, methane constitutes the greatest percentage of the free fluid phase as shown in Figure 2.9.

2.5.5 Carbon dioxide

The majority of carbon dioxide molecules are part of the adsorbed phase, while a small number are present as a part of the free fluid or within the kerogen structure, as

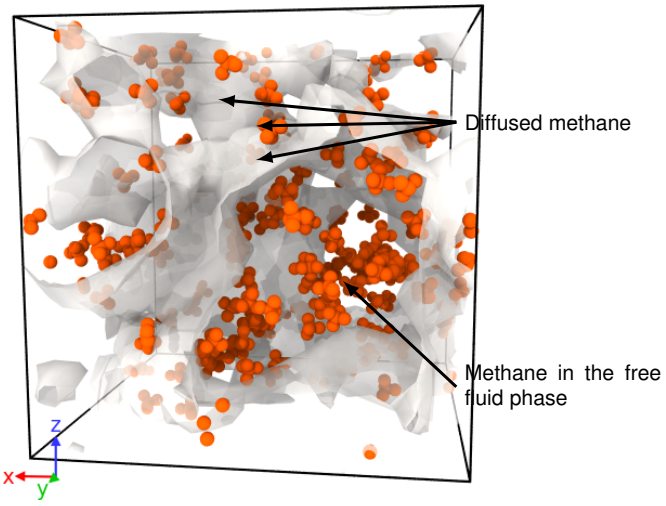


Figure 2.9: Distribution of methane in organic pores. Methane atoms are orange. The length of each box edge is 4.9 nm.

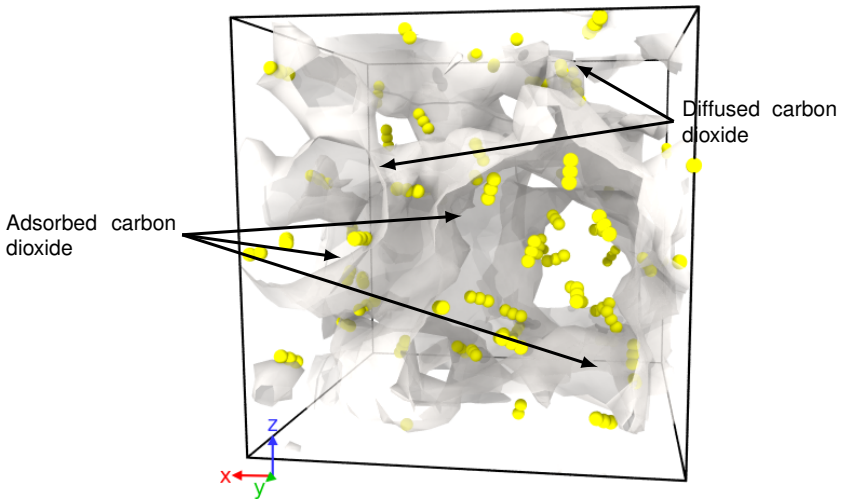


Figure 2.10: Distribution of carbon dioxide in organic pores. Carbon dioxide atoms are yellow. The length of each box edge is 4.9 nm.

shown in Figure 2.10. It is interesting that, due to competitive adsorption with hydrocarbon molecules, preferential adsorption of carbon dioxide is dominant in the organic matter models considered here, as demonstrated in Section 2.6.2.

2.5.6 Water

Due to the hydrophobic environment in kerogen, almost all water molecules cluster together forming a droplet that resides in one of the pore throats (or close to one of them), partially clogging it. Collell et al. (2014) also noticed the formation of small water droplets within organic matter. A few water molecules also diffuse into kerogen as

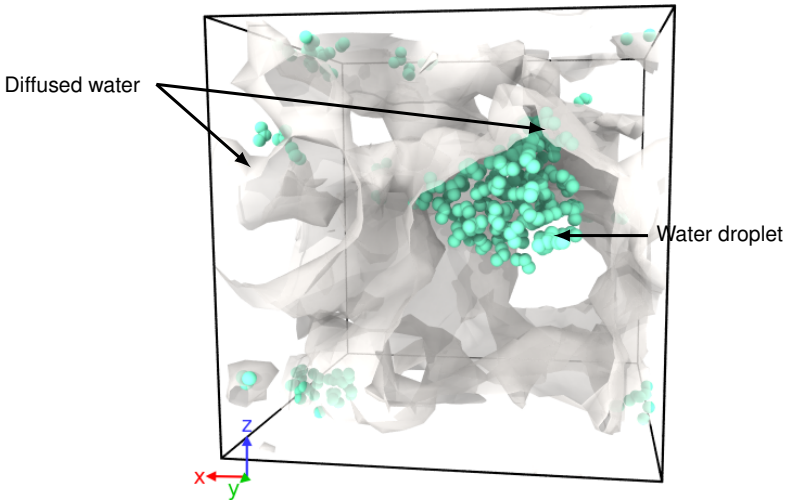


Figure 2.11: Distribution of water in organic pores. Water atoms are green. The length of each box edge is 4.9 nm.

illustrated in Figure 2.11.

2.6 A deeper look into the spatial distribution of fluids

In the previous section the analysis presented was purely descriptive. In this section a more quantitative analysis is carried out. Particularly, it focuses on the evaluation of the macroscopic property called radial distribution function $g(r)$ that describes the spatial organization of atoms about a central atom or molecule, and the estimation of the thickness of the adsorbed phase onto the organic pore surface.

2.6.1 Radial distribution function

The radial distribution function $g(r)$ measures how atoms organize themselves around one another, giving rise to a local structure (Haile, 1992). It is proportional to the probability of finding two atoms separated by a distance $r \pm \Delta r$, as illustrated schematically in Figure 2.12. For atomic systems, it can be determined experimentally using x-ray and neutron diffraction techniques (Haile, 1992; Billinge and Levin, 2007).

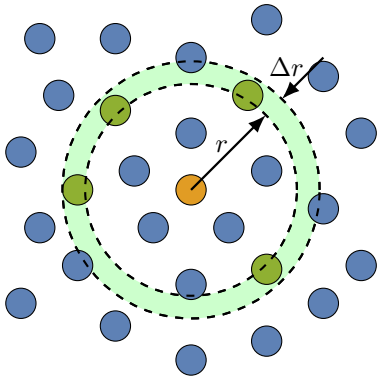


Figure 2.12: Spatial organization of atoms. The center of green atoms are located within the spherical shell of radius r and thickness Δr centered on the yellow atom.

Haile (1992) provides a detailed procedure to obtain $g(r)$ using trajectories generated with molecular dynamics simulations. The mathematical expression is

$$g(r) = \frac{\langle N(r, \Delta r) \rangle}{\frac{1}{2} N \rho V(r, \Delta r)} = \frac{\sum_{k=1}^{M} N_k(r, \Delta r)}{\frac{1}{2} M N \rho V(r, \Delta r)} \tag{2.3}$$

where N is the number of atoms in the simulation, V is the volume of the simulation box, $\rho = N/V$ is the number density, $N(r, \Delta r)$ is the number of atoms found in a spherical shell of radius r and thickness Δr with the shell centered on another atom, $\langle \cdots \rangle$ represents the time average of some function, $V(r, \Delta r)$ is the volume of the spherical shell, $N_k(r, \Delta r)$ is the number of atoms found in a spherical shell at time t_k in the run, and M is the total number of time steps. Equation (2.3) can be interpreted as the ratio of the local number density $\rho(r)$ to the system number density ρ.

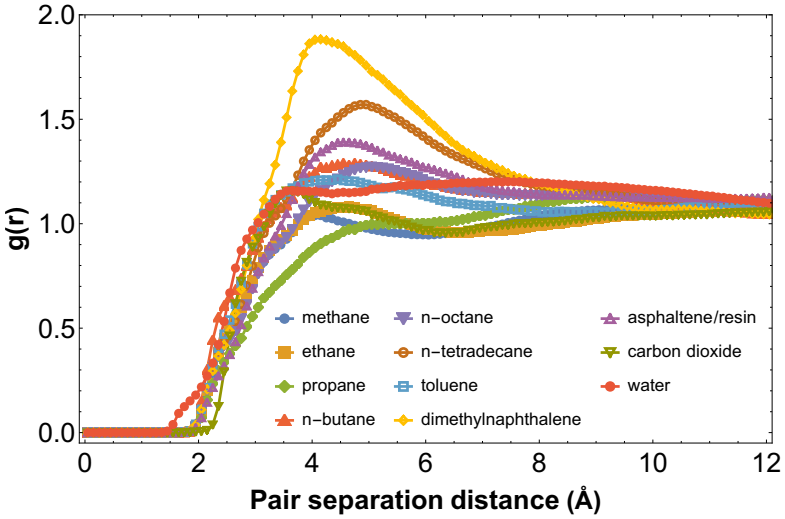
The goal is to provide a measure of the probability of finding the atoms of a certain molecular species at some distance r away from the pore surface, which in turn translates into estimating the adsorbed and free fluid phases in organic pores. To this end, the following procedure was implemented using the Python interpreter OVITOS that comes with OVITO distributions (Stukowski, 2010).

First, the kerogen atoms that lie on the pore surface are selected, then the radial distribution function of the atoms in the pore surface and all the atoms that belong to a fluid species is computed, and these functions are averaged over the last 2000 simulation frames after the system has reached equilibrium. This corresponds to simulation times of

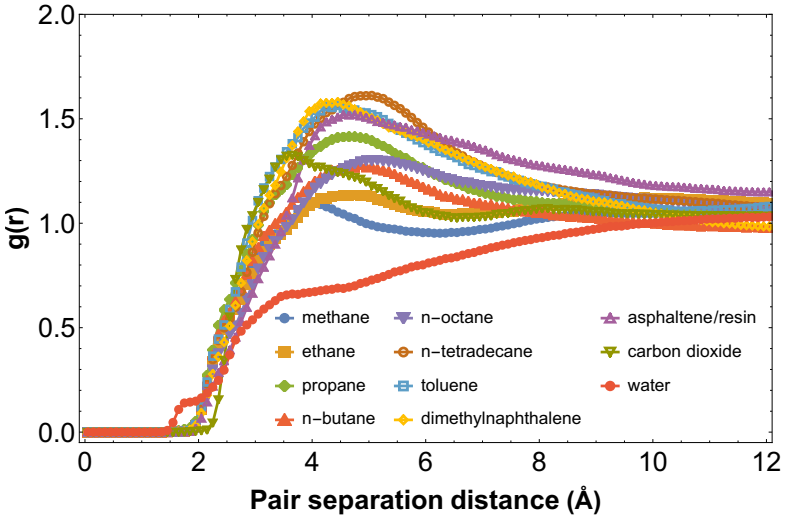
1 ns in equilibrium. The same procedure is repeated for each fluid species and the results are shown in Figure 2.13 where the vertical scale is the same in all cases to allow direct comparisons.

In general the radial distribution functions in Figure 2.13 look similar in the three cases, meaning that on average the spatial distribution of fluid molecules is consistent regardless of the specific geometry of the pores. An interesting observation is that the first curve that develops corresponds to water atoms, which in terms of probabilities means that atoms in water molecules may be found closer to the pore surface than atoms in any other molecule. Thus water droplets may contact the pore surface although water is not the wetting phase. The contact points might correspond to regions where the local chemistry of kerogen is more polar due to the presence of oxygen or nitrogen atoms.

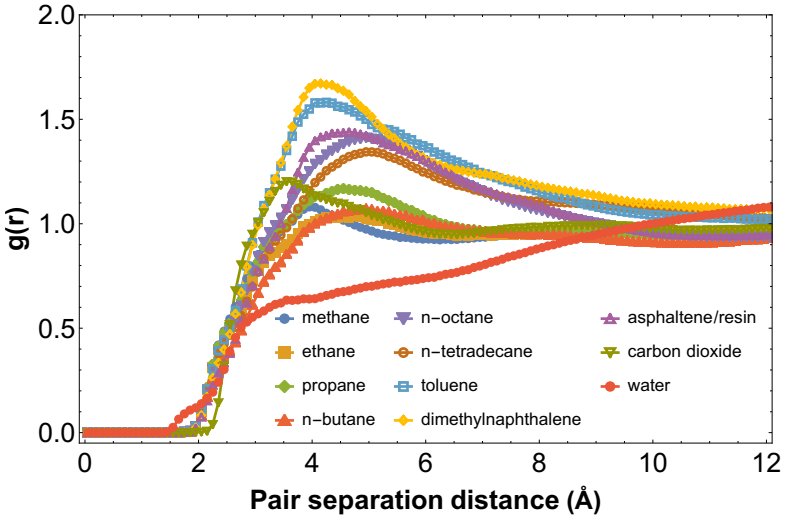
Table 2.6 summarizes the location of the peaks in the radial distribution functions presented in Figure 2.13 and Figure 2.14 displays the average locations listed in the last column in Table 2.6. Interpreting these peaks as the distance from the pore surface at which it is more likely to find atoms of each species, it can be said that the heavier species n-butane, n-octane, n-tetradecane, toluene, dimethylnaphthalene, and asphaltenes and resins may be found closer to the pore surface whereas the lighter components propane, ethane and methane may be found filling the pores, therefore their large error bars. Most atoms in carbon dioxide are likely located at a small distance from the pore surface, which means this species preferentially adsorbs onto kerogen over hydrocarbons. The average location of the peak that corresponds to atoms in water molecules is the farthest from the pore surface, which is expected as the environment is hydrophobic and water does not wet the pore surface although a few water molecules are adjacent to the pore surface. Additionally, the consistency in the location of the peaks of the asphaltene and resin fraction and the heavier components makes their error bars small compared to the lighter species propane, ethane and methane, this being a consequence of the preferential adsorption of the heavier species to the pore surface and the more spread presence of the lighter ones in the free fluid phase. These locations, however, might depend on factors such as the concentration of each species, and the condition of temperature and pressure



(a) Model 1



(b) Model 2



(c) Model 3

Figure 2.13: Radial distribution function of surface kerogen atoms and atoms of every molecular species.

Table 2.6: Location of the peaks (in Å) in the radial distribution functions.

Species	Model			Average
	1	2	3	
methane*	12.75	4.05	3.95	6.92 ± 5.05
ethane*	13.15	4.75	4.75	7.55 ± 4.85
propane	9.55	4.65	4.55	6.25 ± 2.86
n-butane	4.75	4.85	4.95	4.85 ± 0.10
n-octane	5.05	5.15	4.95	5.05 ± 0.10
n-tetradecane	4.95	5.05	4.95	4.98 ± 0.06
toluene	4.45	4.45	4.25	4.38 ± 0.12
dimethylnaphthalene	4.15	4.45	4.15	4.25 ± 0.17
asphaltene/resin	4.55	4.55	4.65	4.58 ± 0.06
carbon dioxide	3.55	3.65	3.55	3.58 ± 0.06
water*	7.45	19.95	13.75	13.72 ± 6.25

*Peaks beyond 12 Å are not shown in Figure 2.13.

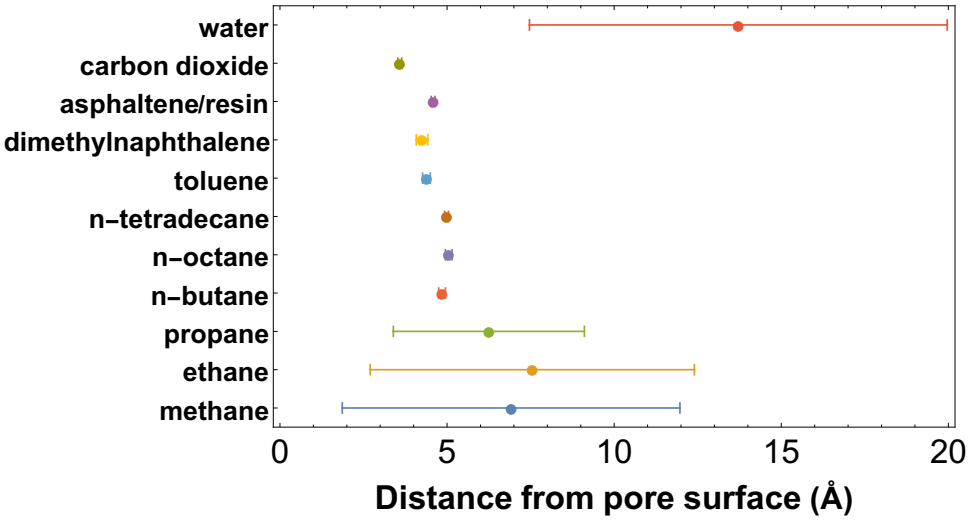


Figure 2.14: Average location of the peaks in the radial distribution functions.

of the system.

2.6.2 Estimation of the thickness of the adsorbed phase

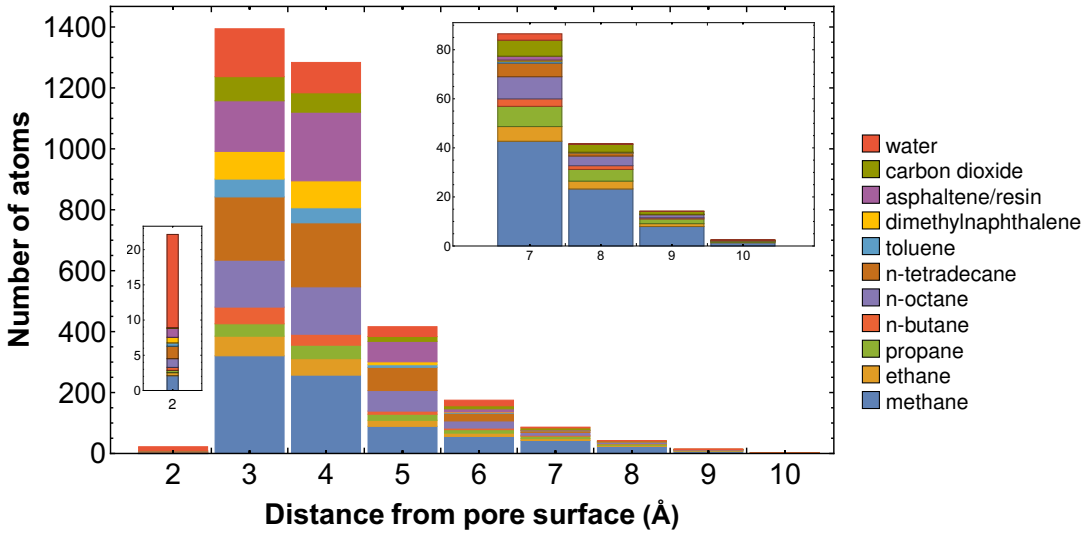
In addition to the radial distribution functions presented in the previous section, it is also possible to count the number of atoms located at a distance range between r and $r+\Delta r$ away from the pore surface using the Python interpreter OVITOS that comes with OVITO distributions (Stukowski, 2010). The kerogen atoms that lie on the pore surface are selected, then the number of atoms in the fluid species that are located at a distance $r < r_c$ away from the pore surface is computed, where r_c is the cutoff distance of interest

and is varied in intervals of 1 Å. The counts are averaged over the last 2000 simulation frames after the system has reached equilibrium. This corresponds to simulation times of 1 ns in equilibrium. The histograms in Figure 2.15 present the raw data obtained in this process. The close-ups of the data between 1 and 2 Å, as shown in the insets, demonstrate that atoms in water molecules are the closest to the pore surface. Figure 2.15 also shows that atoms in methane and water molecules are located across the pores, as discussed in the previous section.

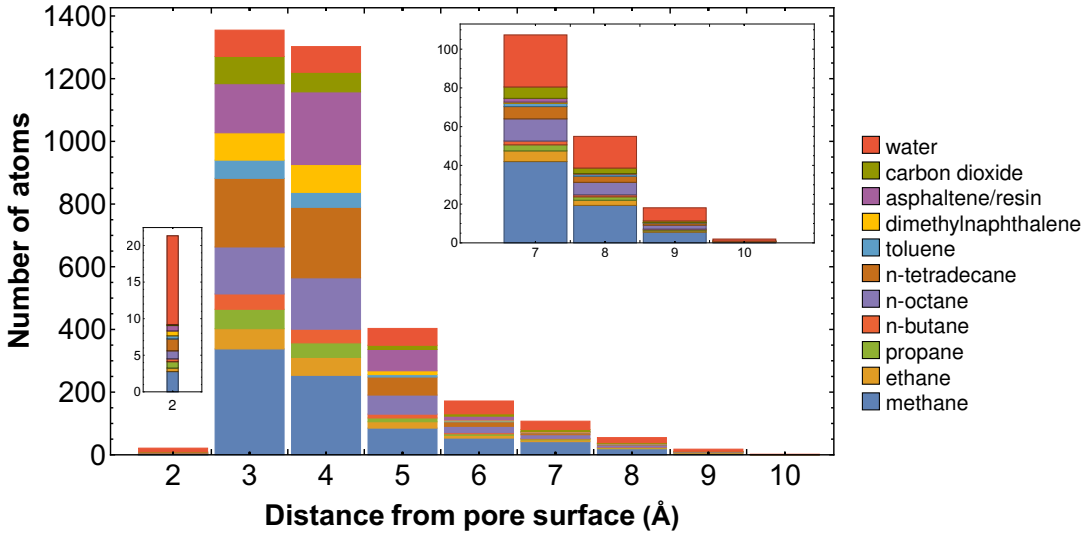
With some manipulations, the same information can be presented in a more informative manner such as the number of molecules for each species as a function of distance from the pore surface, which can be used to quantify adsorbed and free fluid compositions. Figure 2.16 presents the bar charts of the number of molecules located at some interval distance away from the pore surface.

Additionally, the fraction of molecules of each species located at some interval distance away from the pore surface can be also plotted to get a better sense of where most of the molecules per species reside inside the organic pores. This information is presented by the matrix plots in Figure 2.17. Considering Figures 2.16 and 2.17, the most prevalent molecules in the range between 1 and 2 Å away from the pore surface correspond to water and most carbon dioxide molecules are found within 4 Å away from the pore surface. Also, methane and water molecules are located across the pore space, just as stated above.

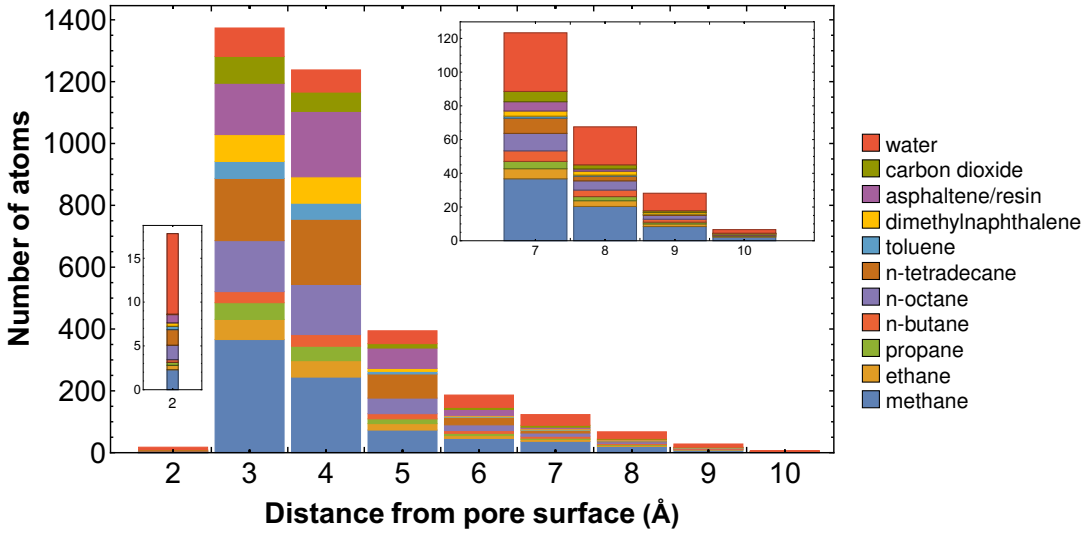
Clearly, Figures 2.15 and 2.16 show that most of the molecules are located within 4 and 5 Å away from the pore surface. Although these histograms do not represent directly number density profiles, they are related because the thickness of the shells used to count particles is 1 Å. However, the geometry of the shells is complicated as the pore surfaces are irregular as illustrated in Figure 2.2. The drastic drop in the heights of the bars between 4 and 5 Å suggests that the adsorbed phase comprises the molecules located no further than 4 Å away from the pore surface. In fact, the matrix plots in Figure 2.17 indicate that roughly 80% of the total number of hydrocarbons and carbon dioxide molecules are found in that range. Water molecules, on the other hand, are more spread



(a) Model 1

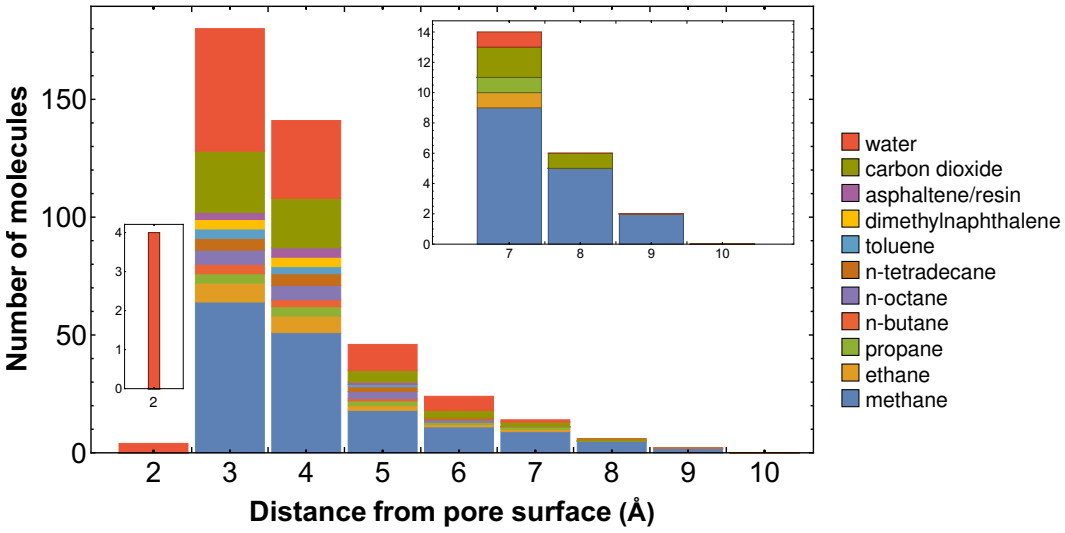


(b) Model 2

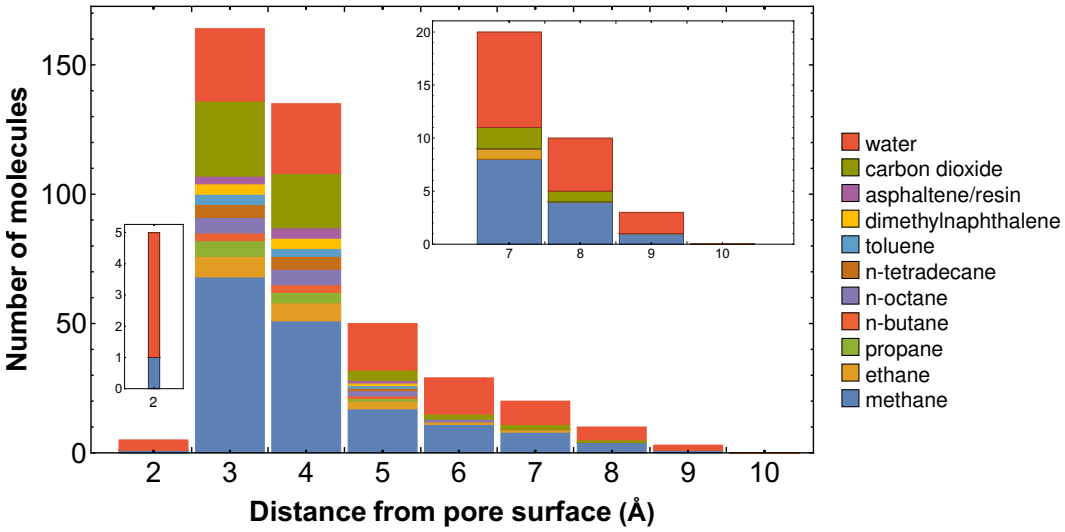


(c) Model 3

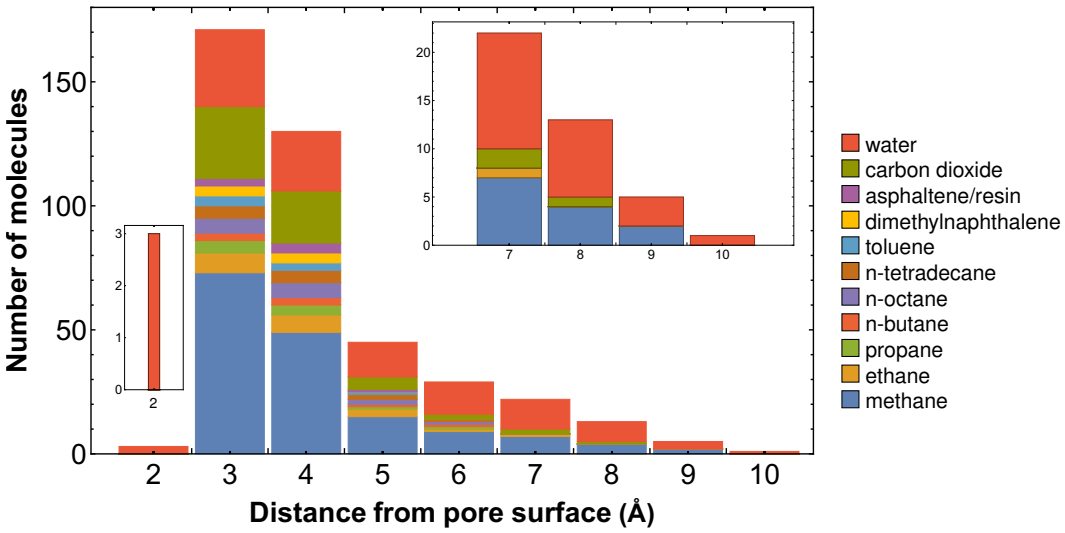
Figure 2.15: Histogram of the number of atoms as a function of the distance away from the pore surface.



(a) Model 1

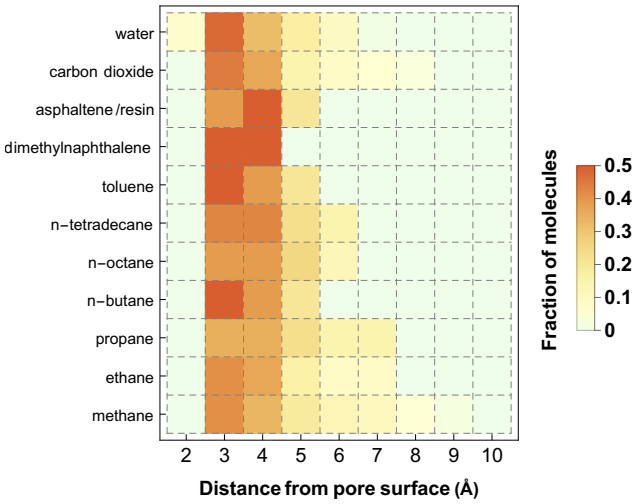


(b) Model 2

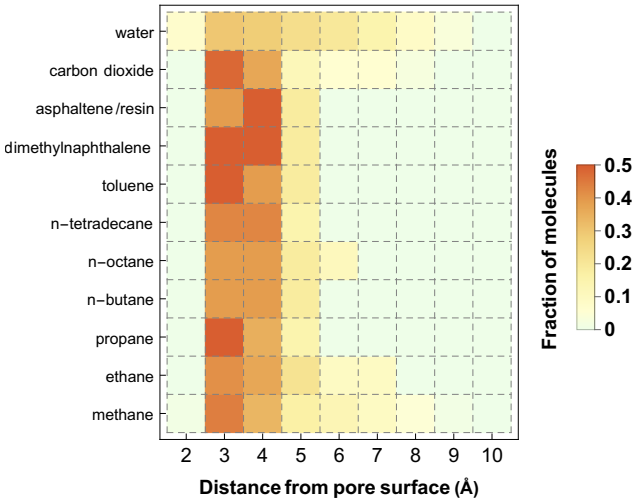


(c) Model 3

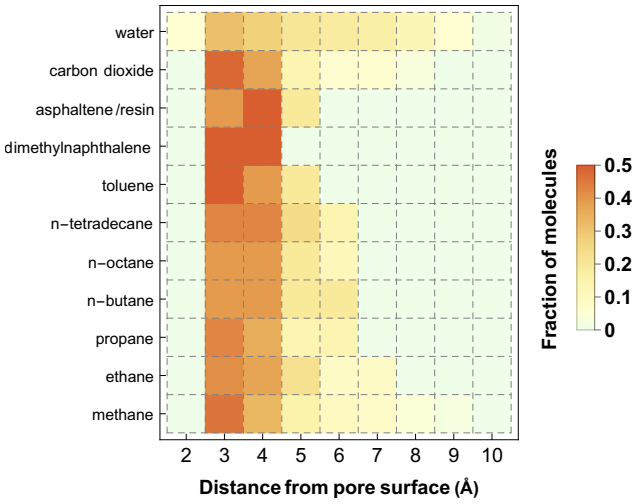
Figure 2.16: Histogram of the number of molecules as a function of the distance away from the pore surface.



(a) Model 1



(b) Model 2



(c) Model 3

Figure 2.17: Matrix plots of the fraction of each species as a function of the distance away from the pore surface.

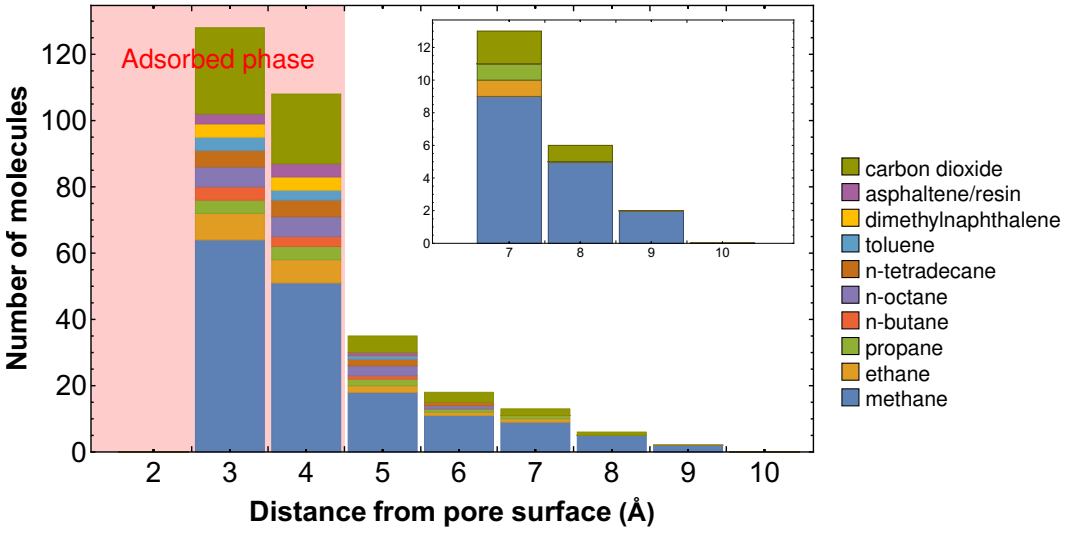
inside the pores.

The discussion above leads to the conclusion that the thickness of the adsorbed phase is 4 Å as highlighted by the light-red regions in Figure 2.18 in which water molecules are not depicted to exclude them from the adsorbed phase, and molecules beyond 5 Å are likely forming the free fluid phase. Thus, assuming an adsorbed phase thickness of 4 Å, the fraction of molecules in the free and adsorbed phases is summarized in Table 2.7 and illustrated in Figure 2.19. It can be said that of all the methane, ethane, propane, n-butane, and n-octane molecules in the pore space, approximately 25% of them are far away from the pore surface to be considered as the free fluid phase filling the pore while more than 80% of the heavier components are very close to the pore surface forming the adsorbed phase. Moreover, comparing the thickness of the adsorbed phase (4 Å) with the kinetic diameter of a methane molecule (3.8 Å) (Ismail et al., 2015), it seems that the adsorbed molecules form just one layer. All these numbers, however, should be used with caution as they might depend on factors such as the pore size and the force fields used to describe the atoms in the system.

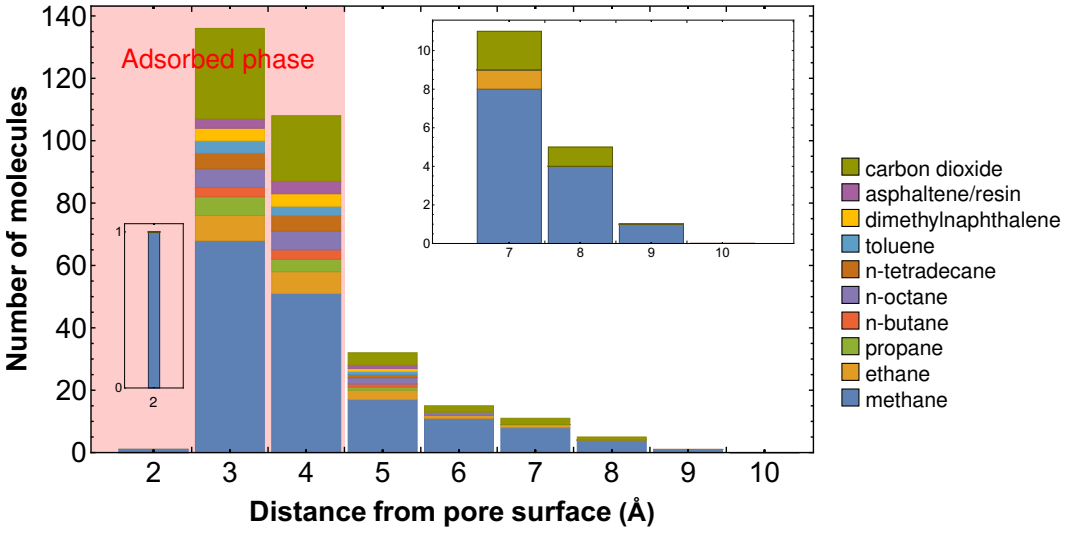
2.7 Final remarks

The fluids in organic pores at reservoir conditions segregate inside the pores, as the existence of the two phases free and adsorbed confirms. Depending on the molecular geometry of the species, the heavier longer chains and larger surface area molecules preferentially adsorb on the pore surface whereas the lighter and smaller ones account for most of the molecules that fill the pores, although a very large fraction of them also remains in the adsorbed phase. As the environment is hydrophobic, water molecules cluster together to form droplets that may touch the pore surface but wet it not. Carbon dioxide shows a higher tendency to adsorb than any hydrocarbon species and diffuse into the kerogen structure, which has important implications in enhanced oil recovery strategies.

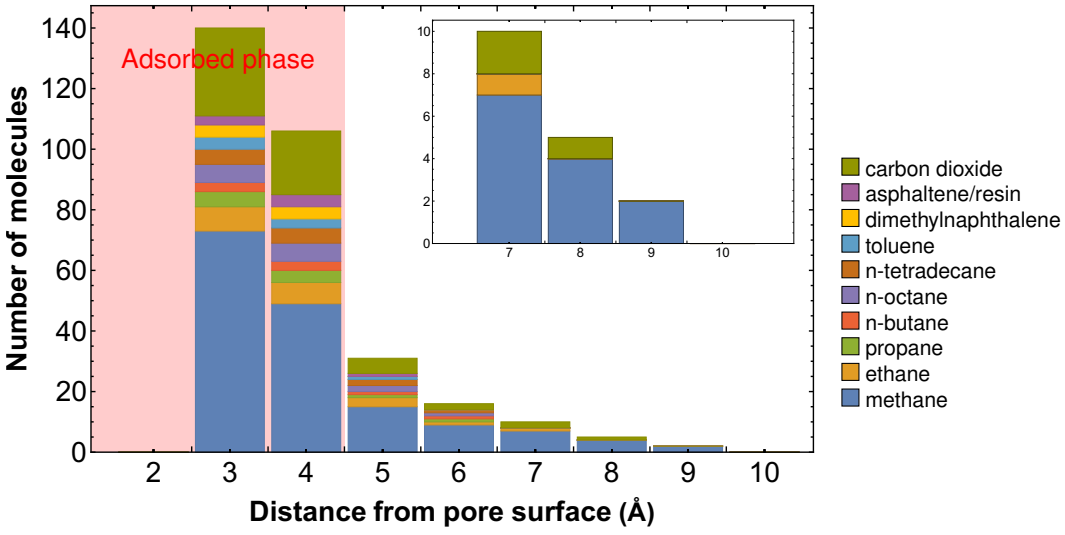
All the observations made in this chapter are static in the sense that the fluids are not subject to concentration or pressure gradients and therefore there is no net flow in



(a) Model 1



(b) Model 2



(c) Model 3

Figure 2.18: Estimation of the thickness of the adsorbed phase.

Table 2.7: Estimation of the amount of fluids in the free and adsorbed phases (assuming the thickness of the adsorbed phase is 4 Å.)

Species	Adsorbed (%)			Free (%)			Average	
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Adsorbed (%)	Free (%)
methane	72.48	74.26	76.59	27.52	25.74	23.41	74.44 ± 2.06	25.56 ± 2.06
ethane	74.24	76.42	74.14	25.76	23.58	25.86	74.93 ± 1.29	25.07 ± 1.29
propane	64.65	82.77	76.62	35.35	17.23	23.38	74.68 ± 9.22	25.32 ± 9.22
n-butane	81.50	82.83	66.01	18.51	17.17	33.99	76.78 ± 9.35	23.22 ± 9.35
n-octane	74.71	75.70	78.98	25.29	24.30	21.02	76.46 ± 2.23	23.54 ± 2.23
n-tetradecane	79.45	84.37	78.34	20.55	15.63	21.66	80.72 ± 3.21	19.28 ± 3.21
toluene	90.42	88.57	89.51	9.58	11.43	10.49	89.50 ± 0.92	10.50 ± 0.92
dimethylnaphthalene	93.83	91.99	89.88	6.17	8.01	10.12	91.90 ± 1.98	8.10 ± 1.98
asphaltene/resin	83.21	82.46	80.21	16.79	17.52	19.79	81.96 ± 1.56	18.04 ± 1.56
carbon dioxide	79.34	83.05	88.99	20.66	16.95	17.01	81.79 ± 2.13	18.21 ± 2.13

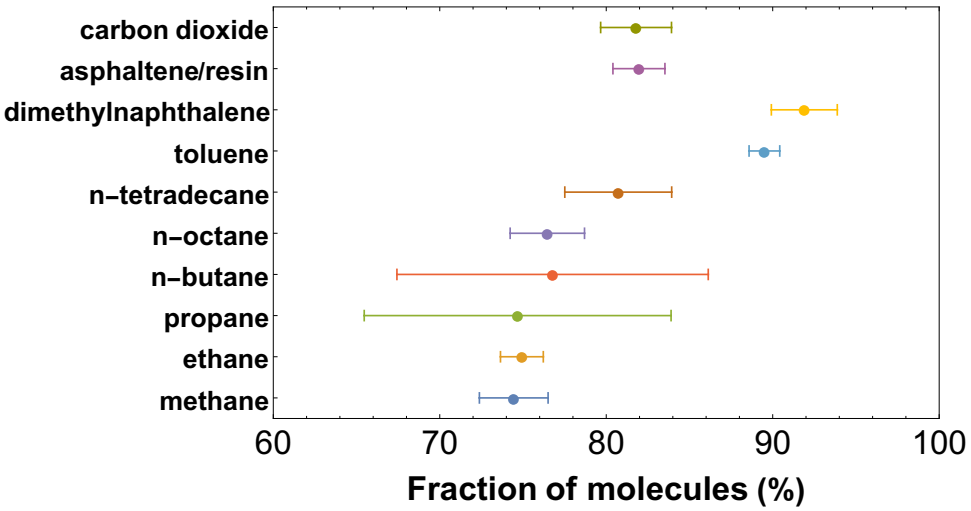


Figure 2.19: Estimation of the amount of fluids in the adsorbed phase (assuming the thickness of the adsorbed phase is 4 Å).

any spatial direction. The next chapters introduce such gradients to study the fluid flow mechanisms through organic nanopores. As will be seen in Section 3.6, the presence of a molecule in the adsorbed phase does not mean that such molecule remains static at that location. Phenomena such as motion along the pore surface and desorption may promote production of some of the heavier molecules and are discussed later.

Chapter 3

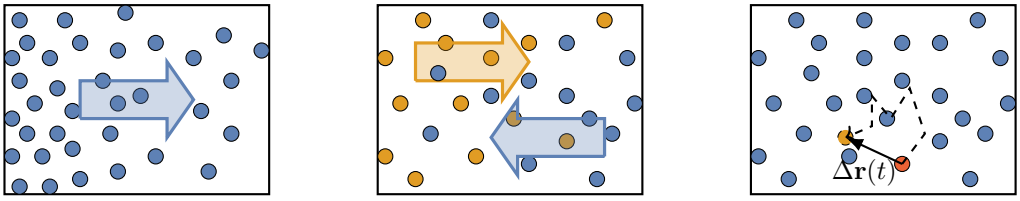
CONFINED FLUIDS IN ORGANIC NANOPORES. THE DYNAMIC PICTURE

3.1 Introduction

In Chapter 2, the spatial distribution of reservoir fluids was discussed. Temporal averages over one nanosecond in equilibrium conditions of the location of different molecular species were reported. It is important to note that the molecules do not remain in fixed positions but instead move randomly because of collisions with other molecules in a process called Brownian motion, even in the absence of external gradients that would cause a mass flux. Brownian molecular motion is also called tracer or self-diffusion and is described by the self-diffusion coefficient $\mathcal{D}$. This chapter concerns the estimation of the self-diffusion coefficients of hydrocarbon species in an oil mixture confined in organic nanopores.

3.2 Transport diffusion and self-diffusion

Diffusion is the process by which matter is transported from one part of the system to another as a result of random molecular collisions (Crank, 1975). Diffusion describes the tendency of matter to migrate in such a way as to eliminate spatial variations in



(a) (b) (c)

Figure 3.1: Microscopic picture of diffusion. (a) Mass transport due to a concentration gradient; (b) self-diffusion by following the flux of tagged particles (yellow balls) counterbalanced by that of untagged ones (blue balls); (c) self-diffusion by tracking the displacement $\Delta\mathbf{r}(t) = \mathbf{r}(t) - \mathbf{r}(0)$ of every particle (the yellow ball represents the final position of the example red particle).

composition and approach a uniform equilibrium state (Kärger et al., 2012). This is a universal property of matter at all temperatures above absolute zero and is a manifestation of the evolution of a system towards a state of maximum entropy to achieve equilibrium (see Appendix A).

The kinetic theory of gases developed in the nineteenth century (Present, 1958; Huang, 1987) provides a microscopic framework to study diffusion. Diffusion phenomena can be divided into two types (Kärger et al., 2012): (1) mass transfer (also called transport, collective or chemical diffusion) that results from a spatial concentration gradient, and (2) Brownian molecular motion which may be traced by either tagging a fraction of the diffusant particles (i.e. the atoms or molecules of a solute that diffuse in a solvent) or following the trajectories of a large number of individual diffusants and determining their mean square displacement (Haile, 1992)

$$\langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle = \int_0^t dt' \int_0^t dt'' \, \langle \dot{\mathbf{r}}(t') \cdot \dot{\mathbf{r}}(t'') \rangle \tag{3.1}$$

where the dot over a variable denotes a time derivative, thus one dot corresponds to d/dt. The different types of diffusion phenomena are illustrated in Figure 3.1. Conventionally, the term diffusivity is reserved for the situation depicted in Figure 3.1(a) in which there is a spatial concentration gradient that drives the mass flux as it is directly related to the macroscopic transport of matter.

The expression (3.1) is a time correlation function. In fact, information about microscopic properties of a system is usually given by correlation functions, which are of

great interest in computer simulations because their time integrals may often be related to macroscopic transport coefficients and their Fourier transforms may often be related to experimental spectra measured as a function of frequency ω (Zwanzig, 1965; McQuarrie, 1976; Thijssen, 2007; Allen and Tildesley, 2017). The discussion below illustrates the connection between time correlations and transport coefficients.

Digression on time correlation functions and transport coefficients

Time correlation functions measure how a dynamic quantity $A(t)$ may be related to some other quantity $B(t)$. Correlation functions can be computed directly from data generated by molecular dynamics simulations, which implies the numerical solution of the equations of motion (A.3) (see Section A.5).

Let $A(t)$ and $B(t)$ be two time-dependent quantities. The time correlation function $C_{AB}(t)$ is defined as

$$C_{AB}(t) = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt_0\, A(t_0)B(t_0+t). \tag{3.2}$$

The integral in (3.2) represents an average over many time origins t_0 with each origin taken from a system at equilibrium. The quantity A is sampled at time t_0 and B is sampled after a delay time t called lag time. Since the correlation function C_{AB} is an equilibrium property, it only depends on the lag time t and is independent of the time origin. In molecular simulations, the time average in (3.2) is replaced by ensemble averages *over initial state points* $\mathbf{\Gamma}(t_0)$ in phase space (see Section A.4),

$$C_{AB}(t) = \langle A(t_0)B(t_0+t)\rangle. \tag{3.3}$$

If A and B are uncorrelated, the correlation function C_{AB} reduces to the product of the individual averages,

$$C_{AB}(t) = \langle A\rangle\langle B\rangle, \tag{3.4}$$

but the converse is not true, that is values of C_{AB} different from the product $\langle A\rangle\langle B\rangle$ do not imply that A and B are related (Haile, 1992).

When A and B are different physical quantities, C_{AB} is called a cross-correlation

function, and when A and B are the same physical quantity, C_{AB} is called an auto-correlation function.

Time correlation functions are stationary, in other words $C_{AB}(t)$ is invariant under translations of the time origin,

$$C_{AB}(t) = \langle A(t_0)B(t_0+t)\rangle = \langle A(t_0+s)B(t_0+s+t)\rangle. \tag{3.5}$$

If $s = -t$ in (3.5), then

$$\langle A(t_0)B(t_0+t)\rangle = \langle A(t_0-t)B(t_0)\rangle. \tag{3.6}$$

In the limit of no lag time, $C_{AB}(0)$ defines a static correlation function

$$C_{AB}(0) = \lim_{\tau\to\infty}\int_0^\tau dt_0\, A(t_0)B(t_0) = \langle A(t_0)B(t_0)\rangle \equiv \langle AB\rangle. \tag{3.7}$$

In the opposite extreme, the long-time limit, the behavior of $C_{AB}(t)$ depends on the periodicity of quantities A and B. For non-periodic functions, A and B normally become uncorrelated after long lag times and

$$\lim_{t\to\infty} C_{AB}(t) = \langle A\rangle\langle B\rangle. \tag{3.8}$$

As mentioned above, correlation functions can be computed using data generated by molecular dynamics simulations. From an operational standpoint (see Appendix A), the computation of $C_{AB}(t)$ requires the selection of the initial points $\mathbf{\Gamma}(t_0)$ in phase space, according to the desired ensemble, over which the average will be subsequently performed. Next, the equations of motion (A.3) are solved to generate the trajectories $\mathbf{\Gamma}(t)$ in phase space. Thus, time-dependent properties may be calculated in any ensemble.

The link between correlation functions and transport coefficients is made through linear response theory. Transfer of mass, energy, or momentum through a system is

described, to first order, by a phenomenological relation of the form

$$\text{flux} = -\text{coefficient} \times \text{gradient}. \tag{3.9}$$

As written, (3.9) conveys the idea that the gradient provides the driving force for a flux that measures the transfer per unit area in unit time, and the coefficient characterizes the resistance to flow (de Groot and Mazur, 1984). The negative sign in (3.9) means that the flux opposes the direction in which the gradient increases. Thus, transport coefficients describe the relaxation of dynamical variables on the macroscopic scale. If the long-time and large-length-scale limits are considered carefully, transport coefficients can be expressed in terms of equilibrium time correlation functions, called Green-Kubo relations, of fluctuating microscopic quantities (Haile, 1992; Allen and Tildesley, 2017). The fluctuations depend on the ensemble utilized, hence the time correlation function of a particular quantity calculated in one ensemble may correspond to the time correlation function of a somewhat different quantity in another ensemble. However, in the thermodynamic limit of sufficiently large systems discussed in Section A.4.2, the ensemble equivalence yields different expressions for transport coefficients that give identical numerical values (Zwanzig, 1965).

The Green-Kubo relation is often written as the infinite time integral of an equilibrium time correlation function of the form (Allen and Tildesley, 2017)

$$\gamma = \int_0^\infty dt \, \langle \dot{A}(t) \dot{A}(0) \rangle \tag{3.10}$$

where γ is the transport coefficient associated with the dynamical variable A.

Linear response theory provides an interpretation of the Green-Kubo relations in terms of the response of the system to a weak perturbation. Just as done in perturbation theory (Landau and Lifshitz, 1977; Byron, Jr. and Fuller, 1992), the perturbation is introduced into the Hamiltonian $\mathcal{H}$ of the system (or directly into the equations of motion (A.3)), producing a time-dependent nonequilibrium ensemble, and thus a nonequilibrium ensemble average may be calculated. Retaining the linear terms in the perturbation

and comparing the equation for the response with a macroscopic transport equation, the transport coefficients may be identified.

Intimately related to any expression of the form shown in (3.10) there is a relation (McQuarrie, 1976; Haile, 1992; Allen and Tildesley, 2017)

$$\lim_{t\to\infty}\frac{d}{dt}\frac{1}{2}\langle[A(t)-A(0)]^2\rangle=\lim_{t\to\infty}\frac{1}{2t}\langle[A(t)-A(0)]^2\rangle=\gamma \tag{3.11}$$

called Einstein relation which is valid for times t larger than the correlation time of A. The three-dimensional analog of (3.11) is

$$\lim_{t\to\infty}\frac{d}{dt}\frac{1}{6}\langle[\mathbf{A}(t)-\mathbf{A}(0)]^2\rangle=\lim_{t\to\infty}\frac{1}{6t}\langle[\mathbf{A}(t)-\mathbf{A}(0)]^2\rangle=\gamma \tag{3.12}$$

where $\mathbf{A}$ is a vectorial quantity.

The significance of Einstein relations is that they relate, in simple ways, macroscopic transport coefficients to molecular quantities (Haile, 1992). ■

3.2.1 Transport diffusion

In light of expression (3.9), a spatial concentration gradient will cause a mass flux from one region to another until the diffusant is uniformly distributed across space. Adolf Fick formulated this observation and proposed what is generally known as Fick's first law of diffusion,

$$\mathbf{J}(\mathbf{r},t)=-D\nabla c(\mathbf{r},t) \tag{3.13}$$

where $\mathbf{J}(\mathbf{r},t)$ is the diffusive flux, $c(\mathbf{r},t)$ is the local concentration of the diffusant, and D is the diffusivity. Expression (3.13) is said to define the diffusivity D. The Fick's second law of diffusion can be derived from the Fick's first law (3.13) using conservation of mass considerations to obtain (Crank, 1975)

$$\frac{\partial}{\partial t}c(\mathbf{r},t)=\nabla\cdot[D(c)\nabla c(\mathbf{r},t)] \tag{3.14}$$

which is the most general form because it allows for concentration dependence of the diffusivity. If D is constant, (3.14) becomes

$$\frac{\partial}{\partial t}c(\mathbf{r},t) = D\nabla^2 c(\mathbf{r},t) \tag{3.15}$$

where ∇^2 is the Laplacian differential operator which in Cartesian coordinates takes on the form

$$\nabla^2 \equiv \sum_{\alpha=1}^{3} \frac{\partial^2}{\partial x_\alpha^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3.16}$$

3.2.2 Self-diffusivity

The quantity that describes the rate of Brownian migration under conditions of macroscopic equilibrium depicted in Figures 3.1(b) and 3.1(c) is referred to as the tracer or self-diffusivity $\mathcal{D}$. The formal definition of the self-diffusivity is given by either (Kärger et al., 2012)

$$\mathbf{J}^*(\mathbf{r},t) = -\mathcal{D}\ \nabla c^*(\mathbf{r},t)|_{c\,=\,\text{constant}} \tag{3.17}$$

where $\mathbf{J}^*(\mathbf{r},t)$ is the diffusive flux of tagged particles and $c^*(\mathbf{r},t)$ is in turn the local concentration of tagged particles, or (Haile, 1992; Allen and Tildesley, 2017)

$$\mathcal{D} = \lim_{t\to\infty} \frac{d}{dt}\frac{1}{6}\langle[\mathbf{r}(t)-\mathbf{r}(0)]^2\rangle = \lim_{t\to\infty}\frac{1}{6t}\langle[\mathbf{r}(t)-\mathbf{r}(0)]^2\rangle, \tag{3.18}$$

where expressions (3.1), (3.10), and (3.12) were used. Notice that in the long-time limit[1] the mean square displacement depends linearly on elapsed time t,

$$\langle[\mathbf{r}(t)-\mathbf{r}(0)]^2\rangle \underset{t\to\infty}{=} 6\mathcal{D}t. \tag{3.19}$$

The definitions (3.17) and (3.18) are equivalent (Kärger et al., 2012). Expression (3.17) shows that the self-diffusivity may vary with the total concentration c but does

[1]This is the reason why it is said in molecular dynamics simulations that simulation time must reach the long-time limit before self-diffusivity may be estimated from mean square displacements. The long-time range ultimately depends on the particular system under consideration.

not vary with the fraction of labeled particles.

Although both diffusion phenomena described above occur by essentially the same microscopic mechanism, the coefficients that describe transport diffusion and self-diffusion, D and $\mathcal{D}$, respectively, are generally not the same (Kärger et al., 2012).

3.3 Simulation details

Two new models of mature kerogen, larger than those built in Chapter 2 and saturated with only a hydrocarbon mixture (no water and carbon dioxide molecules are included in the fluid system), are created using the kerogen type II-C monomer designed by Ungerer et al. (2015) illustrated in Figure 2.1(a). The process followed is similar to that described in Section 2.3, starting with the random placement of 80 kerogen monomers in a $40 \times 40 \times 40$ nm^3 box along with fluid molecules that make up the oil mixture (one corresponds to a black oil and the other to a volatile oil) using PACKMOL (Martínez et al., 2009). Table 3.1 presents the number of molecules of each species in the systems. The force fields utilized to describe the atoms in kerogen and hydrocarbon molecules are cvff (Dauber-Osguthorpe et al., 1988) and OPLS-AA (Jorgensen et al., 1996) with the refinement to the torsional parameters described in Price et al. (2001), respectively, and are summarized in Appendix B. The systems undergo the annealing simulation (Kirkpatrick et al., 1983) carried out in LAMMPS (Plimpton, 1995) described in Table 3.2. Periodic boundary conditions are used in all directions. A cutoff distance of 15 Å is selected in evaluating the non-bonded interactions Lennard-Jones and Coulombic, long distance corrections are used to estimate Lennard-Jones energy beyond the cutoff (Sun, 1998) and the particle-particle particle-mesh method is used for electrostatic interactions (Hockney and Eastwood, 1988). Temperature and pressure are controlled by the Nosé–Hoover thermostat (Nosé, 1984a,b, 1991) and the Parrinello–Rahman barostat (Parrinello and Rahman, 1981), respectively.

After the systems have reached equilibrium, the kerogen model saturated with a black oil undergoes a simulation in the NVT ensemble (constant number of particles

Table 3.1: Number of molecules used to construct the models of mature kerogen saturated with oil mixtures.

Species	Number of molecules	
	Black oil	Volatile oil
kerogen	80	80
methane	800	800
ethane	117	100
propane	85	61
n-butane	67	41
n-octane	180	80
n-tetradecane	150	61
toluene	90	41
dimethylnaphthalene	104	41
asphaltene/resin	106	41

Table 3.2: Simulated annealing campaign to create the saturated kerogen models. NVT means constant number of particles N, volume V, and temperature T, NPT means constant number of particles N, pressure P, and temperature T (see Appendix A).

Ensemble	T (K)	P (atm)	Δt (fs)	t (ns)
NVT	900		1.0	0.1
NPT	900	10	0.5	0.5
NPT	$900 \rightarrow 700$	10	0.5	0.1
NPT	700	10	0.5	0.4
NPT	$700 \rightarrow 500$	10	0.5	0.1
NPT	500	10	0.5	0.6
NPT	$500 \rightarrow 355$	10	0.5	0.1
NPT	355	10	0.5	0.5
NPT	355	300	0.5	0.5
NVT	$355 \rightarrow 2000$		0.5	0.025
NVT	$2000 \rightarrow 355$		0.5	0.5
NPT	355	300	0.5	1.0
NVT	355		0.5	0.5

N, volume V, and temperature T, see Appendix A) for 20 ns using a time step of 1 fs in which all particles in the system are allowed to move with the aim to collect data to estimate the self-diffusion coefficient of the species for which the long-time limit is achieved. In this long simulation, the temperature of kerogen and the fluid molecules is controlled separately implementing two thermostats by explicitly rescaling the velocities of the particles in each subsystem, as discussed in Section A.5.2. The rescaling is applied to only the translational degrees of freedom of the particles. The reasons for choosing this model over that saturated with a volatile oil are that the black oil model contains more fluid molecules that contribute to the ensemble average (hopefully reducing the statistical error) and the computational resources are limited[2]. However, both models are used in the posterior studies described in the next chapters.

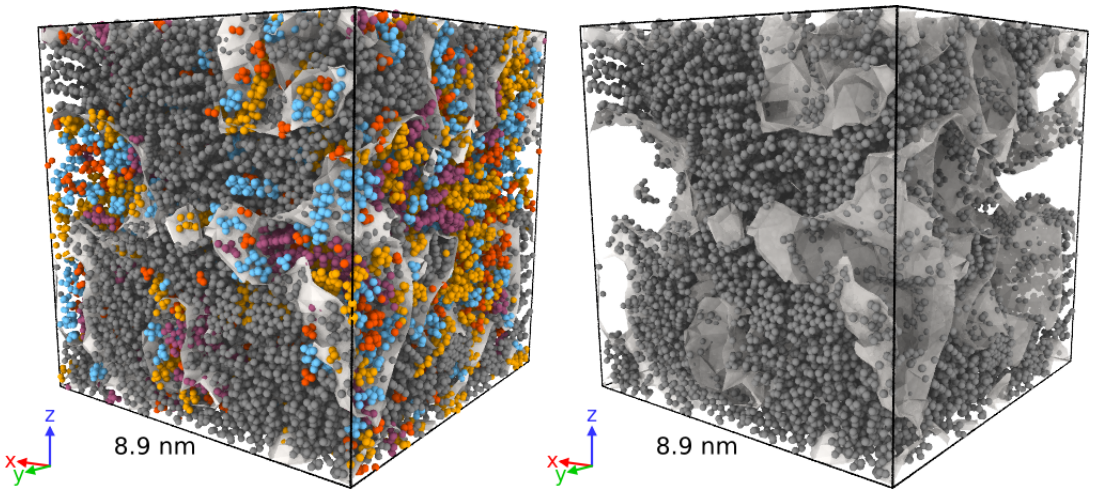
3.4 Saturated mature kerogen models

The models of mature kerogen saturated with oil at the reservoir conditions $T = 355$ K (180°F) and $P = 300$ atm (4400 psi) obtained after the long annealing simulation are shown in Figure 3.2. The systems contain five as many molecules of methane and kerogen monomers as the models created in Chapter 2 to consider larger pores connected through pore throats. According to Holstein (2007), oils with a molecular weight in the range 40 to 70 g/mole and C_{7+} mole fraction between 10 and 30% correspond to volatile oils, while molecular weights in the range 70 to 210 g/mole and C_{7+} mole fraction between 35 and 50% correspond to black oils. Thus, the lumped oil models used in this work represent a black oil and a volatile oil, respectively, as shown in Table 3.3.

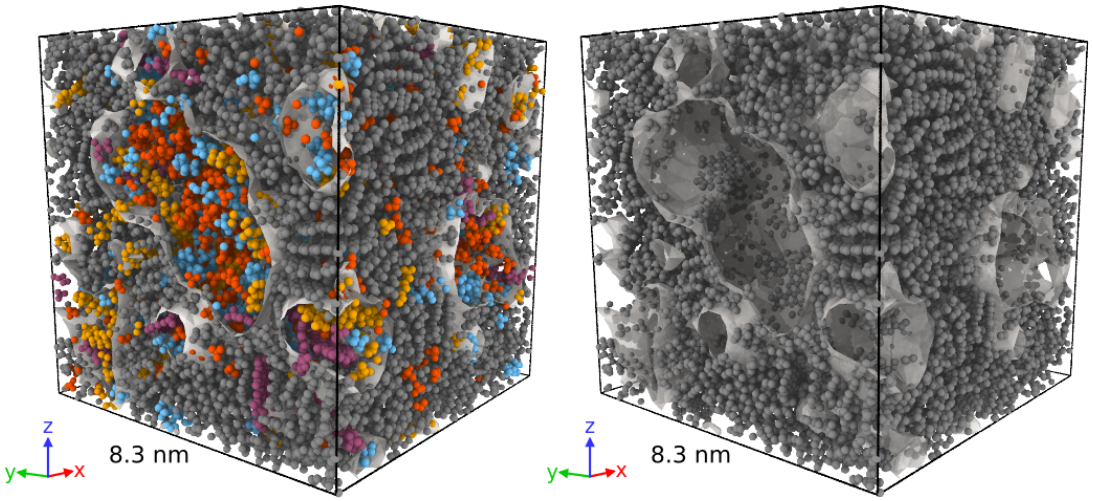
3.4.1 Pore surface, porosity and density

To determine the characteristics of the kerogen structures, the fluids are removed from the systems as illustrated in Figure 3.2 which shows the kerogen frameworks and

[2]These simulations demand long wall times on OSCER, the supercomputer system at the University of Oklahoma used in this work. Roughly, the simulations take a wall time of 35 days, to which the queue times must be also added, giving an effective time of about 6 to 7 weeks.



(a) Saturating fluid: Black oil



(b) Saturating fluid: Volatile oil

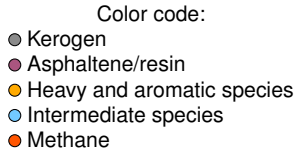


Figure 3.2: Molecular models of mature kerogen saturated with oil. The figures on the left show both kerogen and fluids; in the figures on the right the fluid atoms are removed to visualize the pore geometry (transparent gray surface) and the kerogen framework.

Table 3.3: Composition of the oils that saturate kerogen.

Species	Mole fraction (%)	
	Black oil	Volatile oil
methane	47.1	63.2
ethane	6.9	7.9
propane	5.0	4.8
n-butane	3.9	3.2
n-octane	10.6	6.3
n-tetradecane	8.8	4.8
toluene	5.3	3.2
dimethylnaphthalene	6.1	3.2
asphaltene/resin	6.2	3.2
Total molecular weight (g/mole)	80.68	53.02

Table 3.4: Physical properties of the models of mature kerogen.

Model	Surface area (m^2/g)	Solid volume fraction	Porosity (%)	Density (g/cm^3)
Black oil	1699.40	0.47	53.02	1.41
Volatile oil	1305.11	0.61	38.62	1.30

the pore space. Monte Carlo simulations are used to compute the pore size distributions illustrated in Figure 3.3 using a helium probe and the code PSDsolve-1.01 (Bhattacharya and Gubbins, 2006). The peak at around 5 Å corresponds to the distance between the monomers that make up the kerogen structure, and the largest pore enclosed by the model saturated with a black oil is 25 Å. Using the Python interpreter OVITOS that comes with OVITO distributions (Stukowski, 2010), the average pore surface and solid volume fractions are computed by constructing a polyhedral mesh around the kerogen atoms, generating a geometric representation of the pore surface (Stukowski, 2014), as depicted in Figure 3.2. The radius of the probe sphere is chosen so that it represents a helium probe, and 400 simulation frames are utilized to compute the average values, which corresponds to averages over 4 ns when the systems have reached equilibrium. After subtracting the pore volume as discussed in Section 2.4.1, the kerogen densities are computed using expression (2.2). The values are summarized in Table 3.4. Comparatively, the average density of the kerogen models created in Chapter 2 is 1.32 ± 0.02 g/cm^3 at same pressure and temperature.

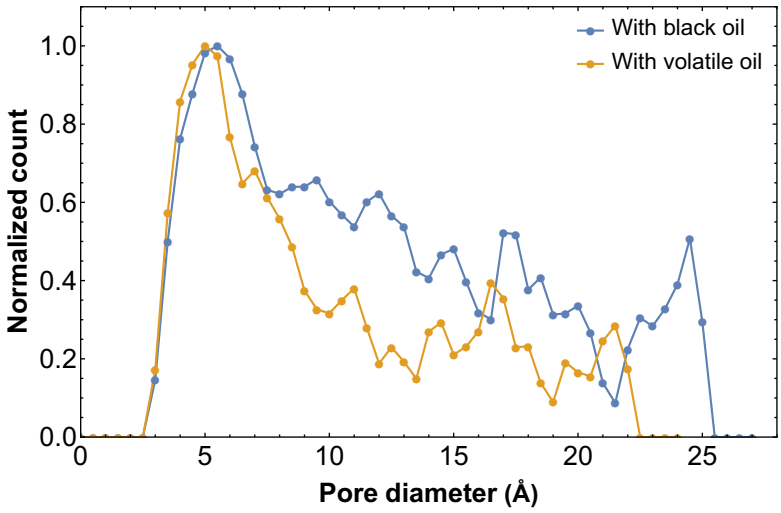


Figure 3.3: Pore size distributions of the mature kerogen models. The colors denote different oil mixtures filling the organic pores.

3.5 Mean square displacement from molecular dynamics simulations

There are two common ways to obtain the self-diffusivity $\mathcal{D}$ from molecular dynamics simulations. The Green-Kubo relation (3.10) and Einstein relation (3.12) for a diffusive flux in the long-time limit $t \to \infty$ yield

$$\mathcal{D} = \frac{1}{6t}\langle[\mathbf{r}(t) - \mathbf{r}(0)]^2\rangle = \int_0^\infty d\tau \, \langle \dot{\mathbf{r}}(\tau) \cdot \dot{\mathbf{r}}(0)\rangle, \tag{3.20}$$

which shows that one could use either molecule positions or velocities. Both methods yield the same result, but the use of molecule positions requires an ensemble average whereas the use of molecule velocities involves integrating the velocity auto-correlation function. There is, however, a long-time tail associated with the latter integral that may cause numerical problems (Haile, 1992; Tuckerman, 2010). The displacement of atom i through the system creates a vortex motion in neighboring atoms that affect the motion of atom i itself (Haile, 1992). Therefore, longer simulations are needed to obtain reliable self-diffusivity from velocities than from positions.

The mean square displacement, on the other hand, requires an ensemble average which involves averaging over all time origins and all molecules in the simulations. In

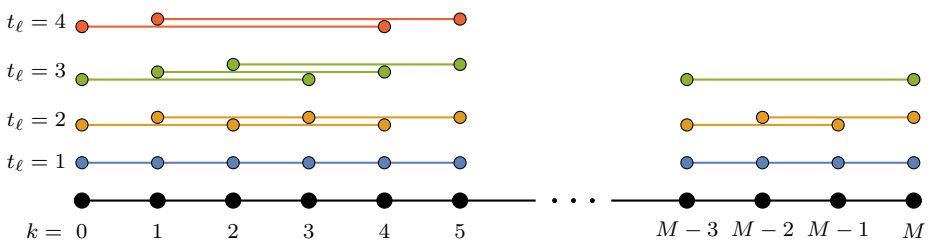


Figure 3.4: Schematic of the information stored in the trajectory of a molecule. The black dots on the bottom represent the time steps at which the coordinates are recorded. The blue dots mark lag times of one time step, the yellow dots lag times of two time steps, the green dots lag times of three time steps, and the red dots lag times of four time steps.

an equilibrium molecular dynamics simulation, any time step can be considered the time zero in expression (3.20) since the variable t refers to observation time (i.e. lag time) rather than some absolute arbitrary time.

The first step to compute the mean square displacement of each species in the black oil mixture is to determine the unwrapped[3] trajectory of the center of mass of each fluid molecule[4]. Even if periodic boundary conditions are used in the simulation, the position of every particle in the system can be recorded as wrapped or unwrapped coordinates.

There are two calculations involved in the mean square displacement. Focus first on molecule i of species j in the system. Assume there are N_j molecules of species j. The trajectory of molecule i is given by the set of coordinates $\{\mathbf{r}_{ij}(k\Delta t)\}$, where the subindex i refers to the molecule ($i = 1, 2, \ldots, N_j$) and the subindex j to the molecule type ($j = 1, 2, \ldots, N_s$), N_s is the number of species, k is the number of the time step ($k = 0, 1, \ldots, M$), $M + 1$ is the total number of time steps, and Δt is the time step at which the simulation was run. A schematic of this information is presented in Figure 3.4.

Any point in an equilibrium simulation can be considered a time origin. It can be seen in Figure 3.4 that there are M time origins for a lag time $t_\ell = 1$ time step, $M - 1$ time origins for $t_\ell = 2$ time steps, $M - 2$ time origins $t_\ell = 3$ time steps, and so on, and only one time origin for a lag time $t_\ell = M$ time steps. However, it is desirable to have the same number of samples per lag time so that each is represented equally in

[3]Since periodic boundary conditions are normally applied in all directions, it is necessary to determine first the actual trajectory in space followed by the center of mass of each molecule to compute the displacement after some elapsed time t. The use of periodic boundary conditions wraps the trajectories back to the primary simulation cell as discussed in Section A.2.

[4]It will be understood hereafter that the position of molecule i refers to the position of its center of mass unless stated otherwise.

the computation of the mean square displacement as a function of elapsed time. Codes like LAMMPS have built in functions to calculate the mean square displacement and the velocity auto-correlation function. Nonetheless, when utilized, neither the mean square displacement nor the velocity auto-correlation function are ensemble averages because they only use the starting point of the simulation as the time origin, falling in the issue exposed above.

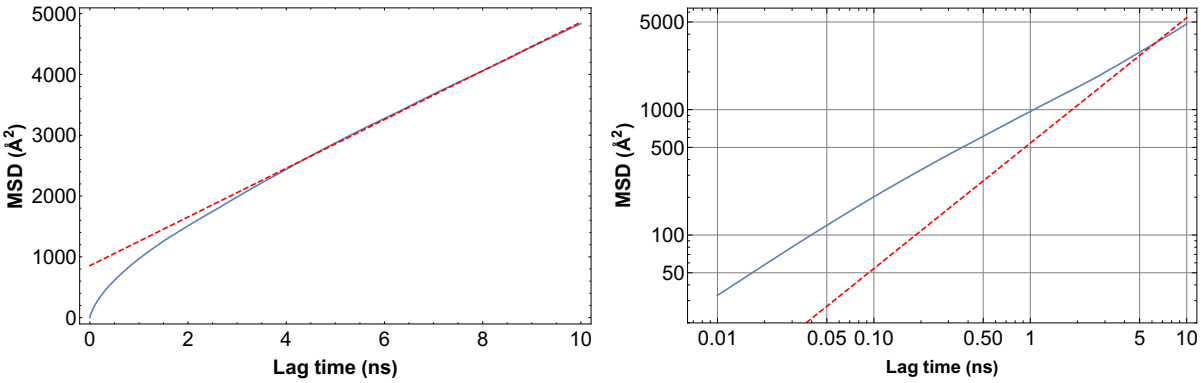
The solution is to evaluate the mean square displacement for lag times t_ℓ from 1 to $N_{\max} = \text{int}(M/2)$ (Keffer, 2002). Only the first $\text{int}(M/2)$ points are considered as time origins. Constraining $t_\ell \leq N_{\max}$ gives the same number of data points contributing to every elapsed time t_ℓ, and therefore they are weighted equally. The drawback, however, is that very long simulations might be needed to reach the long-time limit in order to estimate the self-diffusion coefficient $\mathcal{D}$ from the mean square displacement, since now only half of the data will be used as lag times.

Once the previous calculation is done for one molecule i of species j, it is repeated for the other $N_j - 1$ molecules of the same species, then the results are added together an divided by N_j to improve statistical accuracy.

The considerations just discussed are implemented in Python to be used with the interpreter OVITOS that comes with OVITO distributions (Stukowski, 2010). The mean square displacement of every molecular species is computed and the results are shown next.

3.6 Estimation of self-diffusion coefficients

According to expression (3.19), the mean square displacement (MSD) $\langle[\mathbf{r}(t)-\mathbf{r}(0)]^2\rangle$ is a linear function of the observation time in the long-time limit, intercepting the vertical axis at the origin of the plane and having a slope equal to $6\mathcal{D}$. The mean square displacement of every molecular species is computed in the range of lag times $0 \leq t_\ell \leq t/2$, where $t = 20$ ns is the total duration of the simulation. Figure 3.5(a) is an example of the MSD of methane confined in organic pores. Two regions can be identified in the



(a) (b)

Figure 3.5: Mean square displacement of methane. (a) Linear scale; (b) log-log scale. The grid lines are included as a visual help.

figure: (1) an early non-linear relation that shows the free molecular motion and the start of collisions, and (2) a later linear relation. The latter will be used to estimate the self-diffusion coefficient of methane, considering the portion of data points that have reached the long-time limit. However, determining such a portion is not straightforward from the linear plot in Figure 3.5(a), thus it is also plotted in log-log scale in Figure 3.5(b) in which the long-time limit should show a linear trend with slope one. Although Figure 3.5(b) is fairly straight, the slope is clearly less than one for elapsed times below 5 ns which is indicative of sub-diffusive behavior that is observed due to limited lag times. In fact, it is known that confinement significantly restricts diffusion (Calvo-Muñoz et al., 2011; Kärger et al., 2012). Longer simulation times would eventually capture the diffusive behavior of the fluid.

To determine the time $t_{\ell,\min}$ when the long-time limit begins on the logarithmic scale plot, a series of slope-one linear regression fits to the data is performed. Points at an early lag time are eliminated until the coefficient of determination R^2 crosses a threshold value of 0.9999. Then the best linear fit $\langle[\mathbf{r}(t_\ell) - \mathbf{r}(0)]^2\rangle = 6\mathcal{D}t_\ell + b$ for $t_\ell \geq t_{\ell,\min}$ is determined, from which the self-diffusivity is obtained. The intercept with the vertical axis b is not zero but it has no physical significance (Keffer, 2002). If b were not included, it would change the value of the slope.

In the case of methane, it is determined that $t_{\ell,\min} = 3.28$ ns. The dashed red line in Figure 3.5(b) corresponds to the best linear fit with slope one (in log-log scale), and

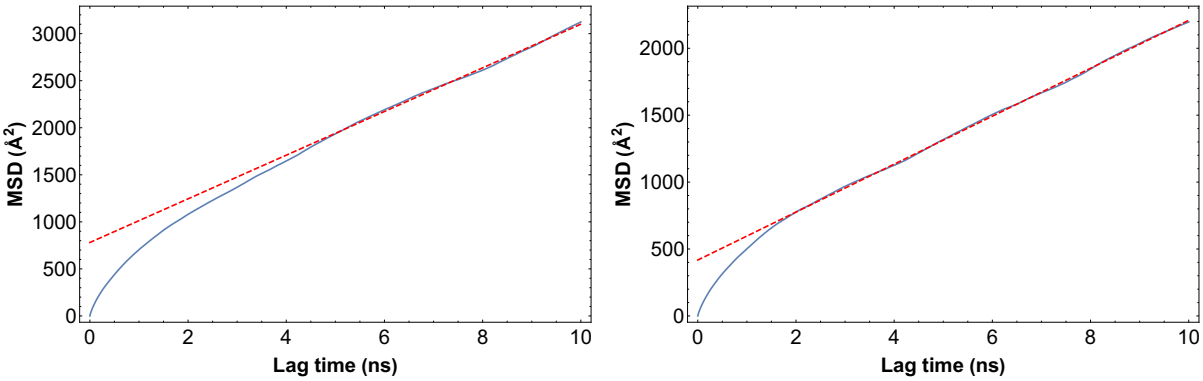
Table 3.5: Self-diffusion coefficients of the fluid species in the black oil.

Species	$t_{\ell,\min}$ (ns)	$\mathcal{D}$ ($\times 10^{-10}$ m^2/s)	$\mathcal{D}_{\text{methane}}/\mathcal{D}$
methane	3.28	6.677 ± 0.006	
ethane	4.39	3.867 ± 0.007	1.7265 ± 0.0013
propane	3.80	2.9858 ± 0.0031	2.2362 ± 0.0030
n-butane	4.51	2.789 ± 0.009	2.394 ± 0.008
n-octane	3.76	2.1463 ± 0.0014	3.1108 ± 0.0034
n-tetradecane	5.29	0.8927 ± 0.0011	7.479 ± 0.011
toluene	3.62	1.7402 ± 0.0021	3.837 ± 0.006
dimethylnaphthalene	5.37	0.6536 ± 0.0025	10.22 ± 0.04
asphaltene/resin	4.97	0.05962 ± 0.00010	111.98 ± 0.22

the corresponding best linear fit using the data in the long-time limit is the dashed red line in Figure 3.5(a). The self-diffusion coefficient of methane, estimated from the slope, is $\mathcal{D}_{\text{methane}} = (6.677 \pm 0.006) \times 10^{-10}$ m^2/s.

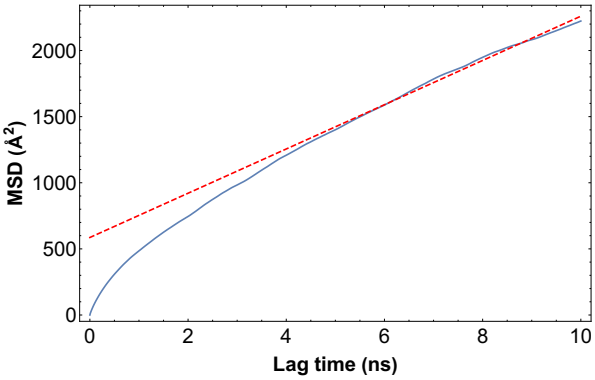
A similar analysis is carried out for all other fluid species in the black oil. The linear plots with the best linear fits are shown in Figure 3.6, the log-log plots are presented in Figure 3.7, and the results are summarized in Table 3.5. It should be pointed out that the later the onset of the long-time domain, the fewer data points are utilized to find the best linear fit to estimate the self-diffusion coefficients, and the less reliable such estimation will be. That is the case with the bigger molecules n-tetradecane, dimethylnaphthalene and asphaltene/resin fraction. Disregarding the specific values of $\mathcal{D}$ given in Table 3.5, what is significant is that the order of magnitude of the self-diffusion coefficients of the confined fluid species is on the order of 10^{-10} m^2/s. Moreover, as the molecular weight of the linear alkanes increases, the self-diffusion coefficient decreases.

Inspired by the Graham's law of diffusion, which states that the diffusion or spontaneous mixing of two gases in contact is, in the case of each gas, inversely proportional to the density of the gas (Kärger et al., 2012), the self-diffusion coefficient of each species is plotted as a function of the inverse of the square root of its molecular weight in Figure 3.8. The linear fit obtained in this work shows that there is in fact a dependency of the diffusivity on the molecular weight of the species even under confined conditions. The positive slope indicates that the self-diffusion coefficient decreases as the molecular weight of the hydrocarbons increase.

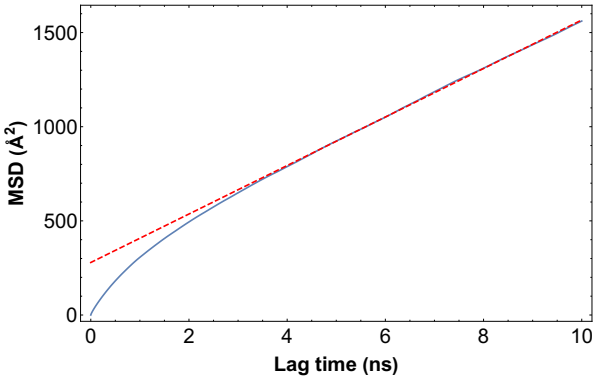


(a) Ethane

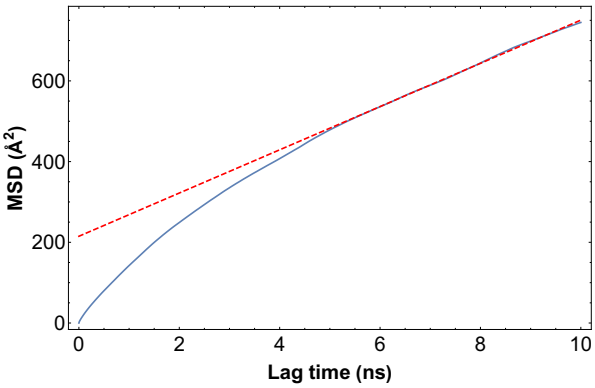
(b) Propane



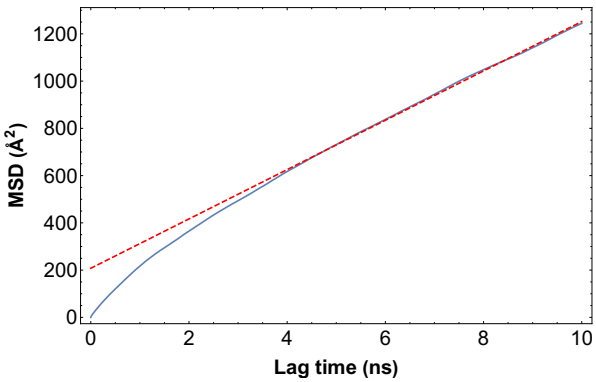
(c) n-butane



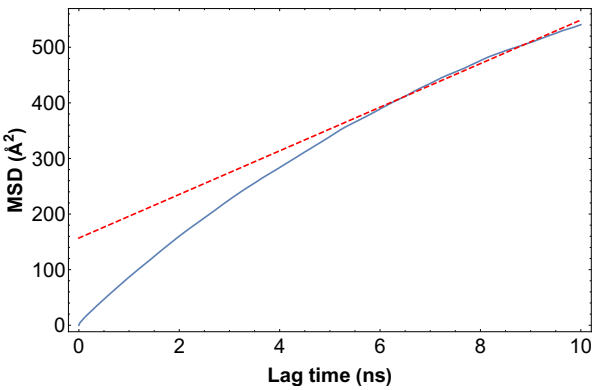
(d) n-octane



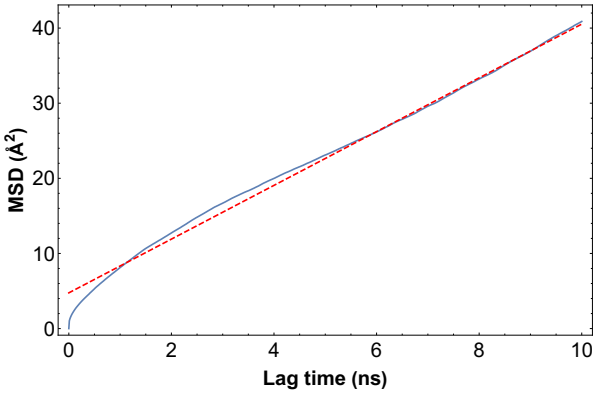
(e) n-tetradecane



(f) Toluene

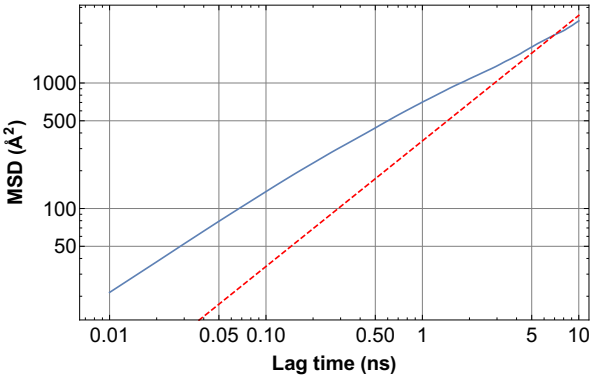


(g) Dimethylnaphthalene

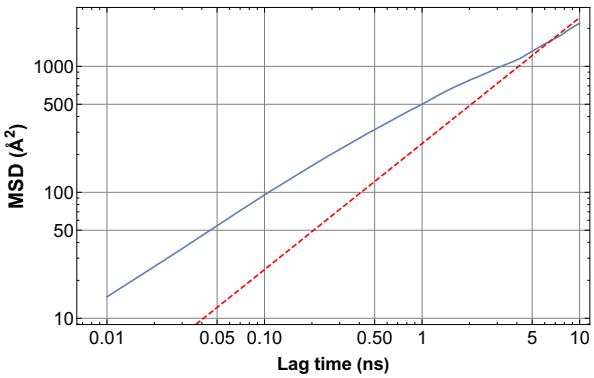


(h) Asphaltene/resin

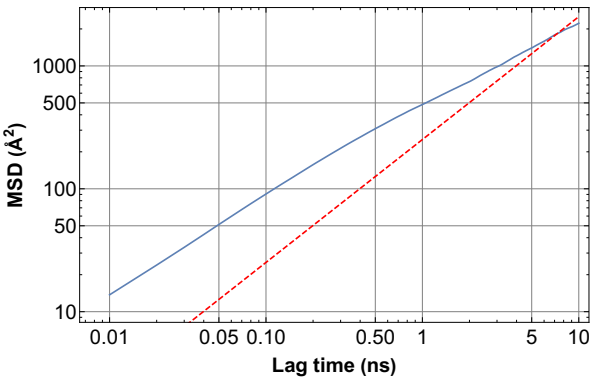
Figure 3.6: Linear plots of the mean square displacement of fluid species other than methane.



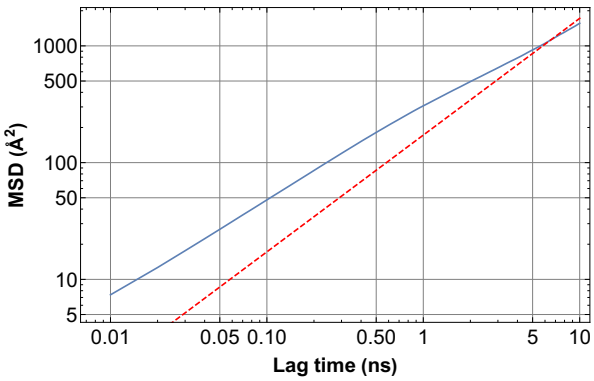
(a) Ethane



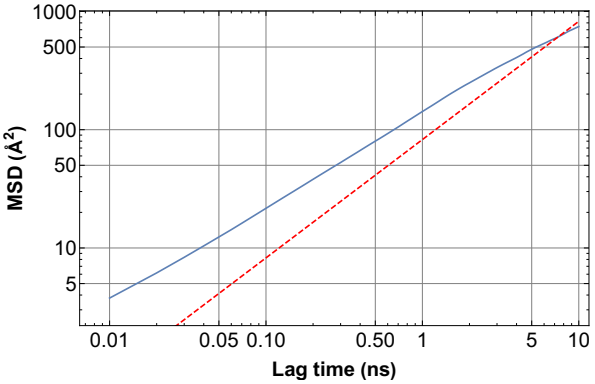
(b) Propane



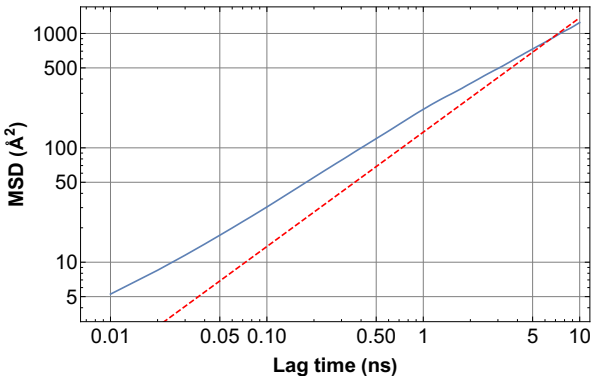
(c) n-butane



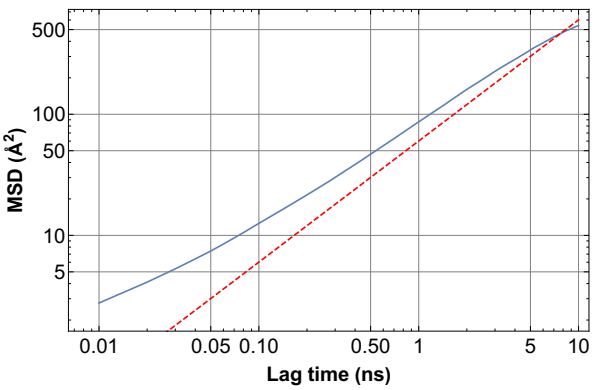
(d) n-octane



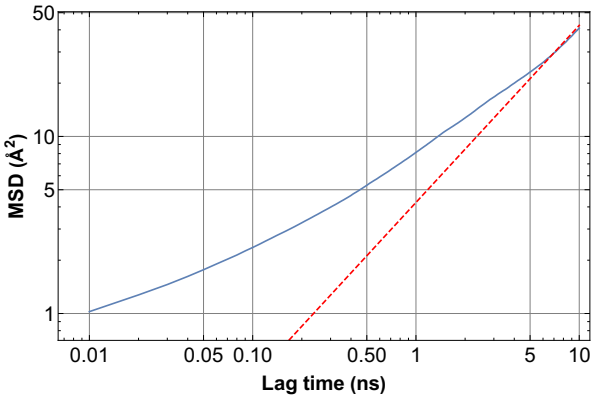
(e) n-tetradecane



(f) Toluene



(g) Dimethylnaphthalene



(h) Asphaltene/resin

Figure 3.7: Log-log plots of the mean square displacement of fluid species other than methane. The grid lines are included as a visual help.

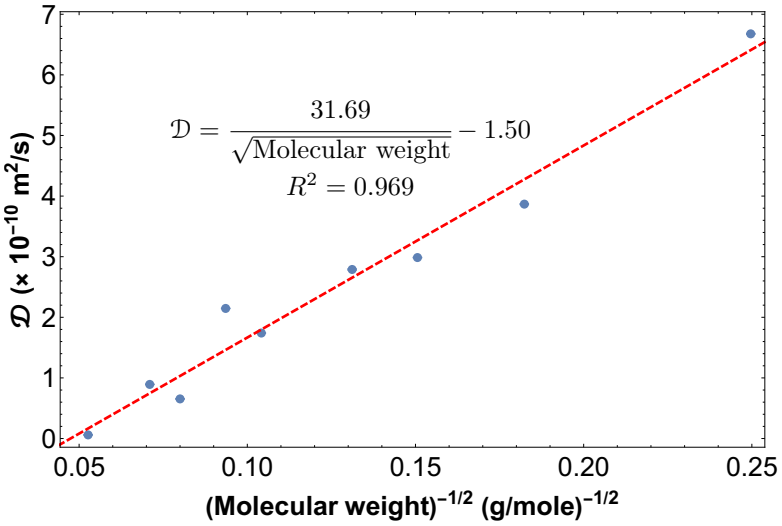


Figure 3.8: Self-diffusion coefficient of each species as a function of the inverse of the square root of its molecular weight. The red-dashed line is the best linear fit described by the given equation.

If self-diffusion measures the mobility of a certain species under thermodynamic equilibrium conditions, it is just expected that methane, the smallest hydrocarbon component in the fluid mixture, presents the highest mobility, and as the molecular weight, chain length, molecule size, or molecule geometry complexity increases, the mobility decreases, as clearly seen in Figure 3.8. Furthermore, molecular surface area, adsorption effects, and pore and throat geometries are also parameters that govern the mobility of molecular species. The very low mobility of the asphaltene/resin fraction in the black oil is a good example of the importance of these factors.

3.7 Final remarks

Diffusion of molecules through the system is the result of random molecular collisions. Matter is transported from one part of the system to another in such a way as to eliminate spatial variations in composition and approach a uniform equilibrium state. As mentioned in Sections 2.5 and 2.6, the different compounds that make up an oil mixture are segregated within organic pores, creating local concentration gradients that might promote molecular diffusion to produce a uniform mixture. However, the enhanced interactions between pore walls and fluid molecules compete with the tendency of matter

to eliminate spatial concentration gradients. What is seen using molecular dynamics simulations is a state of equilibrium between those competitive processes.

The situations discussed in Chapters 2 and 3 assume no external disturbances to the system under study, thus they describe the fluid behavior of confined fluids in organic nanopores, in a manner of speaking. It was mentioned in Chapter 2 that, when averaged over time, the close proximity of a molecule to the pore surface does not mean that such molecule remains still at a fixed position; the diverse self-diffusion coefficients among the fluid species demonstrate that the molecules do move within organic pores at different extents as a result of their differential interaction with the pore walls. Thus, adsorption of hydrocarbon molecules to the organic pore surface tends to increase the local density of the fluid (as compared to the same amount of fluid at same temperature and pressure but under unconfined conditions, which occupies a larger volume), but does not imply the freezing of the molecules at specific locations. This motion mechanism is called *surface diffusion* and will be discussed in Section 5.4.

Chapter 4

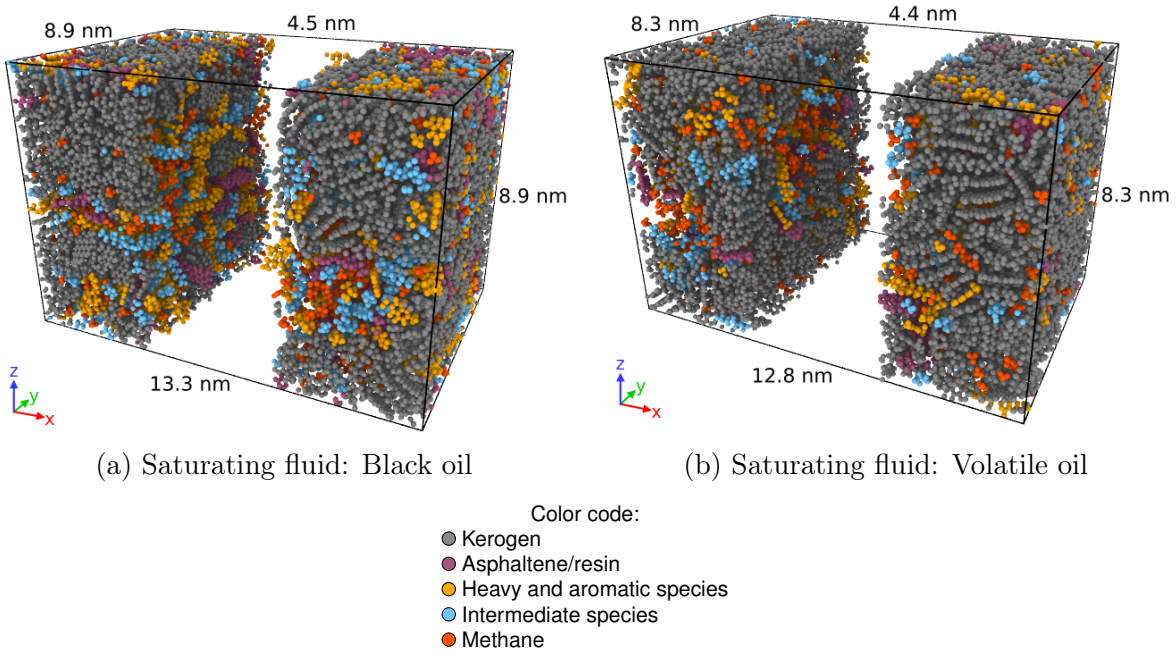
FLUID MIGRATION FROM KEROGEN TO A MICROFRACTURE

4.1 Introduction

This chapter studies the migration of fluid molecules from within organic pores to a microfracture placed next to the kerogen models saturated with oil mixtures created in Chapter 3. This is a follow up of the material presented in the first part of the article *A molecular dynamics study of primary production from shale organic pores* (Perez and Devegowda, 2020b). The article shows that under primary production, the lighter hydrocarbon species are produced more rapidly compared to the heavier compounds. For the present work, the simulation times are extended and provide additional insights.

4.2 Description of the systems

Once the mature models of kerogen saturated with oil mixtures reach thermodynamic equilibrium conditions, an empty space perpendicular to the x-direction is inserted in each system; the width of the empty space is approximately half the length of the simulation box in each case. The dimensions in the y- and z-direction are kept the same as those of the simulation box, as illustrated in Figure 4.1. This empty region can be



(a) Saturating fluid: Black oil

(b) Saturating fluid: Volatile oil

Figure 4.1: Microfracture in contact with saturated kerogen.

thought of as a slit pore in contact with smaller nanopores, or as a microfracture crossing organic matter that is connected to a larger fracture network. The surface area of the slit pore is that in the yz-plane. The implementation of periodic boundary conditions in all directions effectively produces a series of parallel slit pores or microfractures in contact with organic matter.

When the microfracture is introduced, many of the bonds between atoms in the kerogen monomers are broken in order to create the kerogen-pore interface. Consequently, the kerogen structures are kept rigid throughout all subsequent simulations. The rigidity of the kerogen body forbids the ability to study phenomena such as kerogen swelling due to adsorption (Pathak et al., 2017) or pore wiggling (Marbach et al., 2018) and their effects on fluid transport and are not considered in this work.

4.3 Simulation details and thermodynamic considerations

After each microfracture is introduced, the systems are allowed to evolve and relax for 1 ns in the NVT ensemble to investigate how fluid molecules migrate towards the

microfracture. The temperature of the system is kept fixed at $T = 355$ K (180°F) and the initial pressure of the fluids inside the pores is $P = 300$ atm (4400 psi) whereas the initial pressure in the microfracture is virtually zero since no fluid molecules are present in that spatial region. The time step is 1 fs and the temperature of the fluid molecules inside the organic pores in kerogen and in the microfracture are controlled independently by implementing two thermostats that explicitly rescale the velocities of the particles in each region, as discussed in Section A.5.2. The rescaling is applied to only the translational degrees of freedom for the particles. All other parameters (for instance the force fields with which the kerogen and hydrocarbon atoms are described) are enumerated in Section 3.3.

The insertion of the microfractures yields two spatial gradients, namely a pressure gradient and a concentration gradient between the interior of kerogen and the microfractures. These gradients will drive the flow of fluid molecules from the region of high potential (the interior of kerogen) to the one of low (the microfracture). The number and type of molecules in the microfracture are monitored as a function of time to calculate the temporal evolution of the composition of the mixture and the pressure that the fluid would exert in a similar unconfined box under equilibrium conditions. Intuitively, one would expect that both composition and pressure in the two regions equalize after some long enough period of time. However, as discussed in Section A.4.4, if the two regions are considered as two subsystems that are allowed to exchange matter, then thermodynamic equilibrium conditions will be achieved only once the chemical potential of every molecular species equalizes in the two subsystems, not the pressure or the composition separately. Further, this situation is not free expansion where neither heat nor work is exchanged with an external source (Callen, 1985; Schroeder, 2000) because both subsystems are coupled to thermostats that transfer energy from or to them.

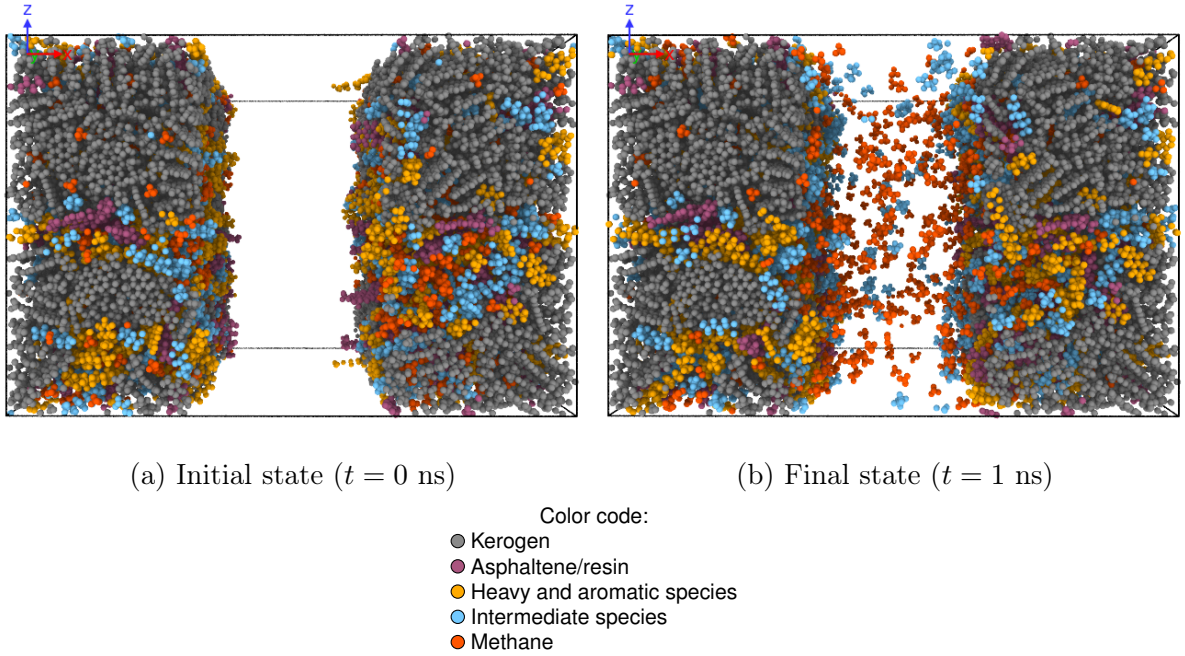
4.4 Composition of the fluid in the microfracture

Several different conditions for fluid flow from kerogen to the microfracture are studied in this work and are described in this section. The first one is considered as the base case. All other results, obtained by varying some specific simulation variable, are compared to the base case. The variables modified for comparison with the base case are (1) simulation time, (2) initial particle velocities (initial conditions), (3) microfracture width and (4) initial in-situ fluid composition.

4.4.1 Base case

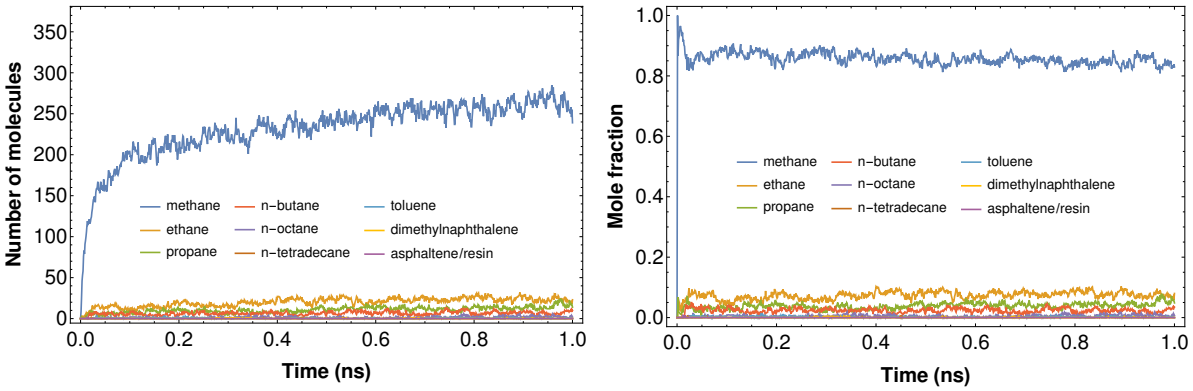
The base case considers the migration of molecules from kerogen saturated with the black oil created in Chapter 3 to the microfracture. Initially, the system is allowed to relax for 1 ns. Figure 4.2 compares the initial ($t = 0$ ns) and final ($t = 1$ ns) states of the system when a few fluid molecules have migrated to the microfracture. Figure 4.3 presents the temporal evolution of the molecules in the microfracture. Noticeably, mostly methane and a few intermediate components such as small alkane chains are found in the microfracture, whereas heavy and aromatic components are completely absent. Figure 4.3(a) shows that methane, the lightest component, is the dominant species in the microfracture. Recall that methane is the most abundant species in the free fluid phase as discussed in Section 2.6.2 and the one with the highest mobility as explained in Section 3.6, factors that promotes its migration. A smaller number of other small linear chains such as ethane, propane, and n-butane are also present in the microfracture. However, Figure 4.3(b) shows that, except for the initial sharp rise in methane migration, the relative amount of each species (i.e. the composition) remains fairly constant throughout the simulation.

Figure 4.4 presents a visual comparison between the composition of the initial hydrocarbon mixture within kerogen with the average composition in the microfracture over the last half of the simulation (i.e. over 0.5 ns). The mixtures are remarkably different, with the oil in the microfracture lighter in comparison to the initial in-situ fluid. The



(a) Initial state ($t = 0$ ns) (b) Final state ($t = 1$ ns)

Figure 4.2: Comparison of the (a) initial and (b) final states in the base case.



(a) (b)

Figure 4.3: Temporal evolution of the (a) number of molecules and (b) mole fraction of every species in the microfracture in the base case.

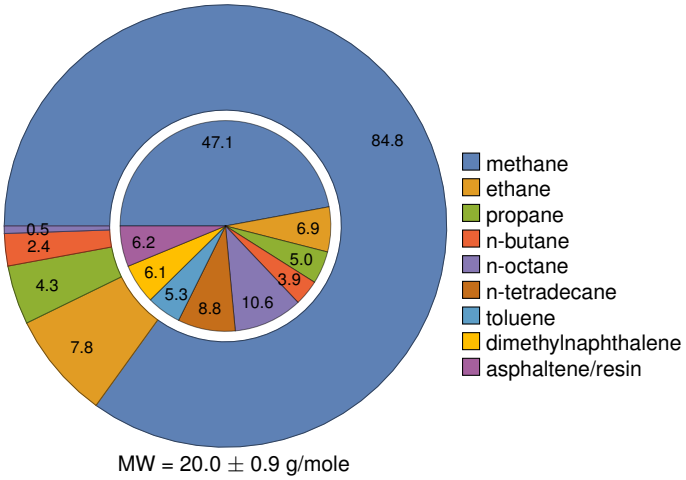
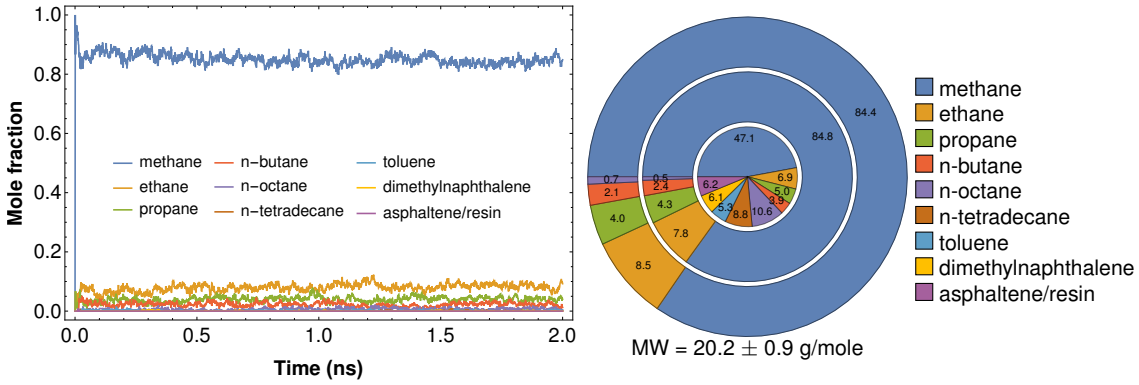


Figure 4.4: Comparison between the composition (mole fraction) of the hydrocarbon mixtures initially in organic pores (innermost pie chart) and in the microfracture (outermost pie chart) in the base case.

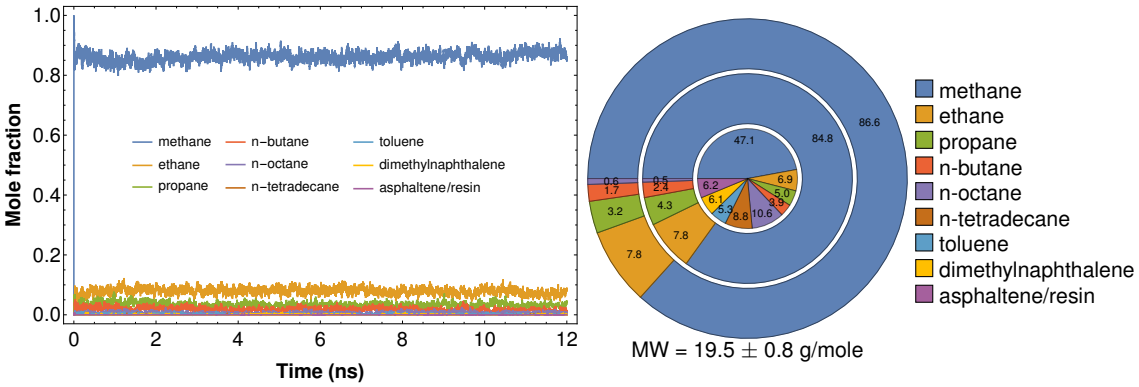
mole fractions of methane and C_{7+} in the microfracture are approximately 85% and less than 1%, respectively, and the average molecular weight is 20.0±0.9 g/mole, therefore the migrated fluid corresponds to a dry or a light wet gas (Holstein, 2007). This observation is not exclusive to the base case, as will be shown below.

4.4.2 Effect of simulation time

As discussed in Section A.2, there is a risk that simulations for 1 ns do not provide representative results. This section presents a simulation run of 2 ns to investigate whether a longer simulation time shows qualitatively similar results as with a 1 ns run. The time evolution of the composition of the fluid in the microfracture and the average composition over the last half of the simulation are presented in Figure 4.5(a). The time evolution on the left exhibits a similar behavior as with the base case, that is the composition of the migrated fluids remains fairly constant over time. The pie chart on the right shows the comparison between the in-situ fluid composition (innermost pie chart) and the average composition over the last half of the simulations (0.5 ns in the base case, 1 ns in the present case) of the migrated fluid in both cases (exterior pie charts), and the average fluid composition in the microfracture is also similar to that in the base case, with an average molecular weight in this case of 20.2 ± 0.9 g/mole.



(a) Simulation time = 2 ns



(b) Simulation time = 12 ns

Figure 4.5: Composition (mole fraction) of the hydrocarbon mixtures in the microfracture after longer simulation times. The figures on the left display the time evolution of the composition over time, and the pie charts show the average composition over the last half of the simulations. The innermost pie chart corresponds to the composition of the fluid initially in organic pores, the middle pie chart is the composition obtained in the base case, and the outermost pie chart represents the composition in the microfracture after longer simulation times.

To further test the importance of simulation time, a much longer simulation is run for 12 ns and the results are presented in Figure 4.5(b). As in the previous case, the time evolution of the fluid composition in the microfracture behaves similar to the base case, and more importantly, the values of mole fraction for every molecular species are very similar to those obtained in the base case, which suggests that the overall composition of the fluid mixture that arrives at the microfracture should be not expected to change after the first few nanoseconds of simulation. This in turn indicates that the composition of the mixture in the microfracture is largely independent on the simulation time. Instead, as will be demonstrated in Section 4.4.5, it depends on the in-situ fluid composition within kerogen. In fact, the average composition and molecular weight of 19.5 ± 0.8 over the last 6 ns of the migrated fluid presented in the pie chart demonstrate that the values are statistically equal to those obtained in the previous cases. Irrespective of the simulation time, the migrated fluid can be classified as a dry or wet gas.

4.4.3 Effect of initial conditions

Section A.2 points out that the choice of initial conditions may have an undesired effect on the results. This section presents a simulation run with different initial velocities, drawn from a Maxwell-Boltzmann distribution with temperature $T = 355$ K, to assess if the initial conditions are influencing the observations made in the previous sections. The results presented in Figure 4.6 highlight the similarities with the base case. The composition remains fairly constant over time, and the average fluid composition and molecular weight of 19.7 ± 0.9 g/mole are statistically equal to those obtained in the base case, which supports the statement that the results are not biased by the choice of initial conditions.

4.4.4 Effect of microfracture width

In this section, the width of the microfracture is increased to 7.9 nm as shown in Figure 4.7(b) to assess the impact of microfracture width on the final results. The results presented in Figure 4.8 underscore the similarities with the base case. The composition

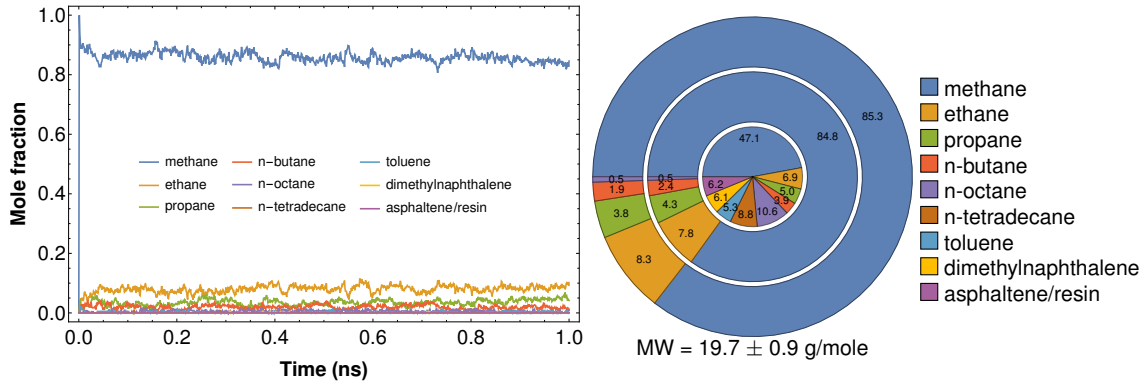


Figure 4.6: Composition (mole fraction) of the hydrocarbon mixture in the microfracture using different initial conditions. The figure on the left displays the time evolution of the composition over time, and the pie chart shows the average composition over the last half of the simulation. The innermost pie chart corresponds to the composition of the fluid initially in organic pores, the middle pie chart is the composition obtained in the base case, and the outermost pie chart represents the composition in the microfracture using different initial conditions.

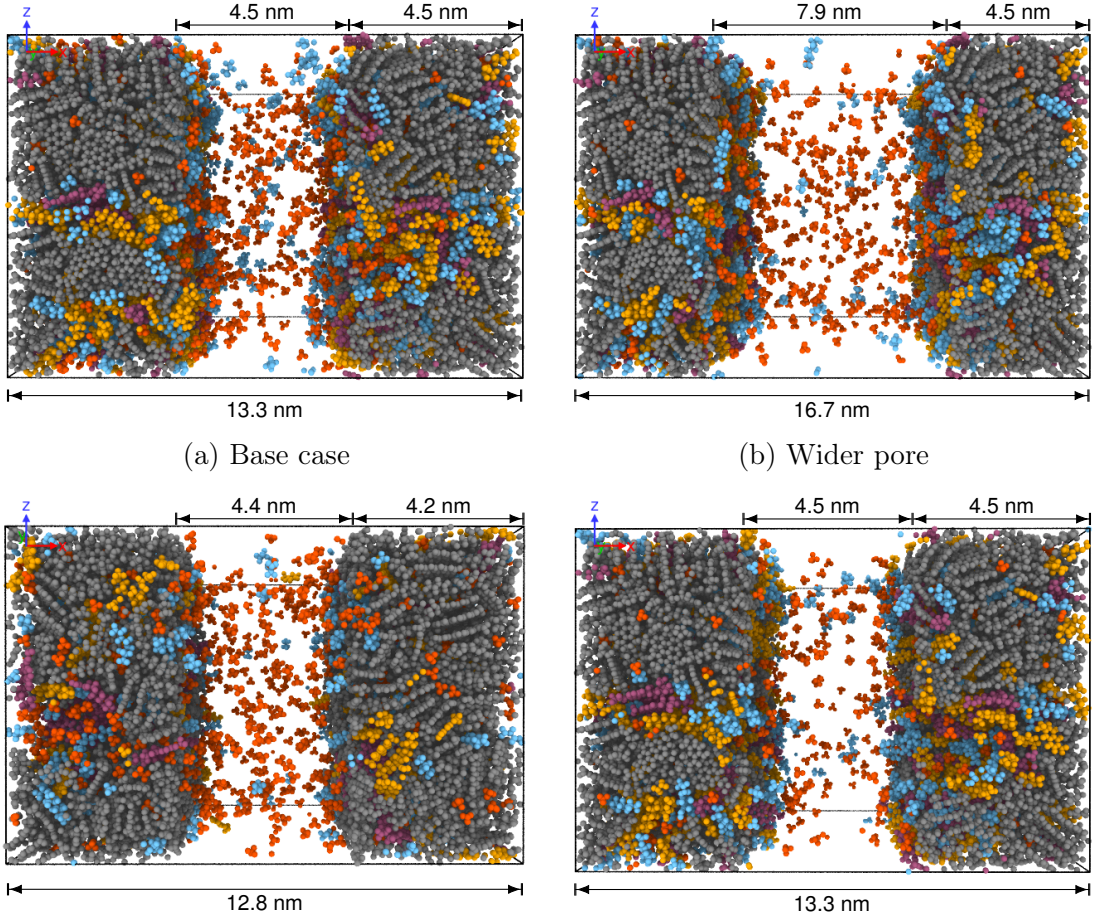
remains fairly constant over time, and the average fluid composition and molecular weight of 19.5±0.7 g/mole lead to the conclusion that the migrated fluid is a dry gas or a wet gas. Therefore, the microfracture width does not influence the produced fluid composition, as seen in these results.

4.4.5 Effect of in-situ fluid composition

Finally, in this section, the initial oil composition is varied to assess its impact on the fluid migrating to the microfracture. The base case, as discussed earlier, is a black oil with molecular weight 80.68 g/mole. The new cases are a volatile oil and a black oil with a heavier molecular weight. The composition of the volatile oil is given in Table 3.3. The heavier black oil corresponds to the fluid remaining within kerogen after the base case simulation for 1 ns and its composition is given in Table 4.1.

Kerogen saturated with a volatile oil

The mature kerogen model saturated with a volatile oil shown in Figure 4.1 is used in this case. The composition of the volatile oil is described in Table 3.3 and the width of the microfracture is 4.4 nm. All other simulation settings are kept the same as in the base case. The final state of the simulation is illustrated in Figure 4.7(c) and the compositions



(a) Base case

(b) Wider pore

(c) Saturating fluid: Volatile oil

(d) Saturating fluid: Heavier black oil

Figure 4.7: Comparison of the final states during kerogen depletion to a microfracture.

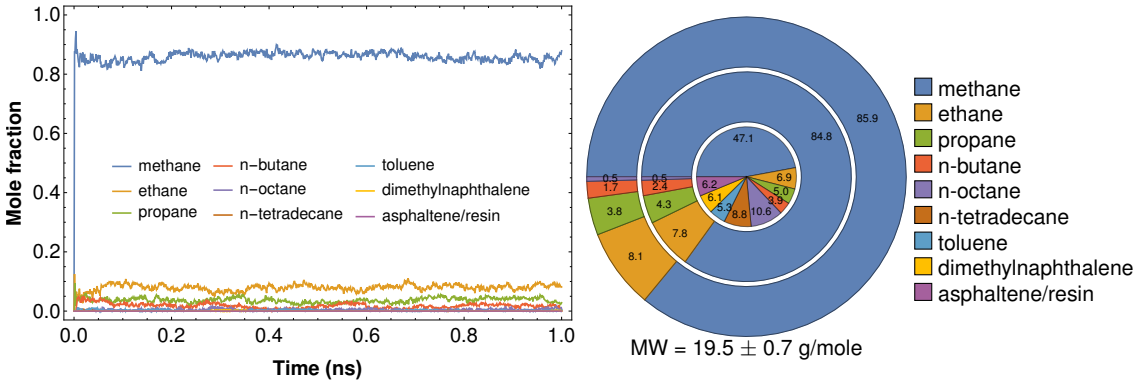
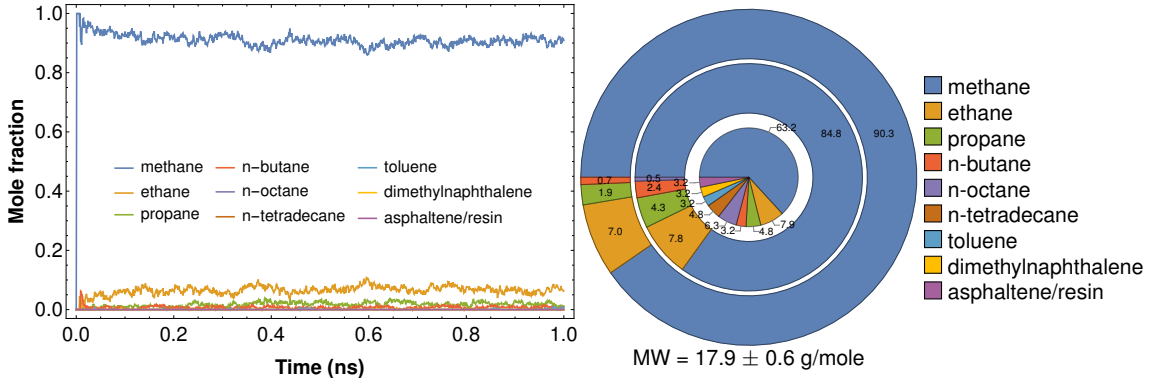


Figure 4.8: Composition (mole fraction) of the hydrocarbon mixture in a wider microfracture. The figure on the left displays the time evolution of the composition over time, and the pie chart shows the average composition over the last half of the simulation. The innermost pie chart corresponds to the composition of the fluid initially in organic pores, the middle pie chart is the composition obtained in the base case, and the outermost pie chart represents the composition in a wider microfracture.

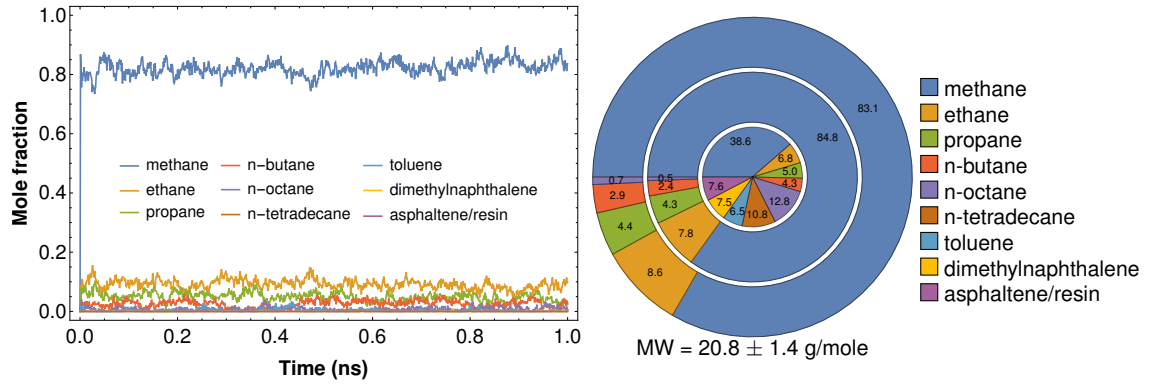
Table 4.1: Molecular composition of the heavier black oil.

Species	Mole fraction (%)
methane	38.6
ethane	6.8
propane	5.0
n-butane	4.3
n-octane	12.8
n-tetradecane	10.8
toluene	6.5
dimethylnaphtalene	7.5
asphaltene/resin	7.6
Total molecular weight	94.33 g/mole

are presented in Figure 4.9(a). The migrated fluid composition shows that the curves become relatively flat after 0.5 ns and that the methane fraction in the microfracture is higher than with the black oil of the base case. This is because the mole fraction of methane in the volatile oil is higher and, except for ethane, the mole fractions of all other species in the volatile oil are lower, as described in Table 3.3. The pie chart clearly demonstrates that, in comparison with the base case, the composition of the migrated fluid is lighter with an average molecular weight of 17.9 ± 0.6 g/mole.



(a) Saturating fluid: Volatile oil



(b) Saturating fluid: Heavier black oil

Figure 4.9: Composition (mole fraction) of the hydrocarbon mixture in the microfracture using different in-situ fluids. The figures on the left display the time evolution of the composition over time, and the pie charts show the average composition over the last half of the simulation. The innermost pie chart corresponds to the composition of the fluid initially in organic pores, the middle pie chart is the composition obtained in the base case, and the outermost pie chart represents the composition in the microfracture using different in-situ fluids.

Kerogen saturated with a heavier black oil

For this study the heavier black oil is the remaining fluid in organic pores at the end of the base case simulation[1] and this simulation is carried out after the fluid molecules in the microfracture have been removed to start off with an empty pore as in all previous cases. The composition of the heavier black oil is presented in Table 4.1 and its molecular weight is 94.33 g/mole. Figure 4.7(d) shows the final state of the system after 1 ns and Figure 4.9(b) presents the compositions of the migrated fluid. The temporal evolution of the composition curves in the figure on the left behaves similar to those in the base case. The methane fraction in the migrated fluid is lower than that of the base case while the ethane fraction is higher. This is a consequence of the fluid composition at the beginning of the simulation, i.e. the in-situ fluid composition. The pie chart on the right summarizes the average mole fraction values and reveals that the concentration of methane is indeed lower than in the base case whereas the concentration of ethane and n-butane are larger, which is due ultimately to the initially fewer molecules of methane available to flow in the free fluid phase. The average molecular weight of 20.8 ± 1.4 g/mole is higher (although comparable) than the base case. The statistical error is larger because the number of fluid molecules is smaller and fluctuations are higher. Nonetheless, this fluid is richer in intermediate components than those obtained in all previous cases and its molecular weight indicates that it is representative of a wet gas or a light gas condensate. This observation supports the conclusion that a heavier in-situ fluid yields a heavier migrated fluid. As seen with the volatile oil, a lighter initial fluid yields a lighter migrated fluid.

The results obtained in this section indicate that the initial composition of the fluids saturating organic pores determines the composition of the fluids in the microfracture, or conversely, the composition of the fluids in the microfracture is a reflection of the composition of the fluids that saturate organic pores. In conclusion, the lighter the in-situ fluid, the lighter the fluid in the microfracture.

[1]For that reason, the initial pressure in the organic pores is not 300 atm (4400 psi) but smaller. This, however, is not expected to affect the *qualitative* observations drawn from the results.

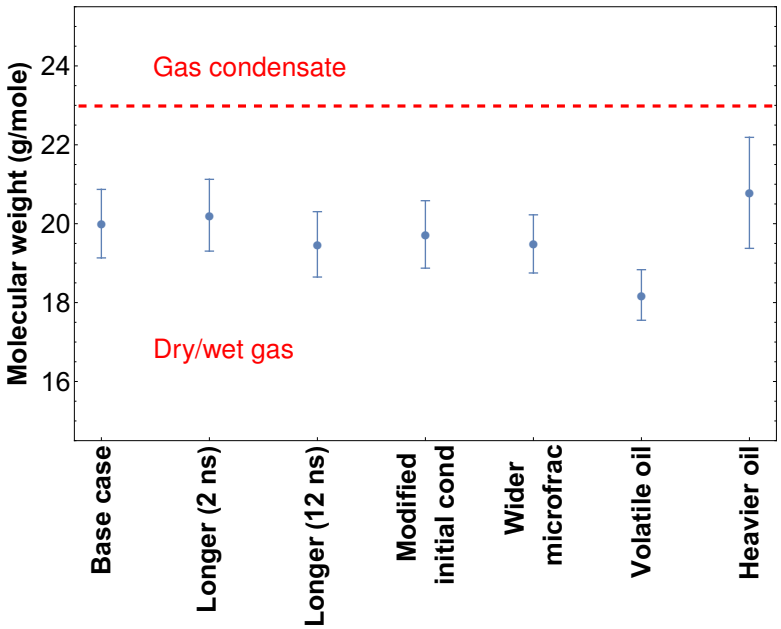


Figure 4.10: Average molecular weight of the fluids in the microfracture.

4.4.6 Summary

The discussion in the previous sections indicates that the average composition and molecular weight of the fluids in the microfracture strongly depend on the initial composition of the fluids in organic pores, but are rather insensitive to other parameters such as initial conditions, simulation time, and pore width. The results are conveniently detailed in Table 4.2 and visually summarized in Figure 4.10.

4.5 Amount of fluids in the microfracture

The previous section discussed the composition of the migrated fluid which was shown to vary with the initial in-situ fluid and remains similar if the initial fluid is the same. However, with longer simulation times or wider microfracture widths, the absolute number of molecules of each species in the microfracture is different. Figure 4.11 shows the temporal evolution of the number of molecules present in the microfracture for all the cases studied, and Figure 4.12(a) presents a compiled summary of the average number of molecules in the microfracture over the last 0.1 ns of the simulations. Since the number of molecules of methane that migrate to the microfracture is one or two orders

Table 4.2: Average composition and molecular weight of the fluids in the microfracture.

Species	Base case[a]	Longer[b] (2 ns)	Longer[c] (12 ns)	Initial conditions[a]	Wider pore[a]	Volatile oil[a]	Heavier black oil[a]
	Mole fraction (%)						
methane	84.8 ± 1.5	84.4 ± 1.5	86.6 ± 1.5	85.3 ± 1.4	85.9 ± 1.2	90.3 ± 1.4	83.1 ± 2.2
ethane	7.8 ± 1.0	8.5 ± 1.1	7.8 ± 1.1	8.3 ± 1.1	8.1 ± 0.9	7.0 ± 1.1	8.6 ± 1.5
propane	4.3 ± 2.4	4.0 ± 0.8	3.2 ± 0.8	3.8 ± 0.9	3.8 ± 0.8	1.9 ± 0.7	4.4 ± 1.3
n-butane	2.4 ± 0.7	2.1 ± 0.7	1.7 ± 0.6	1.9 ± 0.7	1.7 ± 0.7	0.7 ± 0.4	2.2 ± 1.1
n-octane[d]	0.5 ± 0.4	0.7 ± 0.4	0.6 ± 0.4	0.54 ± 0.35	0.46 ± 0.32		
n-tetradecane[d]							
toluene[d]							
dimethylnaphthalene[d]							
asphaltene/resin[d]							
Total molecular weight (g/mole)	20.0 ± 0.9	20.2 ± 0.9	19.5 ± 0.8	19.7 ± 0.9	19.5 ± 0.7	17.9 ± 0.6	20.8 ± 1.4

[a] Averages over 0.5 ns.
[b] Average over 1 ns.
[c] Average over 6 ns.
[d] Quantities for which the error is greater or equal to the value are considered zero and not listed.

of magnitude larger than those of the other species, the same information is presented in semilogarithmic scale in Figure 4.12(b).

With the longer simulations illustrated in Figures 4.11(b) and 4.11(c), the number of migrated molecules is larger. The number of molecules of methane plateaus in Figure 4.11(c) for a 12 ns simulation run. The information in Figure 4.11(c) is shown in Figure 4.13 with an inset picture for the non-methane species showing a leveling off for all hydrocarbon species, indicating that the overall composition stabilizes for a 12 ns run.

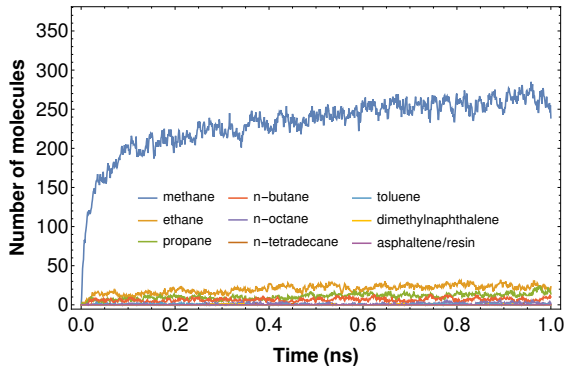
A comparison of the base case run for 1 ns in Figure 4.11(a) and a longer simulation time of 12 ns in Figure 4.11(d) shows that more molecules migrate from the kerogen to the microfracture, but as shown earlier, the composition remains fairly stable. In Figure 4.11(e), a wider microfracture allows for more migration of fluid molecules but again the composition is the same as seen earlier. Figure 4.11(g) shows lower methane numbers which reflects the lower methane mole fraction in the initial fluid which is a heavier black oil.

As indicated in Table 3.1, the number of hydrocarbon molecules utilized to make up the volatile oil is lower than those utilized to build the black oil. This matter does not allow a direct comparison between cases in Figures 4.11(a) and 4.11(f) and therefore nothing is said about it here.

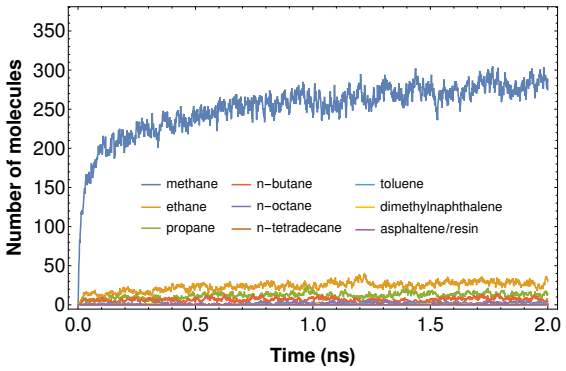
4.6 Pressure exerted by the fluid in the microfracture

As mentioned in Section 4.3, the average number and type of molecules in the microfracture are monitored as a function of time to calculate the pressure that the mixture would exert in a simulation box, the size of the microfracture, under equilibrium conditions. Thus, the average mixture in the microfracture in every case described in Section 4.4 is utilized to run equilibrium molecular dynamics simulations for 2 ns to compute the average pressure over the last nanosecond of simulation, and the results for each case are shown in Figure 4.14.

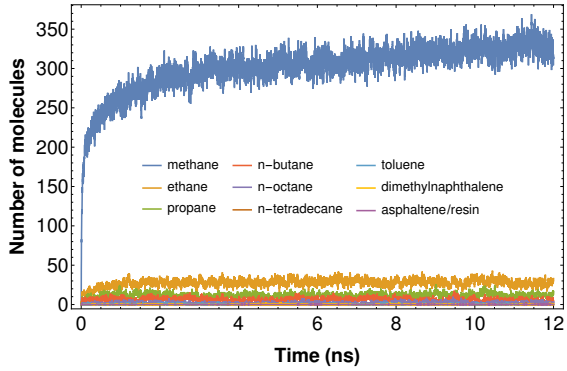
The highest pressure is obtained for the migrated fluid in the longest simulation



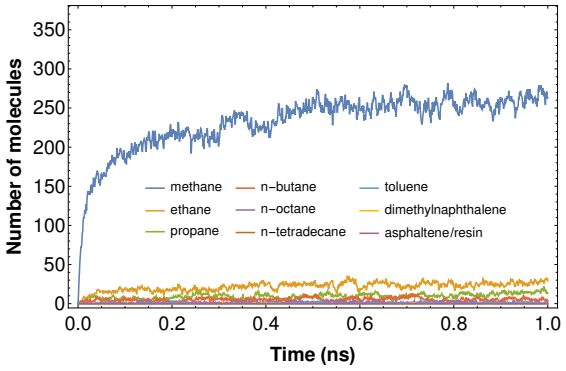
(a) Base case



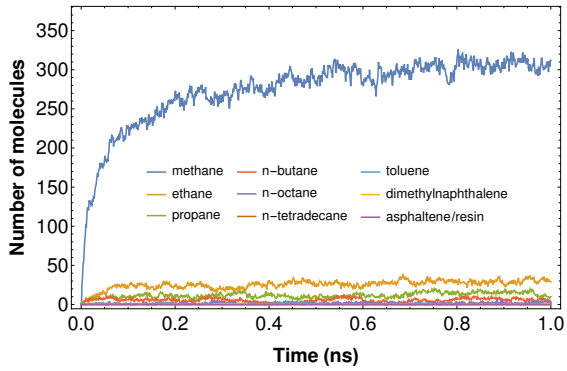
(b) Simulation time = 2 ns



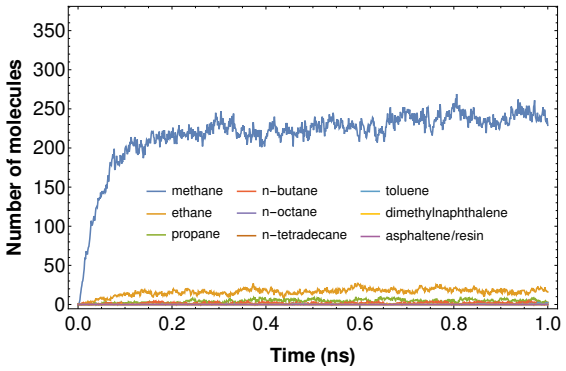
(c) Simulation time = 12 ns



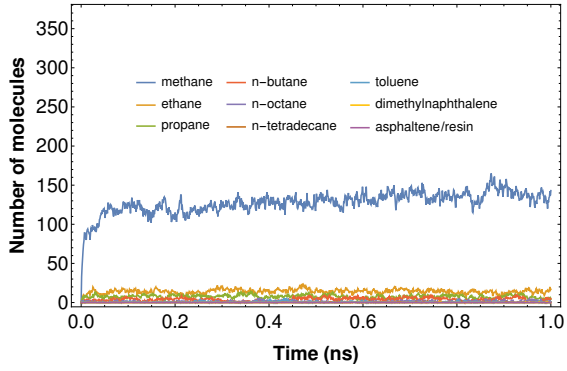
(d) Different initial conditions



(e) Wider microfracture



(f) Saturating fluid: Volatile oil



(g) Saturating fluid: Heavier black oil

Figure 4.11: Number of molecules in the microfracture. The range in the vertical axis is the same in all figures to facilitate direct comparisons.

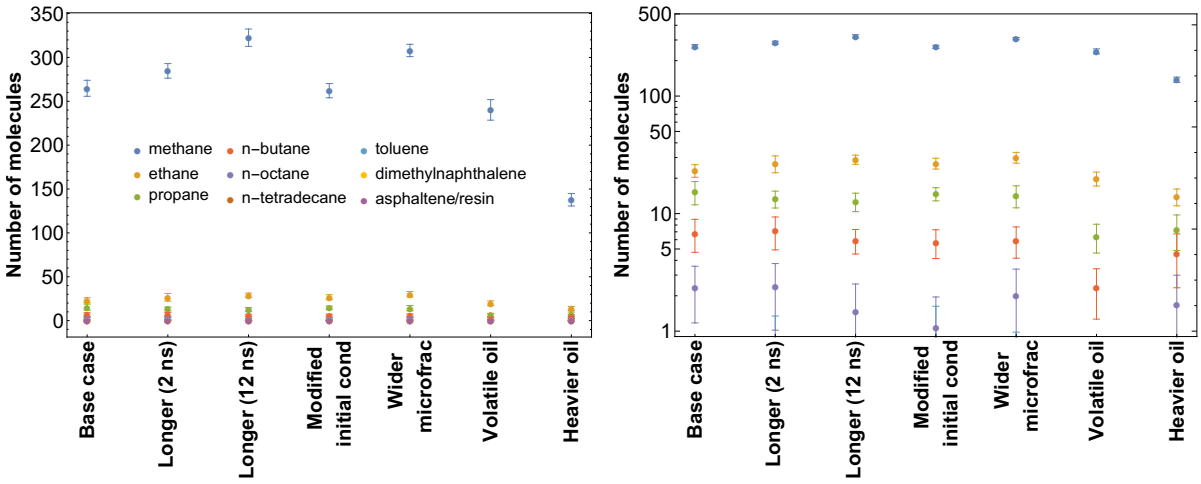


Figure 4.12: Number of molecules per species in the microfracture. (a) Linear scale; (b) semilogarithmic scale.

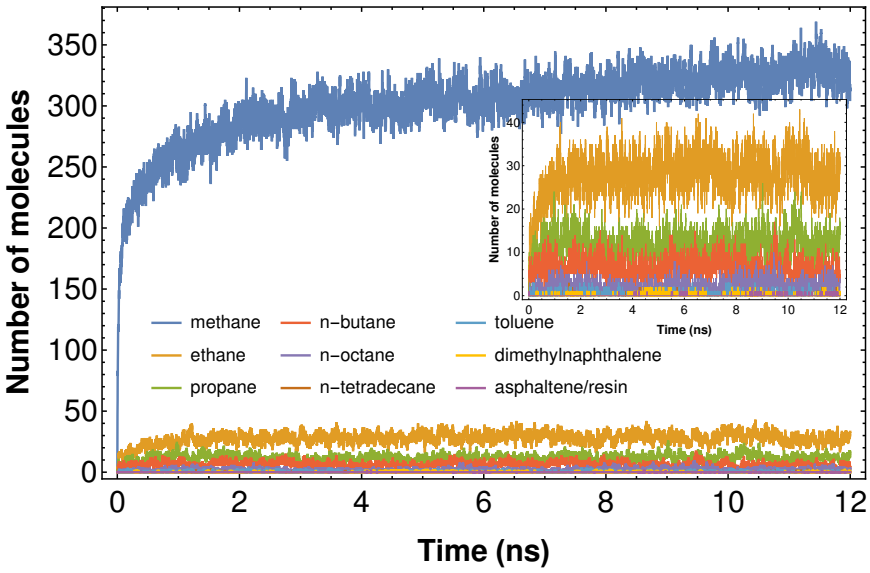


Figure 4.13: Time evolution of the number of molecules in the microfracture during 12 ns. The inset is a close-up of all species but methane.

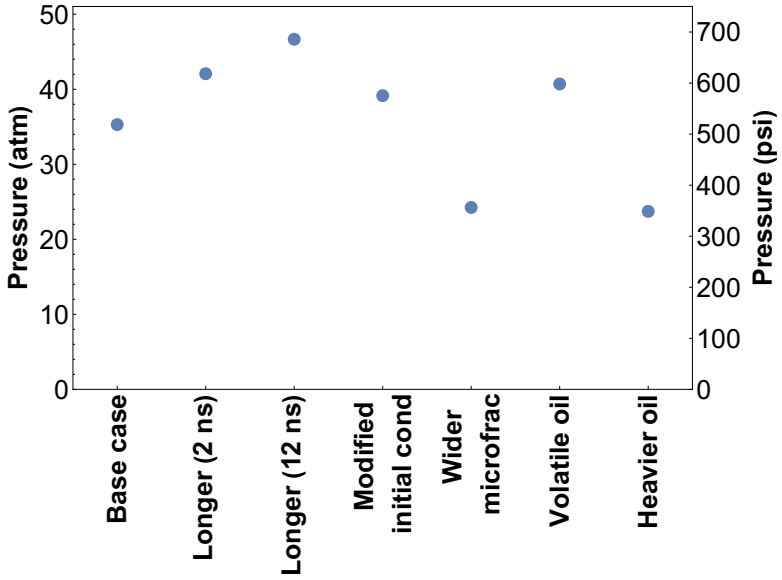


Figure 4.14: Average pressure for every fluid composition in the microfracture.

time of 12 ns and a small microfracture width as expected. The lowest value of final pressure is obtained with a wider microfracture or with the heavier black oil as a starting point. This is ultimately a reflection of the final number of molecules in the microfracture at the end of each case.

4.7 Final remarks

The simulations carried out in this chapter indicate that smaller pores in organic matter feed microfractures with hydrocarbons mixtures that are lighter than the remaining fluids in the pores. Even the longest simulation run produced a fluid composition similar to that of shorter simulations. Though there is no possible way to prove the following generalization using molecular dynamics simulations or field observations, the results suggest that the composition hydrocarbon mixtures in larger organic pores is lighter than those in smaller and smaller organic pores, even if the fluids reach a state of thermodynamic equilibrium (or quasi-equilibrium) over geologic times. The practical implication from a petroleum engineering standpoint is that the oil produced from organic-rich shale reservoirs during primary depletion is lighter than the fluid left behind in the reservoir. In fact, albeit it is impossible to determine the actual in-situ fluid composition in shale wells due to leakage during sample retrieval and the practice of venting cores for safety

reasons at the surface (Freeman et al., 2012, 2013), the results discussed here are observed in shale wells. For instance, Freeman et al. (2012) report real flowing compositional behavior from 5 wells in a North America shale gas play and notice that the produced gas may be composed of 80+% methane but the fraction of heavier hydrocarbons is lower than in the in-situ fluids. This whole idea will be the focus of the next chapter.

Chapter 5

MODELING PRIMARY RECOVERY. FLUID FLOW MECHANISMS IN ORGANIC NANOPORES

5.1 Introduction

Chapter 4 focused on the migration of fluid molecules initially in organic pores to microfracture placed next to the kerogen models created in Chapter 3. In this chapter, the microfracture is regarded as a crack in contact with organic matter in a long fracture network. Fluid molecules in organic matter migrate to the microfracture and this mixture is *produced* over time. This is a follow up of the material presented in the second part of the article *A molecular dynamics study of primary production from shale organic pores* (Perez and Devegowda, 2020b). In that work, only the depletion of the mature kerogen model saturated with a black oil was studied. This work also covers the depletion of the model saturated with a volatile oil to provide more complete picture of primary recovery.

To model primary recovery, a series of molecular dynamics simulations that span a total time of 9 ns are carried out. Every step of the process is a simulation in the NVT ensemble; however, the removal of hydrocarbons from the microfracture which denotes primary production, as implemented in this work, renders the entire process neither an NVT nor a μVT (grand canonical (see Section A.4.4)) ensemble simulation. The process depicts production of hydrocarbons in a step-wise manner. The reasoning behind the

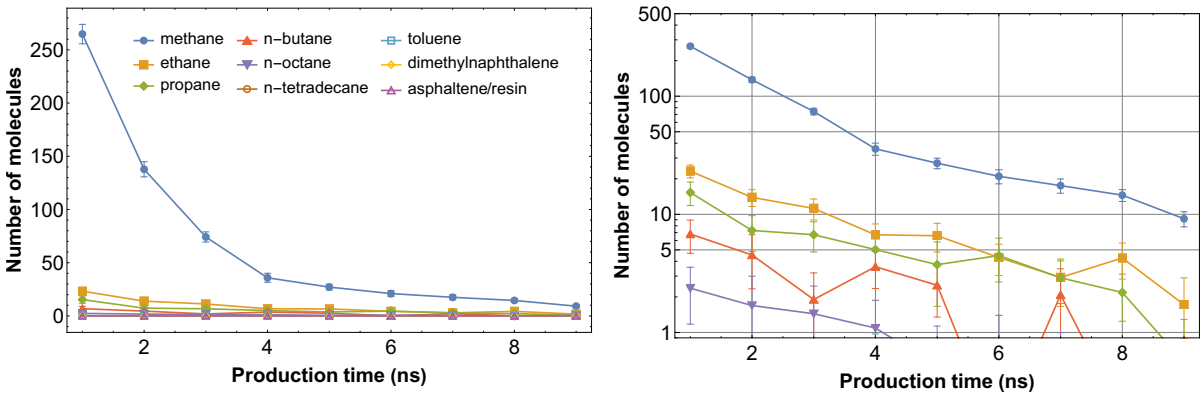
adopted approach is explained below.

5.2 Simulation details

As seen in Section 4.4.2, *NVT simulations for 1 ns give representative samples of the hydrocarbon mixture in microfracture.* Consequently, the systems shown in Figure 4.1 are allowed to relax for 1 ns holding constant temperature T, volume V and number of particles N, and then the molecules that are located in the microfracture at the end of the simulation are removed to mimic production from a high permeability fracture adjacent to organic matter. This process is repeated 8 more times for a total simulation time of 9 ns. This is a discrete, step-by-step version of continuous production.

The composition of the fluids in the microfracture is monitored as a function of time and the average pressure that the mixture would exert in a simulation box the size of the microfracture under equilibrium conditions is computed as described in Section 4.6. The initial conditions of the systems are temperature $T = 355$ K (180°F) and pressure of the fluids in the pores $P = 300$ atm (4400 psi). The time step utilized is 1 fs. The temperature of the fluid molecules inside kerogen and in the microfracture are held constant throughout the simulations and controlled independently implementing two thermostats by explicitly rescaling the velocities of the particles in each region, as explained in Section A.5.2.

According to the discussions in Sections 4.4 and 4.5, the composition of the oil mixture in the microfracture is expected to be representative of what would actually flow and be produced in a continuous fashion, although the number of molecules of each species varies with simulation time, that is if a longer simulation time were utilized to relax the systems in each step of the process, a larger number of molecules per species would be expected to flow to the microfracture to be produced. However, keeping the duration of every stage the same (in this case 1 ns) is considered to give a fair way to compare the number of molecules per species found in the microfracture and removed (produced) at the end of every step.



(a) (b)

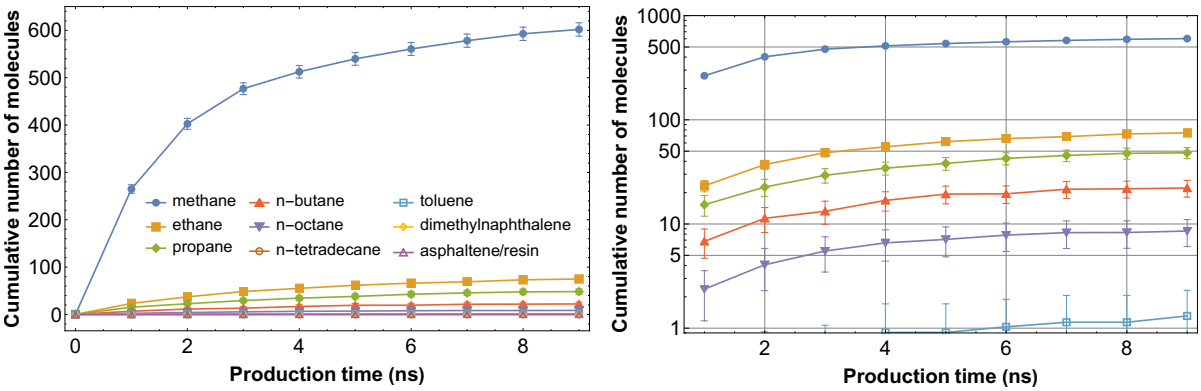
Figure 5.1: Instantaneous production of molecules over time for the black oil case. (a) Linear scale; (b) semilogarithmic scale.

5.3 Primary production

5.3.1 From organic pores saturated with a black oil

This section describes primary production from kerogen saturated with a black oil. Using the step-by-step process described earlier, the number of each of the molecular species produced over each nanosecond step is shown in Figure 5.1(a) on a Cartesian plot and in Figure 5.1(b) on a semilogarithmic plot. In terms of cumulative production over time, the results are depicted in Figure 5.2. Figures 5.1 and 5.2 show that methane is the predominant species produced followed by the alkane chains ethane, propane, n-butane and n-octane, which is the expected behavior according to the discussions in the previous chapters. In addition to these linear alkanes, Figure 5.2(b) shows the production of a few molecules of toluene which, although being aromatic molecules, have more mobility in organic pores than the long chains of n-tetradecane.

Figures 5.1(a) and 5.2(a) indicate that the rate of production of the producible hydrocarbon species is high at early times but decreases rather quickly. The reason is that the free fluid phase, mainly composed of methane, expands and is depleted first. As the free fluid phase is produced, the pressure in the pores drops and light species such as methane, ethane, propane and n-butane in the adsorbed phase desorb and recharge the free fluid phase. In line with this observation, Herdes et al. (2018) also report that



(a) (b)

Figure 5.2: Cumulative production of molecules over time for the black oil case. (a) Linear scale; (b) semilogarithmic scale.

heavier hydrocarbon species preferentially remain adsorbed on to the surface of organic nanopores as small as 1 and 2 nm under depressurization of confined fluids. Furthermore, the number of molecules of each species decreases as the corresponding molecular weight increases. This is in agreement with observations from molecular simulations (Falk et al., 2015; Collell et al., 2015), and with field production data (Freeman et al., 2012) and numerical simulation results (Freeman et al., 2013).

The average molecular weight of the mixture over the last 0.5 ns of every depletion step is represented by the blue curve in Figure 5.3. This figure demonstrates that the average composition of the produced fluid does not remain constant with time but transitions from lighter to heavier as production time increases which has also been observed in field data (Freeman et al., 2012). This is a consequence of the early depletion of the free methane followed by the subsequent desorption of light and intermediate species as time goes by.

The average number and type of molecules produced are utilized to calculate the pressure that the mixture would exert in a simulation box the size of the microfracture under equilibrium conditions. Thus, the average mixture produced at every nanosecond step is utilized to run equilibrium molecular dynamics simulations for 2 ns and the average pressure is computed over the last nanosecond of simulation. The results are presented in Figure 5.4 and reveal that the pressure in the microfracture is far below the initial pressure of 300 atm (4400 psi) of the fluids in organic pores, just as observed in Section

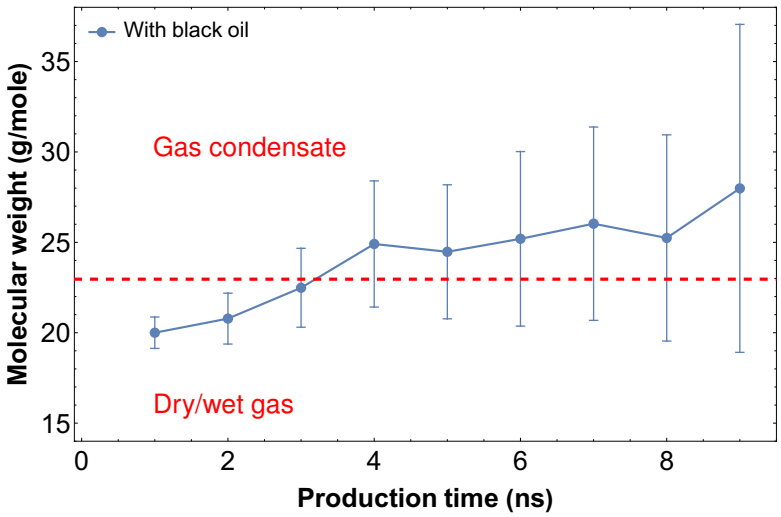


Figure 5.3: Molecular weight of the oil mixture in the microfracture. The points correspond to the average molecular weight over the last 0.5 ns in each simulation.

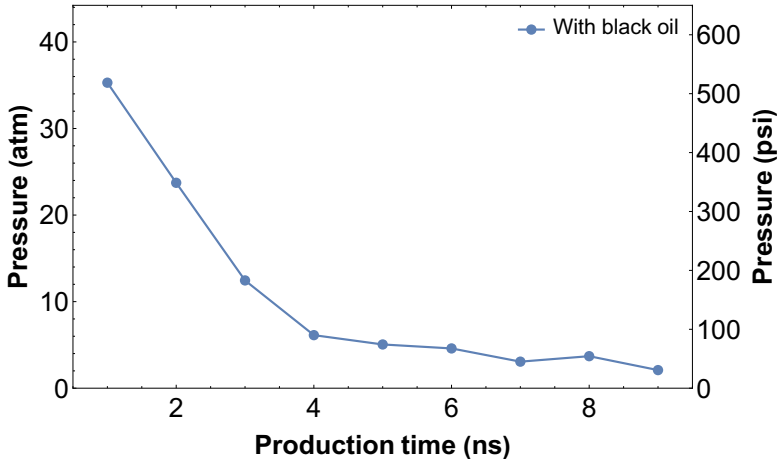


Figure 5.4: Average pressure exerted by the produced oil in the microfracture with time.

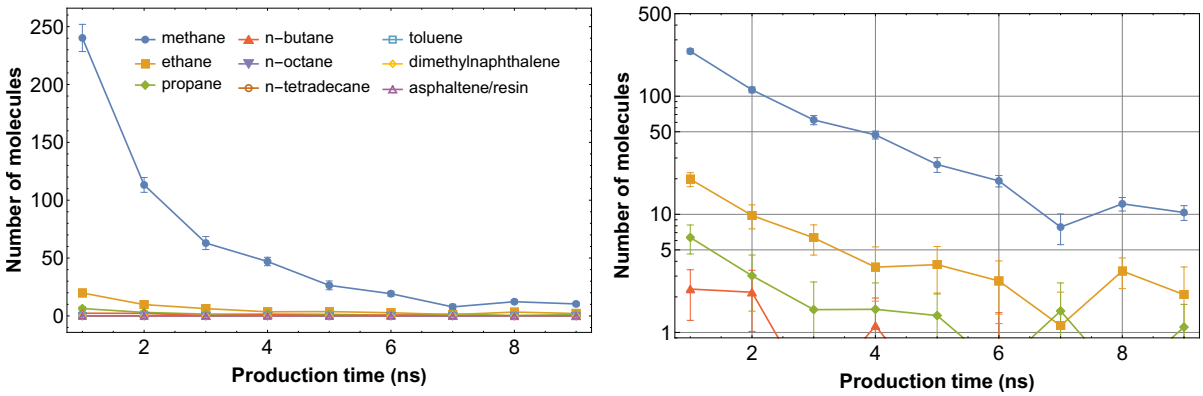
4.6. A pressure drop also occurs inside organic pores as fluid molecules migrate to the microfracture. Such pressure drop promotes desorption of molecules in the adsorbed phase. However, and as mentioned above, no desorption of heavier molecules is seen. Herdes et al. (2018) also report that as the pressure decreases, the free fluid continues to be predominantly comprised of the lighter components because of the preferential adsorption of the heavier species, in agreement with the observations reported in this work.

5.3.2 From organic pores saturated with a volatile oil

A similar analysis as in the previous section is carried out here with the kerogen model saturated with a volatile oil. The instantaneous production of each of the hydrocarbon species is provided in Figure 5.5(a) on a Cartesian plot and in Figure 5.5(b) on a semilogarithmic scale. The cumulative production on Cartesian and semilogarithmic scales are provided in Figure 5.6. The results again demonstrate that methane is the predominant species in the production stream followed by the other small alkane chains. However, unlike production from the kerogen model saturated with a black oil, very few molecules of n-butane migrate to the microfracture to be produced and neither n-octane nor toluene molecules are produced in the present case. There are two reasons for this observation. All hydrocarbon species, except for methane and ethane, constitute a smaller mole fraction in the volatile oil. Second and more importantly, the pore size distribution ranges towards smaller pores (and pore throats) as shown in Figure 3.3, and this restricts the ability of the larger molecules to move towards the microfracture.

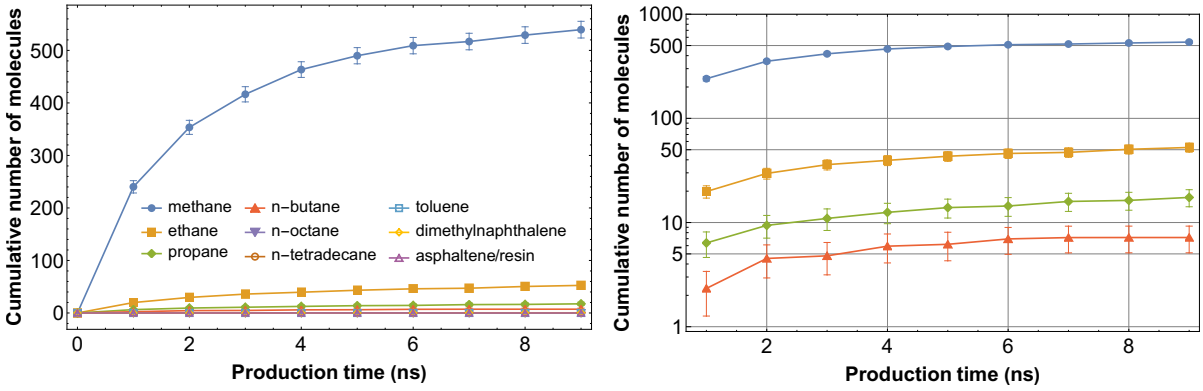
Similarly, Figures 5.5(a) and 5.6(a) indicate that the rate of production of the producible hydrocarbon species is high at early times but decreases rather quickly. The initial production is from the free fluid phase and as the pressure drops, a few of the light and intermediate species desorb and become available. This observation is similar to what was observed with the black oil.

The average molecular weight of the mixture over the last 0.5 ns of every depletion step is represented by the yellow curve in Figure 5.7. The average composition of the



(a) (b)

Figure 5.5: Instantaneous production of molecules over time for the volatile oil case. (a) Linear scale; (b) semilogarithmic scale.



(a) (b)

Figure 5.6: Cumulative production of molecules over time for the volatile oil case. (a) Linear scale; (b) semilogarithmic scale.

produced fluid does not remain constant with time in this case either but transitions from lighter to heavier as production time increases. The produced fluid is lighter when the volatile oil saturates kerogen compared to the case with a black oil which is a consequence of the initial fluid compositions and pore sizes.

To conclude the analysis of oil production from organic pores filled with a volatile oil, the average mixture produced at every nanosecond step is utilized to run equilibrium molecular dynamics simulations for 2 ns and compute the average pressure over the last nanosecond of simulation. The results are shown in Figure 5.8 and, as in the black oil case, reveal that the pressure in the microfracture is far below the initial pressure of 300 atm (4400 psi) of the fluids in organic pores. Interestingly, the average pressure exerted by the produced mixtures in both cases on the faces of the microfracture are similar. This

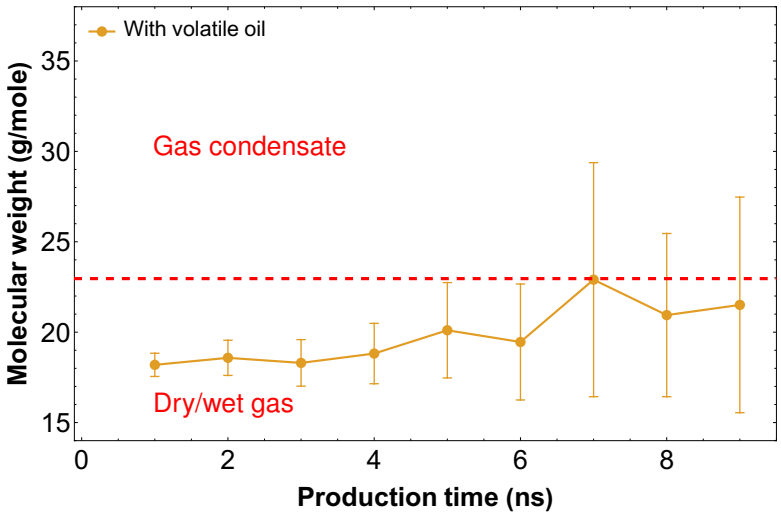


Figure 5.7: Molecular weight of the oil mixture in the microfracture. The points correspond to the average molecular weight over the last 0.5 ns in each simulation.

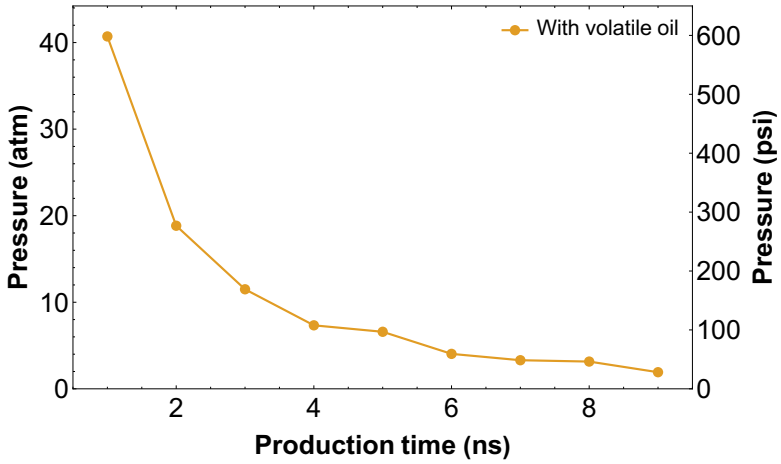


Figure 5.8: Average pressure exerted by the produced oil in the microfracture with time.

observation, nonetheless, must be used with caution since the pressure values reported are not computed directly inside the microfracture but using separate equilibrium molecular dynamics simulations.

5.4 Flow mechanisms during primary depletion

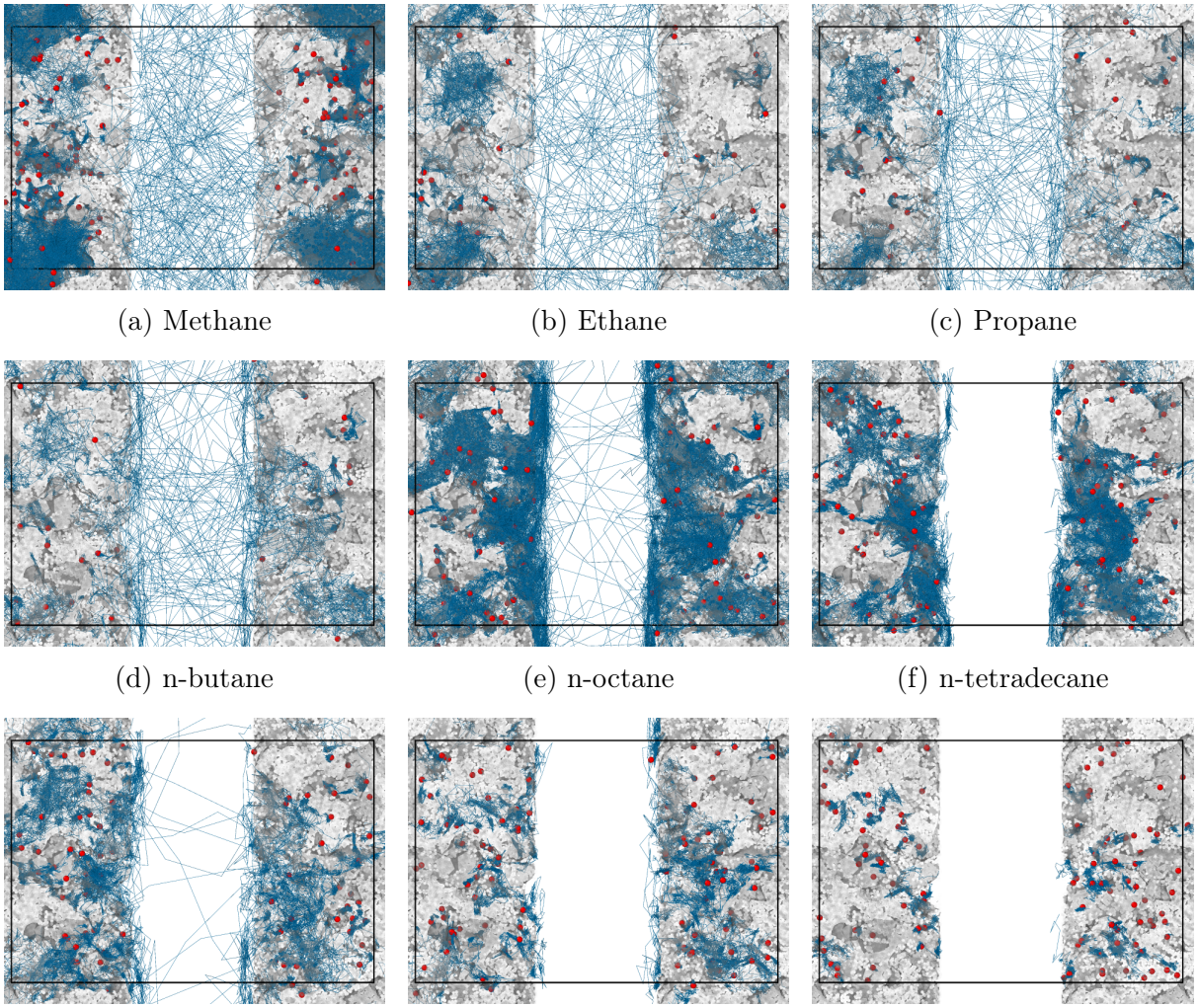
The previous section demonstrated that the produced fluid compositions are lighter in comparison to the in-situ fluids. This is because the intermediate and heavier species tend to be adsorbed and do not contribute appreciably to the production even under depressurization of the pores. The initial production is sustained by the fluid expansion

of the free fluid phase. This section follows up on the previous discussion to address two questions: (1) What factors govern production of a specific hydrocarbon species? (2) What happens to the unproduced molecules while the others migrate to the microfracture?

Recall the coexistence of a free and adsorbed phase inside organic pores. As demonstrated in Section 2.6.2, molecules of a certain species may be found in either phase, depending on the molecular size, geometry and weight of the species. There are light molecules close enough to the pore surface to be counted among the adsorbed molecules, but their mobility is enough to allow a molecule to abandon its location as time goes on. Molecules of heavier species, however, remain close to the pore surface most of the time and do not have the ability to separate from it, but do have the ability to move along the surface. Thermal fluctuations might also enable a molecule to detach from the pore surface and move across the pores for some time, but the enhanced interactions between fluid molecules and the walls will promote its adsorption. In general, the fluid molecules can move inside the pores, some at large extent, others at short.

In the simulations presented in this work, there exists both a pressure and a concentration gradient from the kerogen towards the microfracture. These gradients drive the transport of fluid molecules in both the free and adsorbed phase towards the microfracture. However, once a fluid molecule reaches the microfracture, it may dissolve in the existing fluid within the fracture, or travel back to the pores in kerogen and be re-adsorbed, or get adsorbed at the kerogen-microfracture interface. Note that, in general, the heavier species are preferentially adsorbed over the lighter species; if a heavy molecule reaches the microfracture, it is very likely that it cannot remain in solution and re-adsorb at either the kerogen-microfracture interface or the pore surface. Thus, the answer to the questions is that only the fluid molecules that are able to reach the microfracture and remain in solution in the mixture formed there are produced; the rest of the molecules either cannot desorb from the pore surface or get re-adsorbed.

To illustrate all this, Figure 5.9 shows the trajectories described by the center of mass of each of several hydrocarbon species that remain unproduced at the end of 9 ns



(a) Methane (b) Ethane (c) Propane

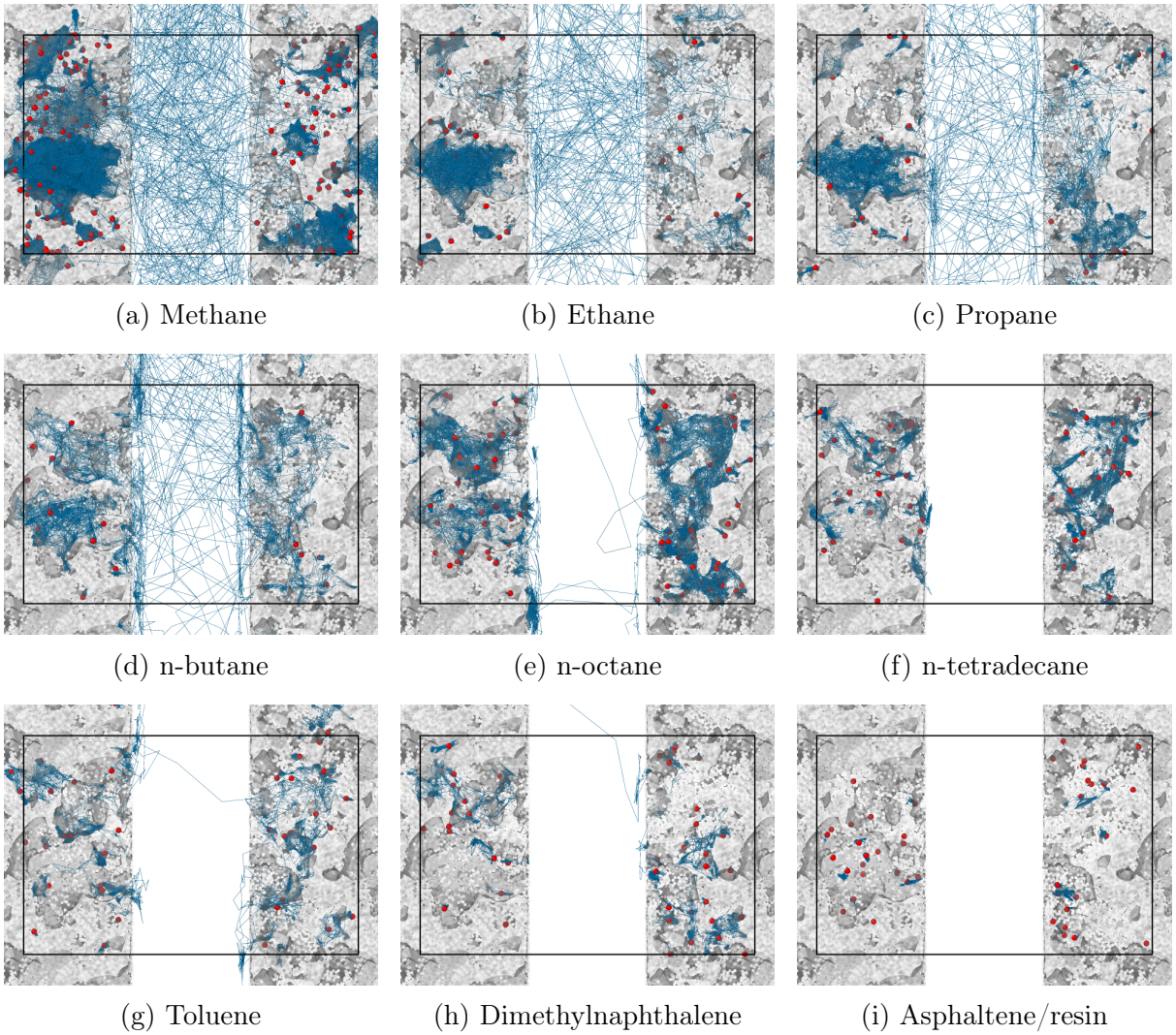
(d) n-butane (e) n-octane (f) n-tetradecane

(g) Toluene (h) Dimethylnaphthalene (i) Asphaltene/resin

Figure 5.9: Trajectory (blue lines) of the center of mass (red circles) of molecules that are not produced after 9 ns of primary depletion for the black oil case. Black box: simulation box. Light-gray: atoms in kerogen. The dark-gray surface is a render of the pore surface.

for the kerogen saturated with a black oil and Figure 5.10 presents similar plots for the kerogen saturated with a volatile oil. The observations from the figures are described below:

- The main point is that the unproduced molecules do move in the system as time goes on.
- The plots for methane, ethane, propane and n-butane show dense trajectories of these unproduced lighter hydrocarbon species within the microfracture. Although the production stream is largely comprised of the lighter species, there is a large proportion of these species that remain unproduced. This is a consequence of the



(a) Methane (b) Ethane (c) Propane

(d) n-butane (e) n-octane (f) n-tetradecane

(g) Toluene (h) Dimethylnaphthalene (i) Asphaltene/resin

Figure 5.10: Trajectory (blue lines) of the center of mass (red circles) of molecules that are not produced after 9 ns of primary depletion for the volatile oil case. Black box: simulation box. Light-gray: atoms in kerogen. The dark-gray surface is a render of the pore surface.

step-wise production model adopted here. In reality, if the fluid remains free, it is likely to be produced.

- There is a low density of trajectories for methane, ethane, propane and n-butane within kerogen in the vicinity of the microfracture, indicating that a large percentage of the free fluid has been produced.

- There is a high density of trajectories in the pores far from the microfracture for methane, ethane, propane and n-butane, indicating that most of the unproduced free fluid is located away from the microfracture.

- There is a high density of trajectories for unproduced n-octane, n-tetradecane and

toluene in kerogen adjacent to the microfracture in the same regions where there is little light species, indicating that while n-octane, n-tetradecane and toluene molecules are mobile, adsorption of these species renders them unable to be produced. This observation is supported by the absence of an appreciable number of these molecules moving in the microfracture.

- The heaviest species dimethylnaphthalene and asphaltene/resin are not immobile, although they are largely unproduced. There is a modest motion of these molecules towards the fracture.

- A closer look reveals that n-octane, n-tetradecane, toluene, dimethylnaphthalene and asphaltene/resin are adsorbed and move along the pore surfaces. Figure 5.11 shows a composite picture for n-octane, tetradecane, toluene and dimethylnaphthalene illustrating this point for a volatile oil.

- The motion of ethane and n-octane in an organic pore is compared in Figure 5.12. While ethane molecules move across the entire pore, n-octane molecules move primarily along the pore surface.

- Figure 5.13 displays the motion of methane, ethane, propane and n-octane molecules in a different organic pore. It illustrates the existence of pores in which methane accounts for almost all the free fluid phase and that the motion of light species such as ethane and propane occurs mostly along the surface. The practical consequence is that not the entire fraction of light hydrocarbons can be produced during primary recovery.

Based on these observations, it is proposed that there are two main mechanisms of fluid transport in organic nanopores: (1) the expansion of the free fluid phase, and (2) the displacement along the surface of the adsorbed molecules. The molecules that move along the surface are said to exhibit surface diffusion.

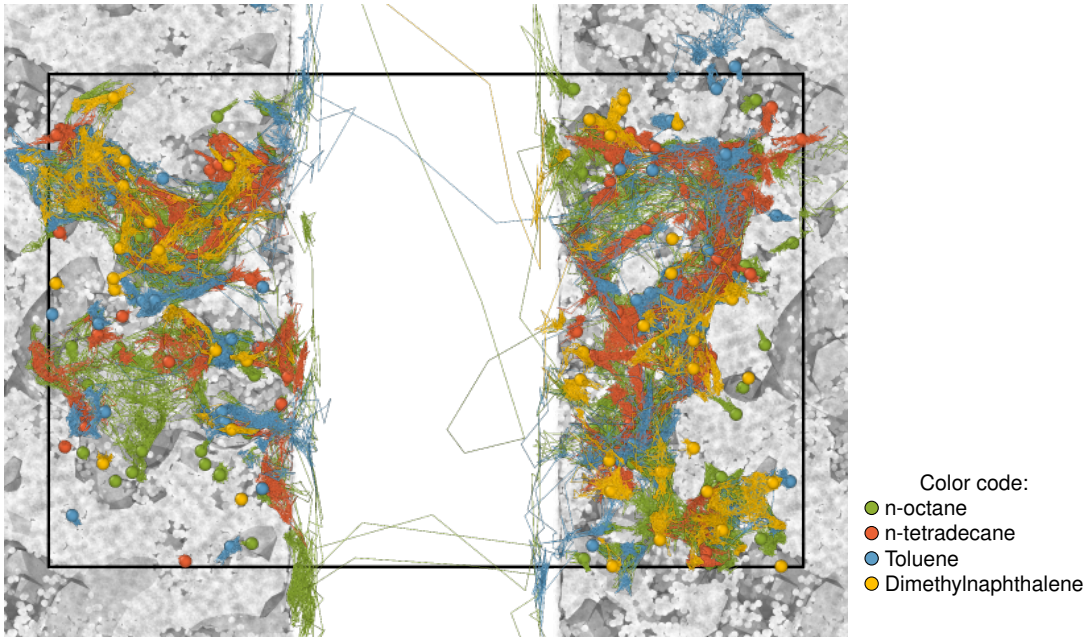
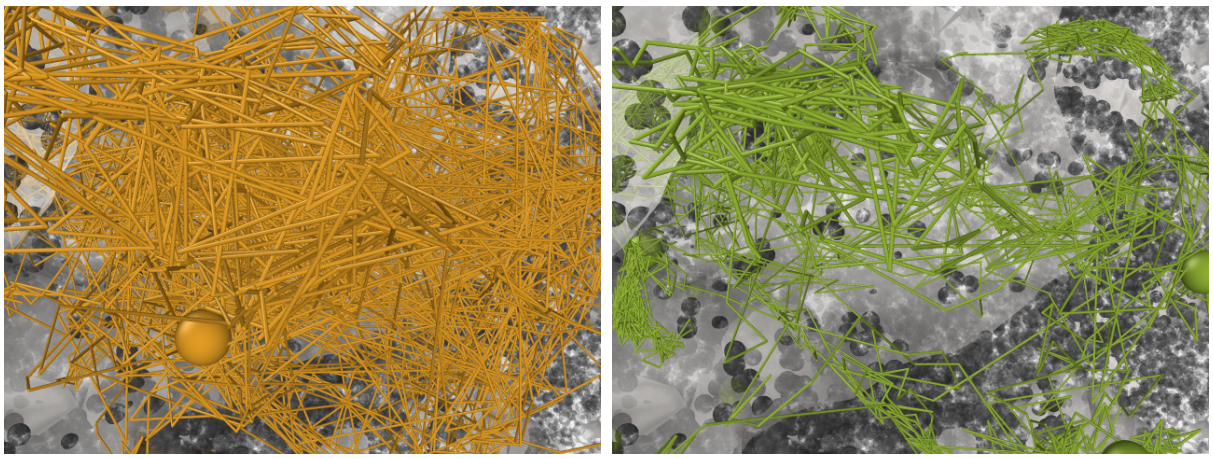


Figure 5.11: Surface diffusion of the heavier species in organic pores in the kerogen model saturated with a volatile oil. The color circles denote the center of mass of the molecules of each species that remain unproduced after 9 ns of primary depletion, and the lines trace their trajectories. Black box: simulation box. Light-gray: atoms in kerogen. The dark-gray surface is a render of the pore surface.

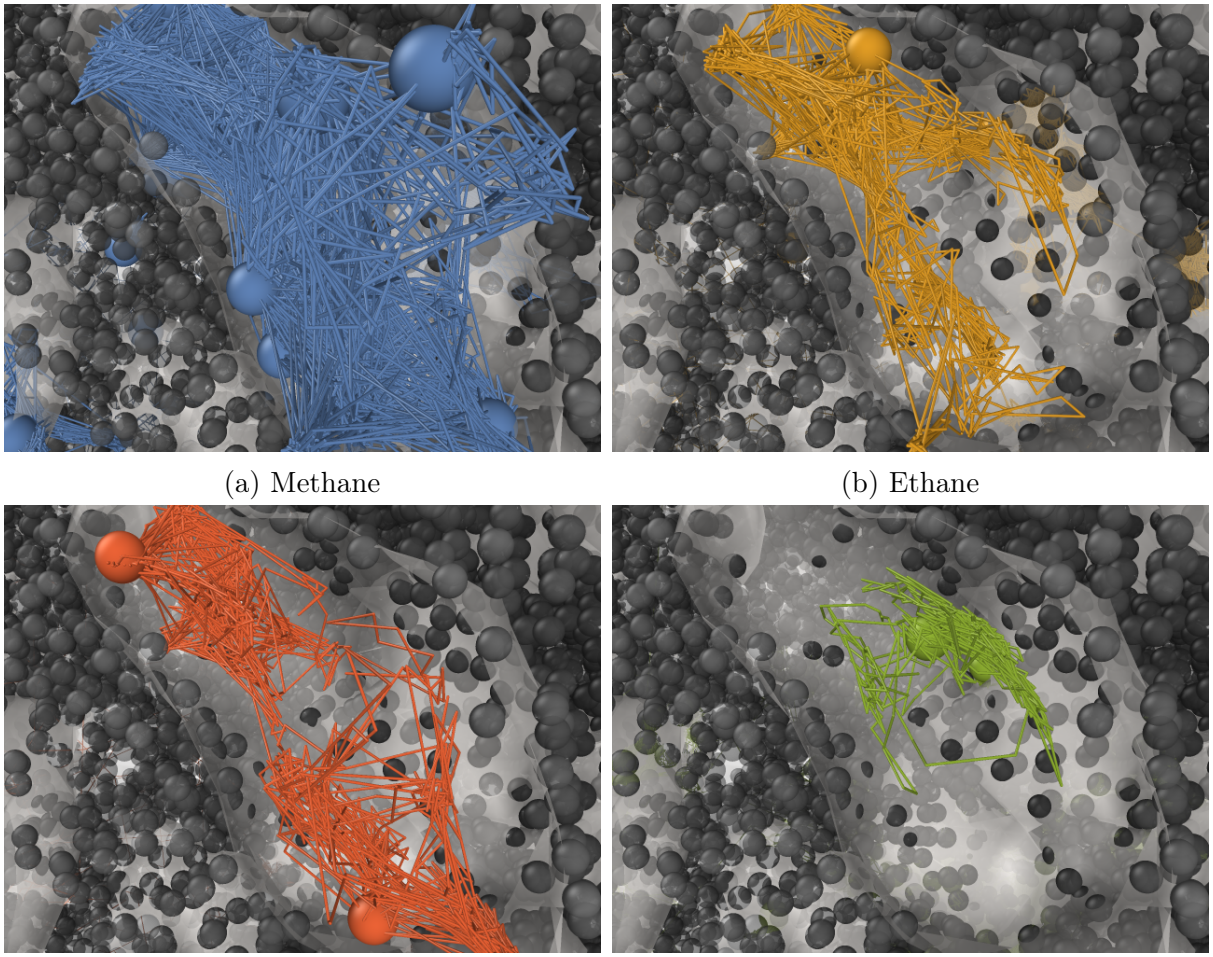


(a) Ethane (b) n-octane

Figure 5.12: Trajectories of the center of mass of (a) ethane and (b) n-octane molecules inside an organic pore in the kerogen model saturated with a volatile oil (the observation location is the same in both cases). The gray surface is a render of the pore surface.

5.5 Making the case for enhanced oil recovery

Section 5.3 presents the production of hydrocarbon molecules from organic pores. Although fluid flow is driven by pressure and composition gradients, and despite the pressure drop inside the organic pores, not all molecules are able to migrate from the



(a) Methane (b) Ethane

(c) Propane (d) n-octane

Figure 5.13: Trajectories of the center of mass of (a) methane, (b) ethane, (c) propane and (d) n-octane molecules inside an organic pore in the kerogen model saturated with a volatile oil (the observation location is the same in all cases). The gray surface is a render of the pore surface.

interior of kerogen to the microfracture to be produced. Regardless, Section 5.4 illustrates that the fluid molecules do move inside the pores, the lighter species across the entire pore volume while the heavier species move along the pore surface. This is true even when the pressure in the microfracture almost never happens in field operations.

To quantify the recovery of hydrocarbons during primary depletion, the amount of carbon atoms in the produced fluid at every step is compared to the initial number of carbon atoms present in the organic pores. Since far more molecules of methane are produced than any other species, this approach is adopted to normalize by the number of carbon atoms in the molecules. The results are presented in Figure 5.14. The recovery is computed at every step for the entire simulation, but taking into account the practical inconvenience to lower the pressure in the fractures below a certain threshold value,

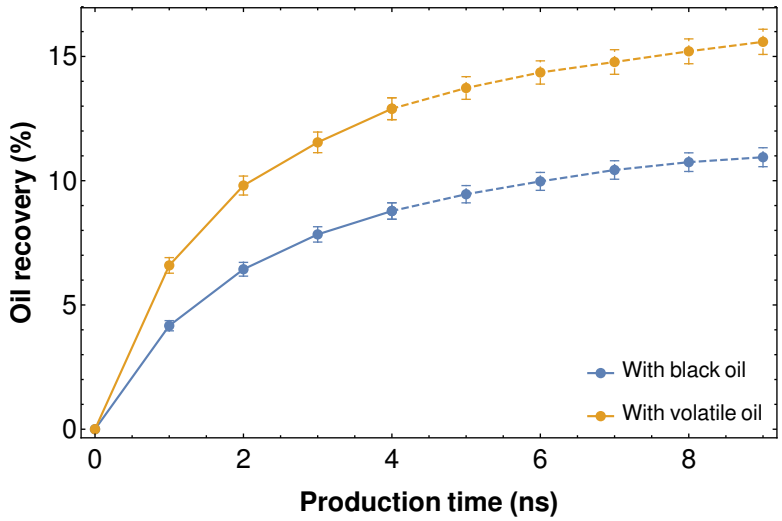


Figure 5.14: Oil recovery during primary depletion. The colors denote different oil mixtures filling the organic pores.

primary depletion will be stopped in this work before the pressure in the microfracture falls below 7 atm (100 psi), which corresponds to a production time of 4 ns represented by the solid curves in Figure 5.14. The results exhibit a recovery below 13% if the in-situ fluid is a volatile oil, and this number is below 9% in the case of a black oil, which corresponds to field observations (Clark, 2009; King, 2014). Therefore, a very large amount of oil is left behind in the pore space during primary depletion. The next chapter will discuss strategies to remediate this.

5.6 Final remarks

The presence of a microfracture in contact with porous organic matter saturated with an oil mixture introduces pressure and concentration gradients that drive fluid flow from the organic pores to the microfracture. The higher permeability in the fracture network as compared to that of the rocks facilitates the production of hydrocarbons from unconventional reservoirs. This process is mimicked in this work by periodically removing from the simulation box the fluid molecules that are found in the microfracture. The results discussed in this chapter exhibit qualitative agreement with both laboratory scale and field scale observations.

The production of different molecular species is not uniform, nor is the composition

of the oil mixture with production time. Recalling the coexistence of the two phases free and adsorbed in organic pores, the results suggest that the free phase expands like a gas to fill the available space whereas the molecules in the adsorbed phase slide along the pore surface. Furthermore, some intermediate sized molecules (represented here by n-octane and toluene) are able to reach the kerogen-microfracture interface. The bottom line is that adsorbed molecules also present some displacement in the pores instead of lingering still at fixed positions, as nicely illustrated in Figure 5.13. This idea underscores the importance of creating new surface area during hydraulic fracturing stages since intermediate species might migrate to the microfracture and, even if they tend to remain adsorbed on the microfracture faces, the injection of a solvent might desorb and dissolve part of those molecules to enhance oil production, in other words those molecules become the easiest targets for enhanced oil recovery. The feasibility of this concept is the subject of the next chapter.

Chapter 6

ENHANCED OIL RECOVERY. MODELING CYCLIC GAS INJECTION

6.1 Introduction

The previous chapter focused on primary production from organic pores in kerogen and demonstrated that hydrocarbon recovery is low. Additionally, the production stream is predominantly composed of the lighter components and to a lesser extent, the intermediate species. This presents a huge opportunity for enhanced oil recovery. However, the whole idea is neither new nor exclusive of this work. It has been long noticed that the primary recovery factor from unconventional reservoirs remains somewhere between 5 and 10% (Clark, 2009; King, 2014), necessitating some form of improved oil recovery (EOR). Shoaib and Hoffman (2009) present one the earliest attempts to simulate and prove improved recovery using gas flooding with carbon dioxide in the Bakken formation. Wan et al. (2013) also use a commercial numerical simulator to propose cyclic gas injection for pressure maintenance and achieving miscibility with oil. In another work, Meng et al. (2017) conduct experiments and use numerical modeling to examine the potential of cyclic injection to recover condensates from gas condensate reservoirs. There have been many more approaches to identify the physical mechanisms and quantify their relative impact when a solvent contacts reservoir fluids. Alfarge et al. (2017) compile

a comprehensive summary of the research efforts by formation, improved oil recovery method/mechanism and mode of implementation (analytical, experimental, simulation, pilot tests). More recently, Tovar et al. (2018) experimentally injected carbon dioxide and molecular nitrogen in organic rich shales and report that the governing process behind improved oil recovery is vaporizing gas drive. Hoffman and Rutledge (2019) evaluate oil swelling caused by diffusion and advection processes for gas injection in unconventional reservoirs using numerical reservoir simulations. In addition to that work, Hoffman and Reichhardt (2019) assess the relative significance of the four recovery mechanisms oil swelling, viscosity reduction, vaporization, and pressure support using a commercial numerical reservoir simulator. Although the simulation works mentioned show additional recovery and demonstrate the promise of enhanced oil recovery, the results entirely depend on the adequacy of the internal models implemented by the simulators. Dang (2019) introduces a workflow to evaluate the enhanced oil recovery potential of a play and proposes key factors that govern the success of cyclic gas injection in shale reservoirs. He notices that the enhanced oil recovery potential of a play strongly relies on how much removable hydrocarbons are unproduced in a formation after primary depletion.

From a microscopic standpoint, very few works have made use of molecular simulations to study enhanced oil recovery from organic rich shales using carbon dioxide as the injection gas (Pathak et al., 2016; Takbiri-Borujeni et al., 2019; Kazemi et al., 2019), but no work has attempted to simulate an injection/production cycle. Perez and Devegowda (2020c) present a molecular study of soaking with an organic solvent, but do not consider the production step during cyclic gas injection (also known as huff-n-puff).

This chapter focuses on the feasibility of enhanced oil recovery by periodically exposing the rock, and therefore the remaining fluids in organic pores, to some organic solvent of choice. This is a systematic and more thorough and detailed exposition than the work presented in *Molecular signatures of solvent injection in shale organic nanopores* (Perez and Devegowda, 2020d).

Table 6.1: Molecular composition of the oils after primary depletion for 4 ns.

Species	Mole fraction (%)	
	Produced oil	Remaining oil
methane	81.8	26.8
ethane	8.8	5.8
propane	5.5	4.7
n-butane	2.7	4.7
n-octane	1.1	16.2
n-tetradecane	0.0	14.0
toluene	0.2	8.3
dimethylnaphtalene	0.0	9.7
asphaltene/resin	0.0	9.9
Total molecular weight	21.1 g/mole	115.5 g/mole

6.2 Where are we after primary depletion?

As mentioned in Section 5.5, primary depletion stops before the pressure in the microfracture drops below 7 atm (100 psi), which corresponds to a production time of 4 ns. Figure 6.1 summarizes the results for primary recovery from the kerogen model saturated with a black oil. Figures 6.1(a) and 6.1(b) show the number of molecules produced on Cartesian and semilogarithmic scales, respectively, and Figures 6.1(c) and 6.1(d) show the cumulative number of molecules produced on Cartesian and semilogarithmic scales, respectively. The time evolution of the molecular weight of the produced fluid is presented in Figure 6.1(e) and it shows a progressive increase as more of the intermediate species desorb and are produced when the pore pressure decreases. Figure 6.1(f) compares the compositions of the produced fluid (outer pie), remaining fluid (middle pie) and original fluid (inner pie) showing that a large proportion of the heavier species remain unrecovered. The produced fluid has a molecular weight of 21.1 g/mole which is significantly lower than that of the original reservoir fluid of 80.68 g/mole. The molecular wight of the remaining fluid is 115.5 g/mole. The composition of the produced and remaining fluids are shown in Table 6.1.

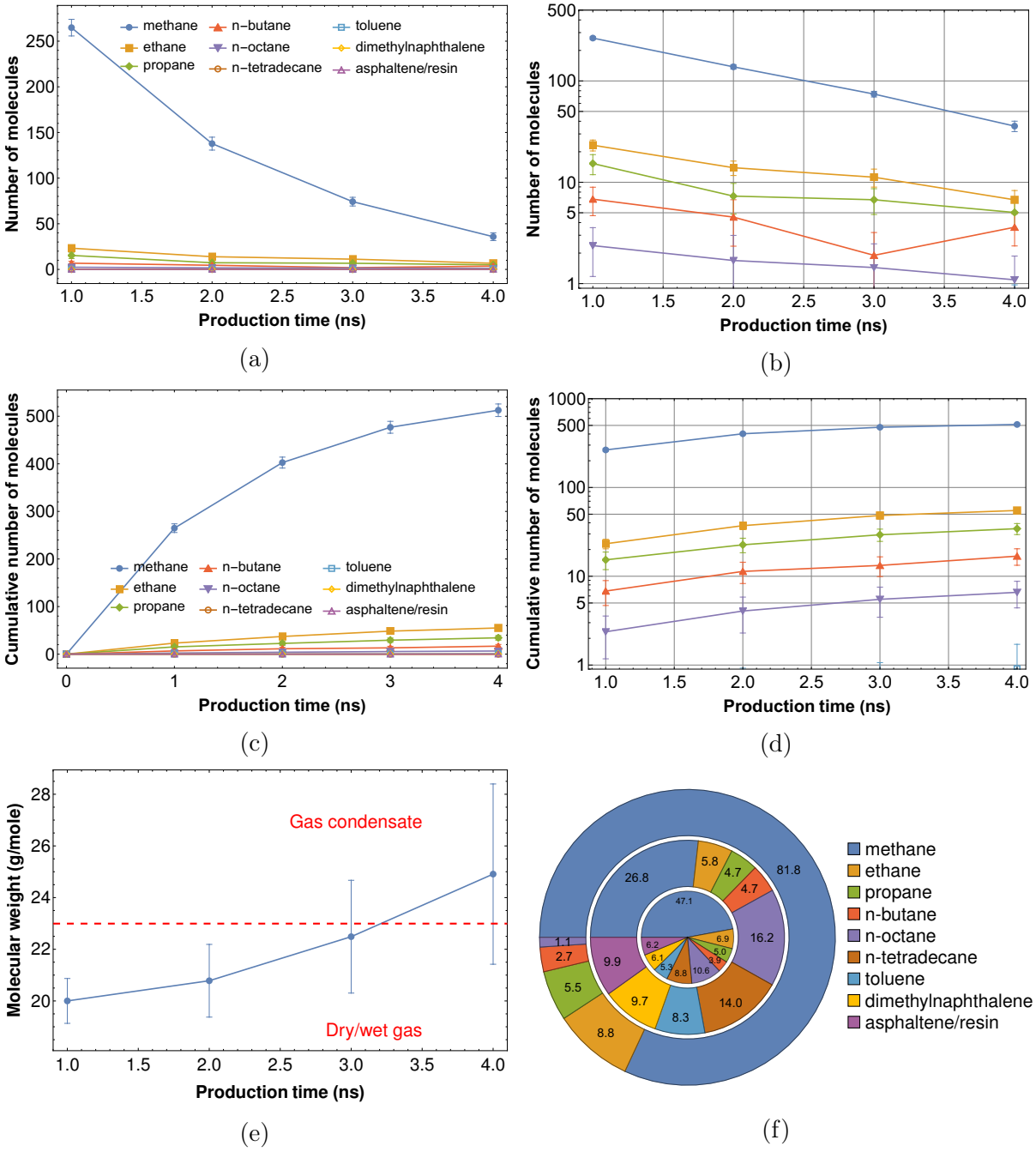


Figure 6.1: Summary of primary depletion for the organic pores initially saturated with a back oil. (a) Instantaneous production of molecules over time; (b) instantaneous production of molecules over time (semilogarithmic scale); (c) cumulative production of molecules over time; (d) cumulative production of molecules over time (semilogarithmic scale); (e) time evolution of the molecular weight of the oil mixture in the microfracture; and (f) comparison between the initial composition of the in-situ fluid (inner pie), the remaining fluid (middle pie) and produced fluid (outer pie) after primary depletion ends.

6.3 Simulation details

The simulations carried out in this part of the work mimic a huff-n-puff process. Huff-n-puff enhanced oil recovery involves injection of a solvent for a specified period of time (days to weeks) followed by a shut in or soak period (days to weeks) and finally the same well is put back on production (weeks to months) (Wang et al., 2017). The purpose of the soak period is to allow the solvent to come into contact with the oil. This recovery method is more difficult to optimize than the conventional solvent injection because of the larger number of parameters, namely solvent composition, injection pressure, injection rate, injection time (huff period), soaking time, and production time (puff period).

After depleting the mature kerogen model saturated with a black oil for 4 ns, a gas mixture composed of methane/ethane (72:28 mole fraction) is injected into the fracture as illustrated in Figure 6.2 and the fluids are allowed to interact for a certain period of time (soaking time) holding temperature T, volume V and number of particles N constant. The composition of the injected gas is comparable to that of field gas produced and is the solvent of choice utilized in experimental work carried out in the Integrated Core Characterization Center (IC^3) at the University of Oklahoma[1]. Then the system is put back on production and the molecules that are found entirely in the microfracture at the end of the soaking period are removed to mimic production from a fracture adjacent to organic matter. Next, additional simulations similar to those described in Section 5.2 are run in steps of 1 ns to mimic production time. The procedure just described is one huff-n-puff cycle.

Then the whole process is repeated one more time to consider two huff-n-puff cycles. The injection pressure of the solvents, which is 300 atm (4400 psi) in both cycles, is chosen to be above the experimentally determined minimum miscibility pressure (MMP) of the oil-solvent system at a temperature of 335 K (180°F)[1]. The composition of the fluids in the microfracture and the penetration of solvent molecules into kerogen are monitored as a function of time. The temperature of the entire system is kept constant at $T = 355$

[1]Courtesy of Sidi Mamoudou, personal communication.

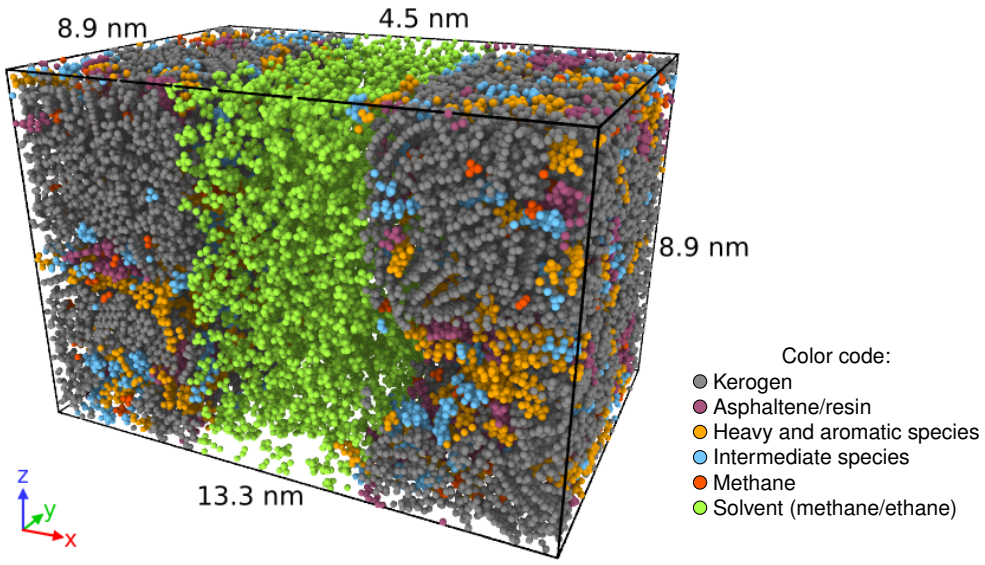


Figure 6.2: Injection of an organic solvent into the microfracture.

K (180°F) but the temperature of the fluid molecules inside the organic pores in kerogen and in the microfracture are controlled independently implementing two thermostats by explicitly rescaling the velocities of the particles in each region, as explained in Section A.5.2. The time step utilized in every simulation is 1 fs.

6.4 Modeling huff-n-puff

The huff-n-puff enhanced oil recovery process simulated in this work comprises two cycles that consider 12 production designs with the aim of evaluating the importance of soaking time and production time in cyclic injection. Moreover, this work investigates the relevance of the residence time concept at molecular scale, which is conveniently defined as the sum of soaking and production time (Dang, 2019)

$$\text{residence time} = \text{soaking time} + \text{production time}, \tag{6.1}$$

thus it accounts for the total interaction time between the solvent and the rock[2]. Figure 6.3 sums up the cases and the respective duration of each step. After primary depletion

[2]The simulations implemented in this work consider injection as an instantaneous process, thus the total interaction time between the fluids in organic pores and the solvent is fully described by the duration of the soaking and production steps. In field applications, the solvent also interacts with the in-situ fluids during the injection period.

for 4 ns, the organic mixture methane/ethane is injected into the fracture as illustrated in Figure 6.2, and the remaining hydrocarbons contained within kerogen are exposed to the solvent in a soak step. Two different soaking periods are considered here, one short (3 ns) and one long (5 ns), followed by four different production stages that span 1 ns, 2 ns, 3 ns, and 4 ns. The end of every production scheme marks the termination of the first cycle, and the entire process is repeated.

The overall process over two cycles can be summarized as follows:

1. Soaking for 3 ns, production for 1 ns, 2 ns, 3 ns, 4 ns; soaking for 3 ns, production for 1 ns, 2 ns, 3 ns, 4 ns.

2. Soaking for 5 ns, production for 1 ns, 2 ns, 3 ns, 4 ns; soaking for 3 ns, production for 1 ns, 2 ns, 3 ns, 4 ns.

3. Soaking for 5 ns, production for 1 ns, 2 ns, 3 ns, 4 ns; soaking for 5 ns, production for 1 ns, 2 ns, 3 ns, 4 ns.

These strategies consider (1) short soaking stages in both cycles, (2) a long soaking period in the first cycle followed by a short soaking period in the second cycle, and (3) long soaking periods in both cycles, and are referred to as short/short (3 ns/3 ns), long/short (5 ns/3 ns), and long/long (5 ns/5 ns), respectively. The production times are kept fixed in both cycles, thus if a system produced for 2 ns in cycle 1, it also produces for 2 ns in cycle 2.

The sections below recount the observations made during each step of one huff-n-puff cycle and the additional oil recovery achieved after two cycles.

6.4.1 Cycle 1

Soaking time

The first step in the cycle is the soaking of the organic matter that hosts the residual fluids. The solvent, initially at high pressure and miscible with the unrecovered fluids, interacts with the hydrocarbon molecules that remain adsorbed on the microfracture faces.

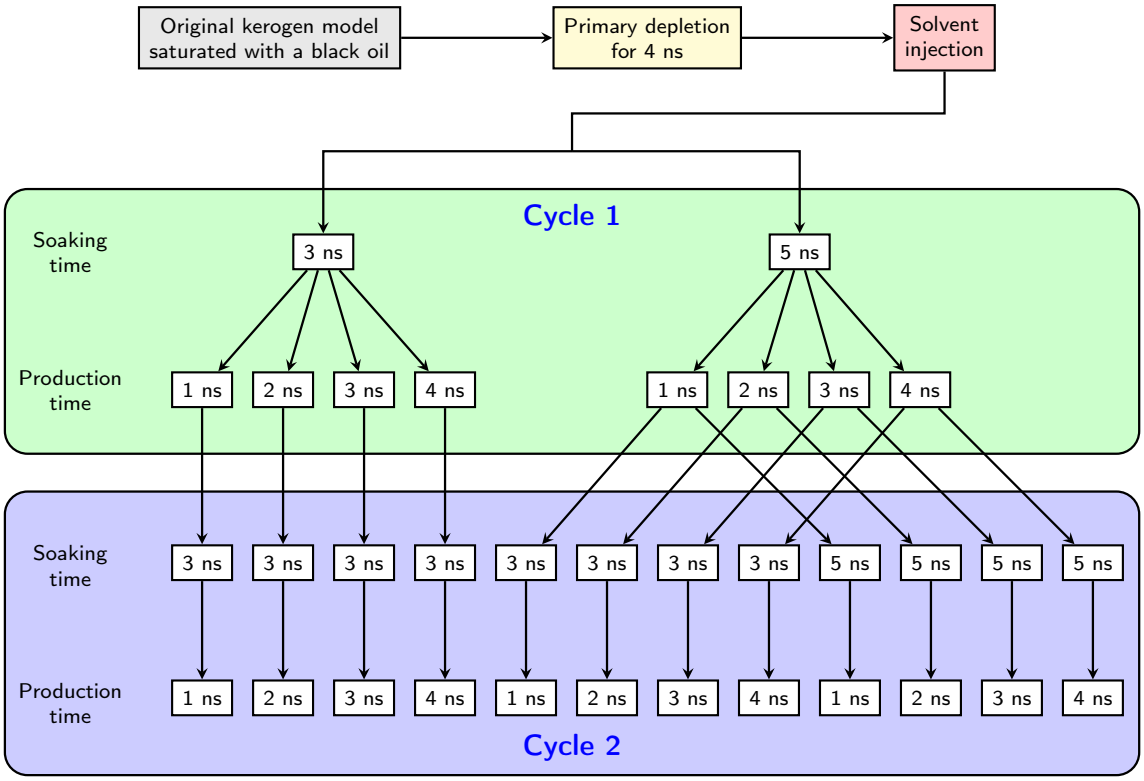
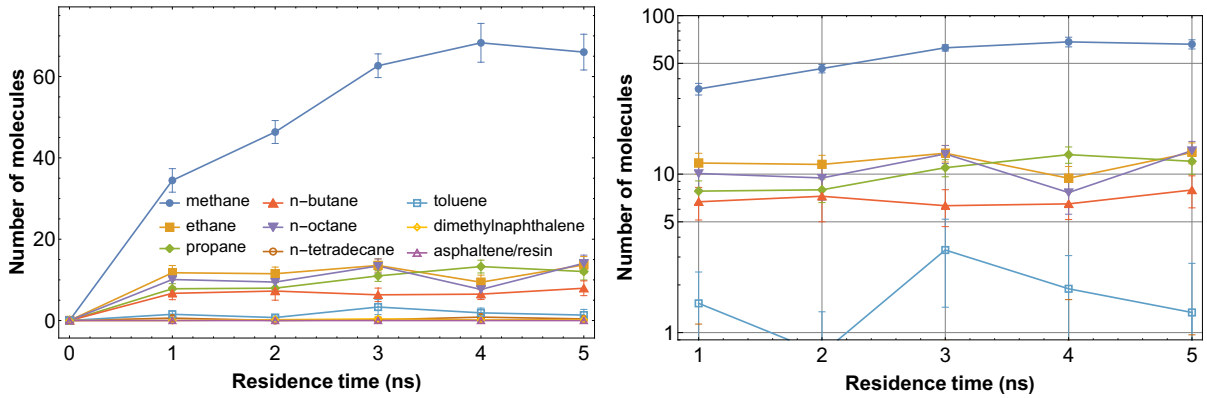


Figure 6.3: Huff-n-puff cases considered.

The solvent also expands rapidly and penetrates the neighboring zone that was depleted during primary production and interacts with fluid molecules there. Furthermore, as soaking time increases, molecules located deeper within kerogen may move to the zone invaded by the solvent and be dissolved in it. During this step the number of fluid molecules of every species in the microfracture is recorded, differentiating between the molecules that are originally part of the in-situ fluid and those belonging to the solvent.

Figure 6.4(a) shows the number of molecules that have migrated from the kerogen to the microfracture during a soaking period of 5 ns. The presence of hydrocarbon species in the microfracture provides evidence of a counterflow from the rock to the fracture network against a pressure gradient. This observation has been also reported by Alharthy et al. (2018), Dang (2019) and Perez and Devegowda (2020c). Since there is a difference of one order of magnitude between the number of molecules of methane and the other species, the same information is presented in semilogarithmic scale in Figure 6.4(b). Note the increased presence of the species n-octane and toluene in the microfracture compared to the amount of these species produced during primary depletion (see Figure 6.1). The



(a) (b)

Figure 6.4: Number of molecules of in-situ fluids in the microfracture during a soaking period of 5 ns in cycle 1. (a) Linear scale; (b) semilogarithmic scale. The residence time coincides with the soak time in this case.

presence of the solvent does enhance the ability of these molecules to migrate to the microfracture and remain in solution, as will be discussed in Section 7.3.2.

Furthermore, the presence of heavier molecules in the microfracture makes the oil mixture heavier than that produced during primary depletion. This is illustrated in Figure 6.5, which shows the molecular weight of the oil mixture dissolved in the organic solvent in the microfracture[3]. The figure suggests two important things: a heavier oil mixture may be potentially produced once the system is put back on production, and as the soaking time increases, the oil mixture dissolved in the microfracture becomes lighter, probably because of the migration of lighter species from deeper pores in kerogen to the microfracture and signal the existence of the counterflow of hydrocarbons against the pressure gradient.

Production time

After the soaking period ends, the system is put back on production and the fluid molecules that are entirely in the microfracture are removed from the simulation box to mimic production from a high permeability fracture adjacent to organic matter as explained in Section 5.2. Next, extra production steps are simulated to achieve the production times listed in Figure 6.3. Production after two different soaking times of 3

[3]The composition of the oil mixture does not take into account the methane and ethane molecules of the solvent, thus the oil mixture is the solute dissolved in the solvent.

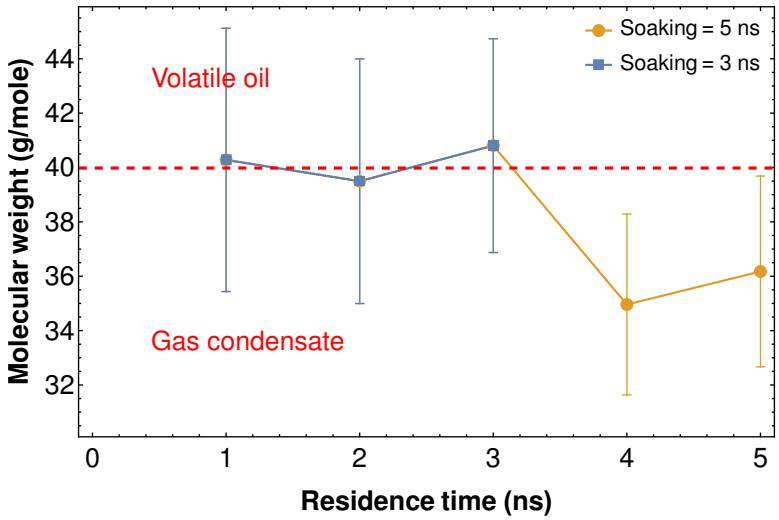
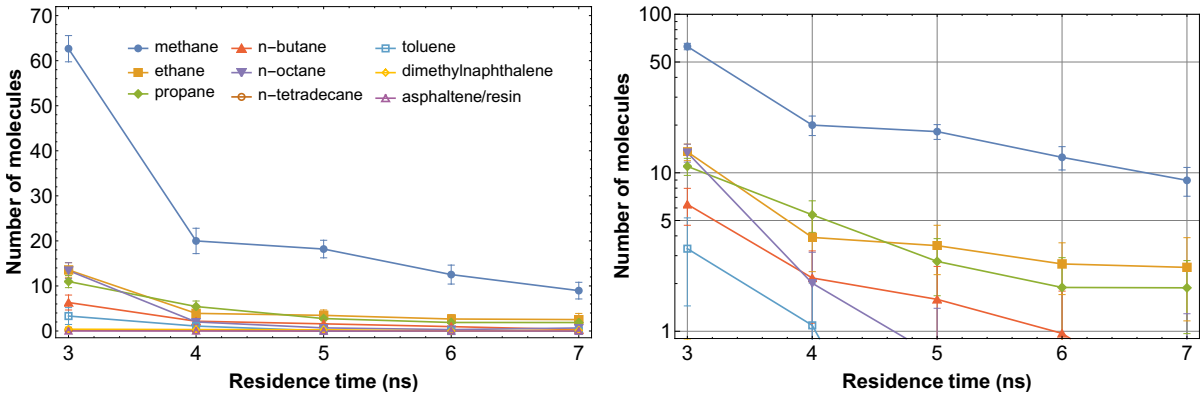


Figure 6.5: Molecular weight of the oil mixture in the microfracture during the soaking period in cycle 1. For reference, the molecular weight of the oil produced during primary depletion is 21.1 g/mole, out of the vertical range of the figure. The residence time coincides with the soak time in this case.

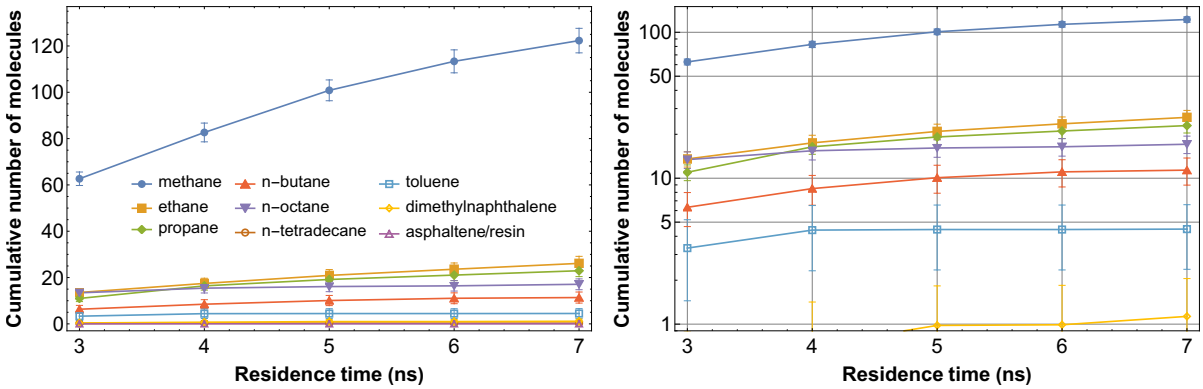


(a) (b)

Figure 6.6: Instantaneous oil production in cycle 1 after soaking for 3 ns. (a) Linear scale; (b) semilogarithmic scale.

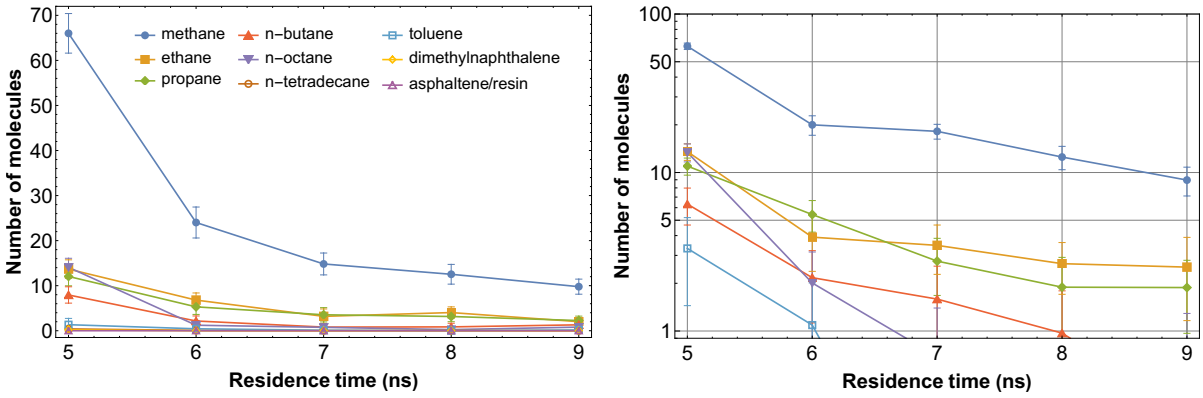
ns and 5 ns are considered, and the results are discussed below.

In the case of a short soaking time of 3 ns, the fluid that has migrated to the microfracture corresponds to a light volatile oil and is produced right after the system is put back on production. The molecular weight of this oil is shown in Figure 6.5. Figure 6.6(a) shows the number of molecules of every species recovered as a function of residence time in the case of production for 4 ns. Since there is a difference of one order of magnitude between the number of molecules of methane and the other species, the same information is presented in semilogarithmic scale in Figure 6.6(b). What is produced next depends on the additional production time, therefore on the total residence time.



(a) (b)

Figure 6.7: Cumulative oil production in cycle 1 after soaking for 3 ns. (a) Linear scale; (b) semilogarithmic scale.



(a) (b)

Figure 6.8: Instantaneous oil production in cycle 1 after soaking for 5 ns. (a) Linear scale; (b) semilogarithmic scale.

Figure 6.7(a) depicts the cumulative production as a function of residence time. Due to the difference of one or two orders of magnitude between the number of molecules of methane and the other species, the same information is presented in semilogarithmic scale in Figure 6.7(b).

Production of the lighter molecules can be sustained for longer production times (i.e. longer residence times). This is because the lighter molecules constitute the free fluid phase, which also has a higher compressibility. Additionally, there is a continuous recharge from within kerogen towards the microfracture. The intermediate species are produced only initially just immediately following the introduction of the solvent as seen in Figure 6.6, indicating that those species are produced in solution when the concentration of solvent is high.

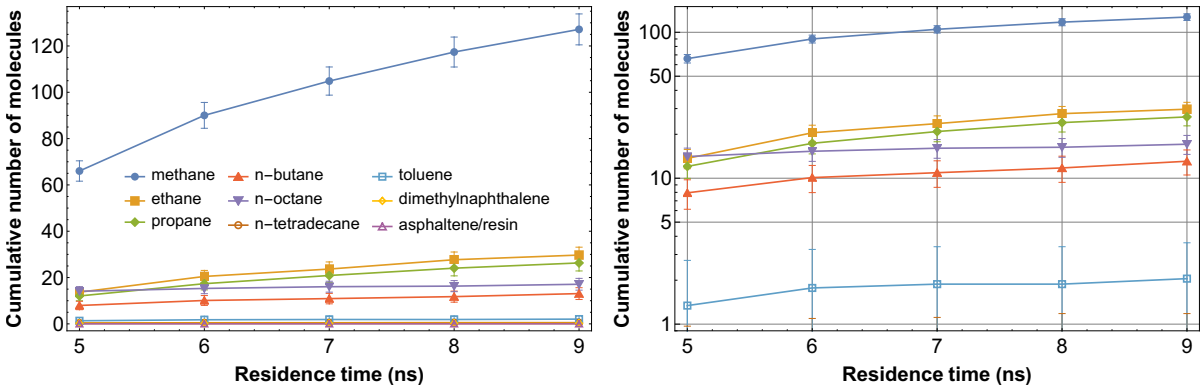


Figure 6.9: Cumulative oil production in cycle 1 after soaking for 5 ns. (a) Linear scale; (b) semilogarithmic scale.

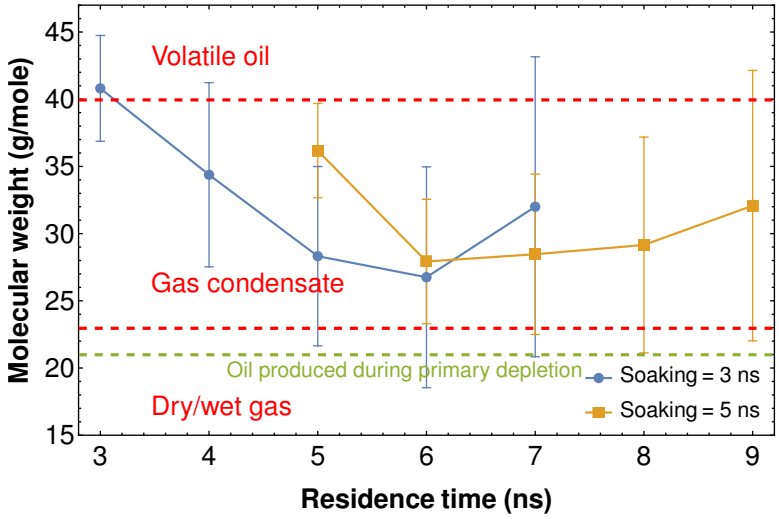


Figure 6.10: Time evolution of the molecular weight of the oil mixtures produced in the first cycle. The dashed-green line displays the molecular weight of the oil mixture produced during primary depletion.

In the case of a long soaking time of 5 ns, Figure 6.5 shows that the oil mixture in the microfracture becomes marginally lighter for a longer soak time. This is because of the displacement and subsequent mixing of additional lighter fluid molecules from within kerogen. Figures 6.8 and 6.9 present respectively the instantaneous and cumulative production for 4 ns after soaking for 5 ns.

The time evolution of the molecular weight of the oil mixtures produced during 4 ns after different soaking times is given in Figure 6.10 as a function of residence time. With a short soak time of 3 ns, the molecular weight of the initial produced fluid is 41 g/mole and then progressively decreases with increasing production time until it stabilizes. The

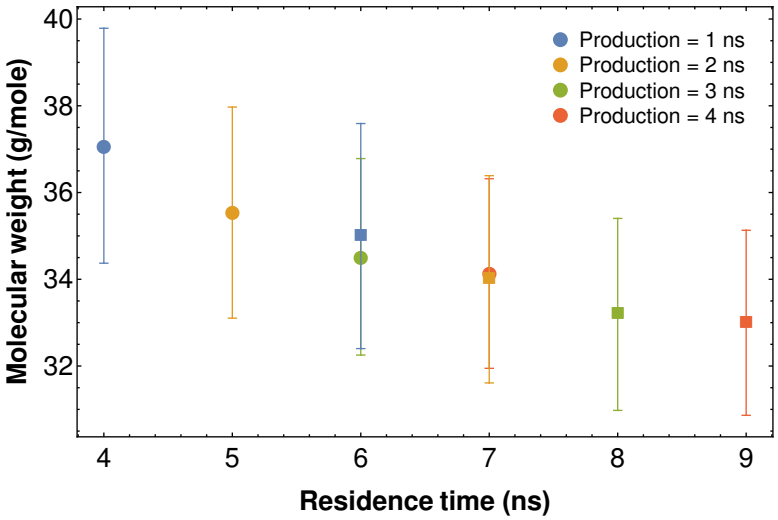


Figure 6.11: Molecular weight of the oil mixtures produced in the first cycle. Circles denote a soaking time of 3 ns; squares denote a soaking time of 5 ns; colors denote different production times. For reference, the molecular weight of the oil produced during primary depletion is 21.1 g/mole, out of the vertical range of the figure.

reasons are the recharge of the lighter species as well as the relatively low mobility of the intermediate species. A similar trend is observed for a longer soak time of 5 ns. This is a level of detail achieved with molecular dynamics simulations. In order to mimic experimental measurements, Figure 6.11 displays the molecular weight of the recovered oil at the end of cycle 1 for every production scheme. It is observed that as residence time increases, the produced mixture becomes lighter, in accordance with the discussions just presented, but nonetheless heavier than the oil produced during primary depletion shown in Figure 6.1(e).

To finalize the analysis of cycle 1, the additional oil recovery accomplished at the end of the first huff-n-puff cycle for every production scheme is presented in Figure 6.12. The recovery is computed with respect to the original black oil present in organic pores. Thus, at the end of the first cycle with a production time of 4 ns, an extra recovery of approximately 5% (i.e. in addition to the 9% recovery achieved after primary depletion, thus the total oil recovery up to this point is 14%) is accomplished by the cyclic injection of a miscible organic solvent. In comparison with Figure 5.14, if primary depletion had continued for the entire range of 9 ns, the final recovery would have been about 11%. However, it is important to note that the produced fluid in the context of enhanced oil recovery is significantly different from that during primary production and arguably, of

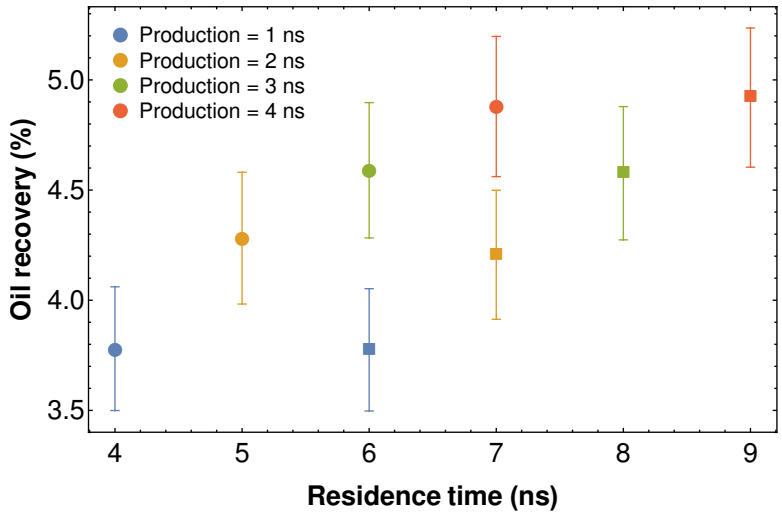


Figure 6.12: Oil recovery during the first cycle in a huff-n-puff process. Circles denote a soaking time of 3 ns; squares denote a soaking time of 5 ns; colors denote different production times.

more value as well because of the increased yield in the intermediate components.

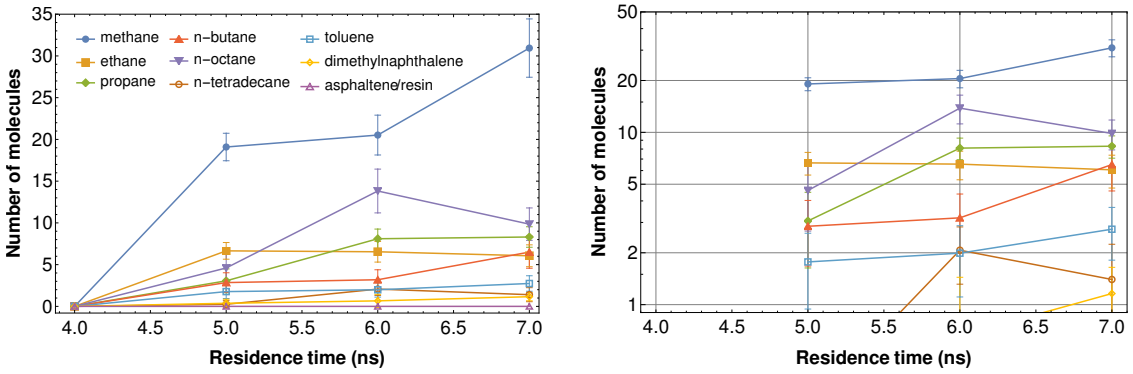
The arguments above indicate that the implementation of just one huff-n-puff cycle improves oil recovery, but no evident technical observations regarding the design of the cycle can be drawn at this point; similar oil recoveries and fluid compositions are achieved regardless of the soaking time as long as the production time is the same. The execution of a second cycle will help draw more general and robust conclusions.

6.4.2 Cycle 2

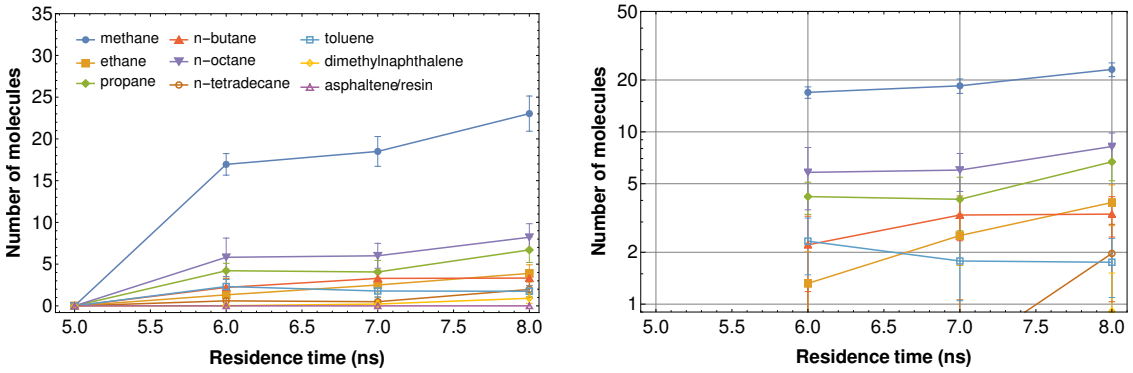
Following the first cycle of production, the solvent is injected again into the microfracture at a high pressure of 300 atm (4400 psi) and the overall process is repeated. The three different soaking schemes (1) short/short, (2) long/short, and (3) long/long are possible now (see Figure 6.3). This approach provides the means to compare different huff-n-puff design parameters.

Soaking time (short/short – 3 ns/3 ns)

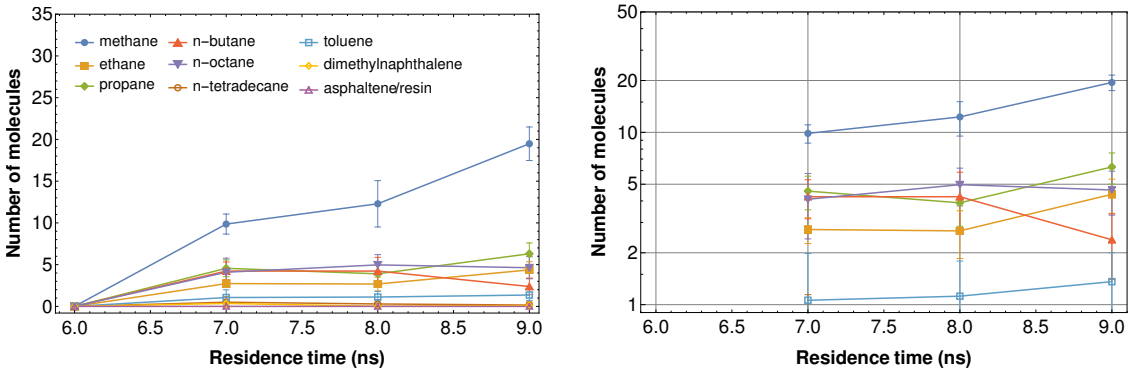
In this part of the study, the first soak cycle of 3 ns is followed with an initial production time of 1 ns, 2 ns, 3 ns, or 4 ns, then each case is followed up by a soak time of 3 ns (see Figure 6.3). The number and type of molecules that are dissolved in the microfracture are recorded as a function of time for all cases and the results are



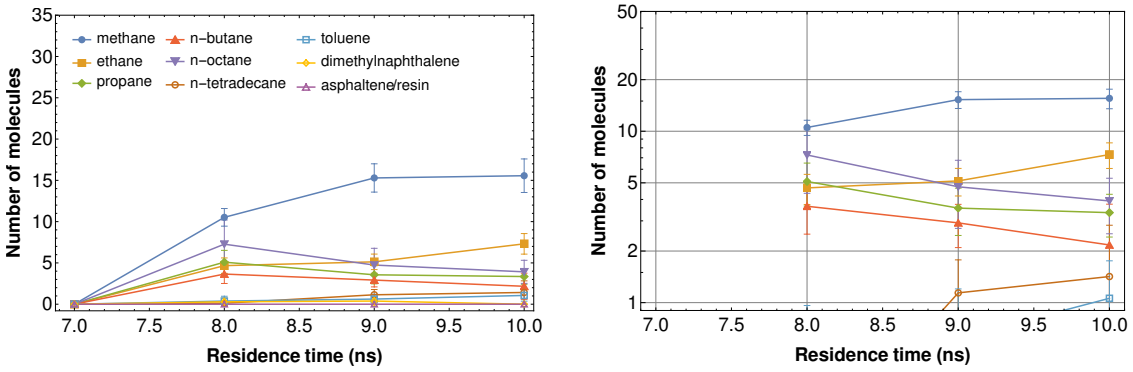
(a) After production for 1 ns



(b) After production for 2 ns



(c) After production for 3 ns



(d) After production for 4 ns

Figure 6.13: Number of molecules of in-situ fluids in the microfracture during the soaking period in cycle 2 (for the short/short case). The figures on the left are linear, the ones on the right present the same information in semilogarithmic scale. The range in the vertical axes are the same throughout the figures to allow direct comparisons.

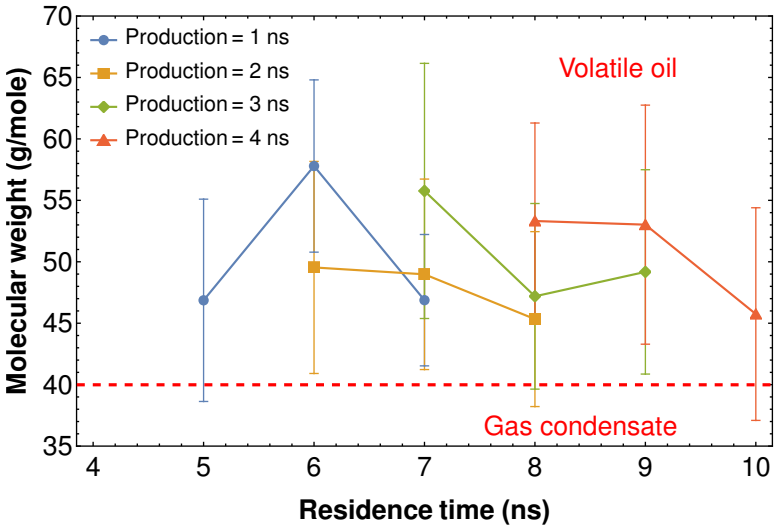


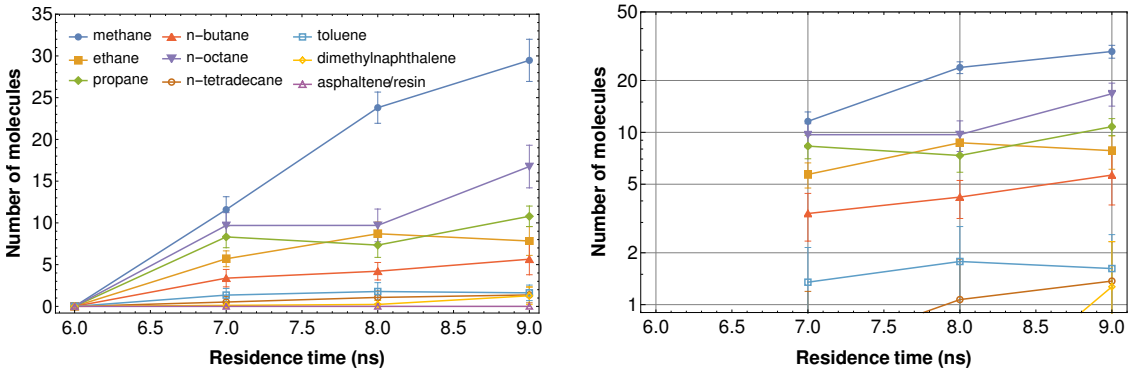
Figure 6.14: Molecular weight of the oil mixture in the microfracture during the soaking period in cycle 2 (for the short/short case). The colors denote different production times in cycle 1.

summarized in Figure 6.13.

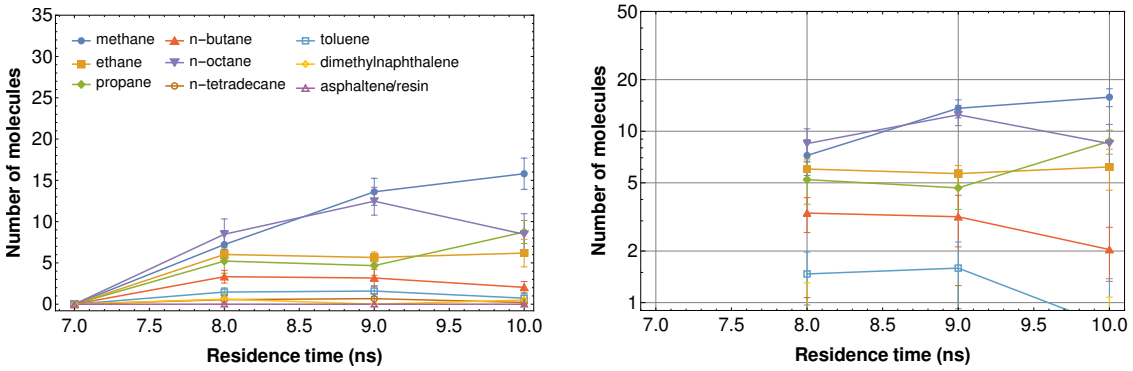
Notice that the fluid molecules that arrive at the microfracture are moving against a pressure gradient and in the same direction as a concentration gradient during the soak period. An evident observation is that the number of methane molecules arriving at the microfracture decreases as production time during cycle 1 increases. The reason is the increased depletion with increased production time in cycle 1. The molecular weight of the oil mixture in the microfracture is displayed in Figure 6.14, distinguishing the mixtures according to the production time during cycle 1. The different cases yield mixtures with a similar molecular weight, corresponding to a volatile oil. A comparison with the molecular weight of the mixtures obtained during soaking in the first cycle in Figure 6.5 shows that contact with the solvent mobilizes a heavier oil in cycle 2.

Soaking time (long/short – 5 ns/3 ns)

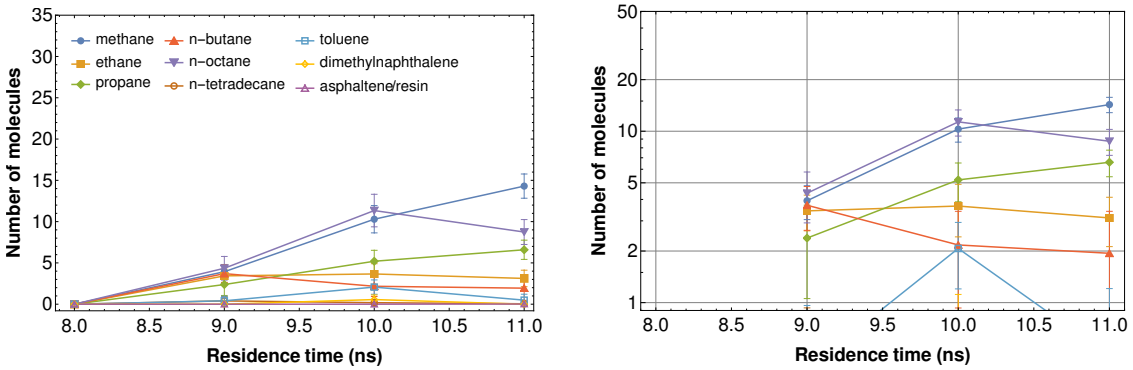
In this part of the study, the first soak cycle of 5 ns is followed with an initial production time of 1 ns, 2 ns, 3 ns, or 4 ns, then each case is followed up by a soak time of 3 ns (see Figure 6.3). The number and type of molecules that are dissolved in the microfracture are recorded as a function of time for all cases and the results are summarized in Figure 6.15. As with the previous case, an increased production time in



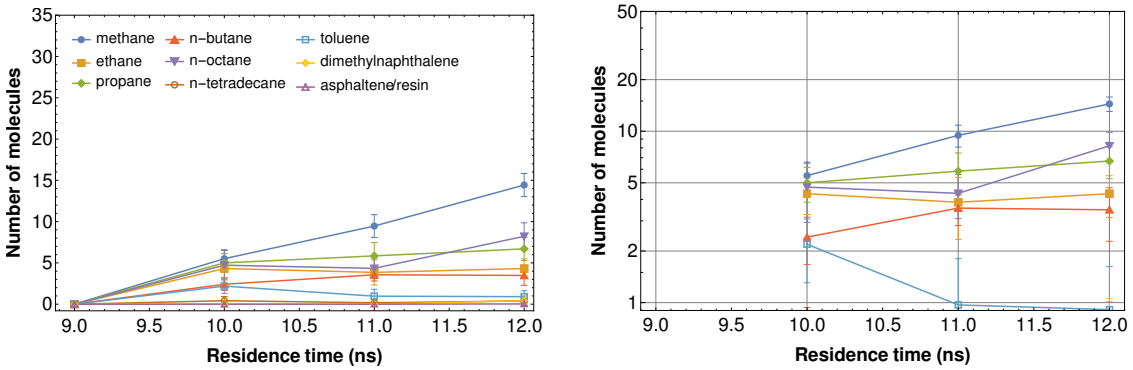
(a) After production for 1 ns



(b) After production for 2 ns



(c) After production for 3 ns



(d) After production for 4 ns

Figure 6.15: Number of molecules of in-situ fluids in the microfracture during the soaking period in cycle 2 (for the long/short case). The figures on the left are linear, the ones on the right present the same information in semilogarithmic scale. The range in the vertical axes are the same throughout the figures to allow direct comparisons.

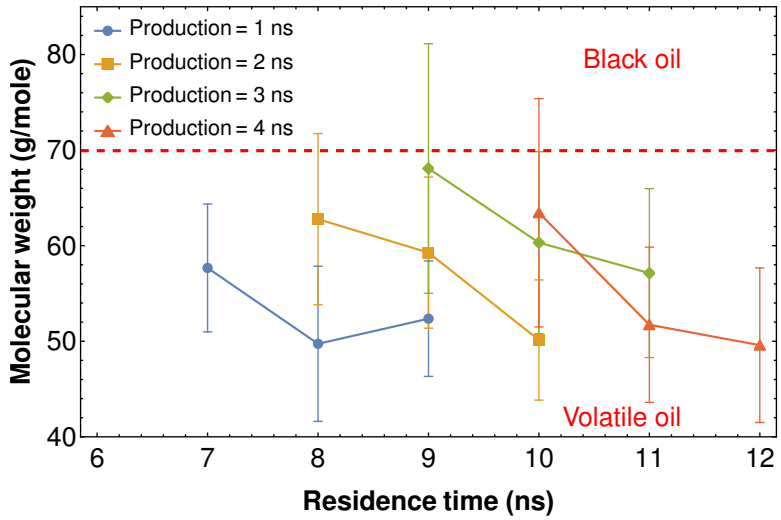
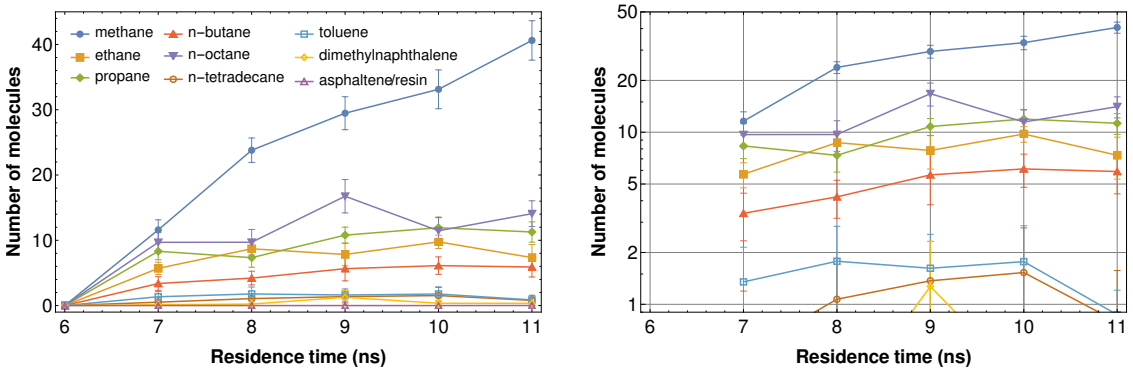


Figure 6.16: Molecular weight of the oil mixture in the microfracture during the soaking period in cycle 2 (for the long/short case). The colors denote different production times in cycle 1.

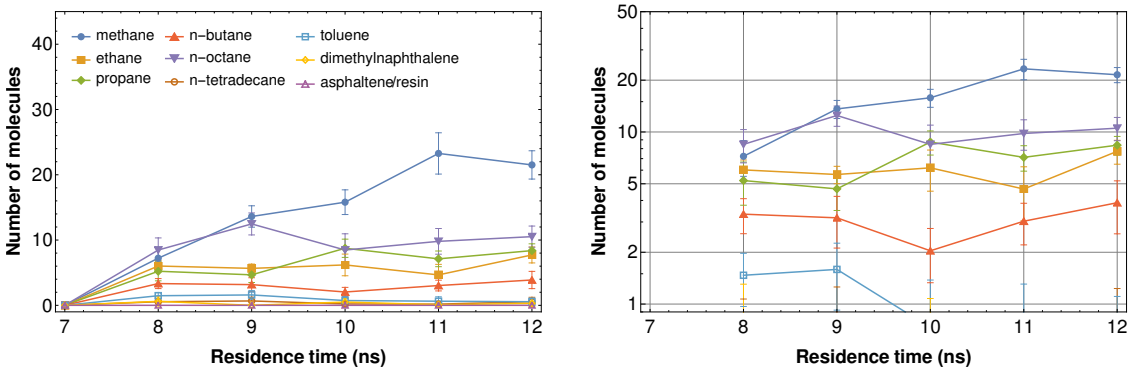
cycle 1 decreases the amount of methane produced in cycle 2. The molecular weight of the oil mixture in the microfracture is displayed in Figure 6.16, distinguishing the mixtures according to the production time during cycle 1. The different cases yield mixtures with a similar molecular weight, corresponding to a volatile oil, or perhaps even a black oil. Compared to the molecular weights of the fluids in the microfracture during the soak period in cycle 1, the mixtures in cycle 2 are heavier.

Soaking time (long/long – 5 ns/5 ns)

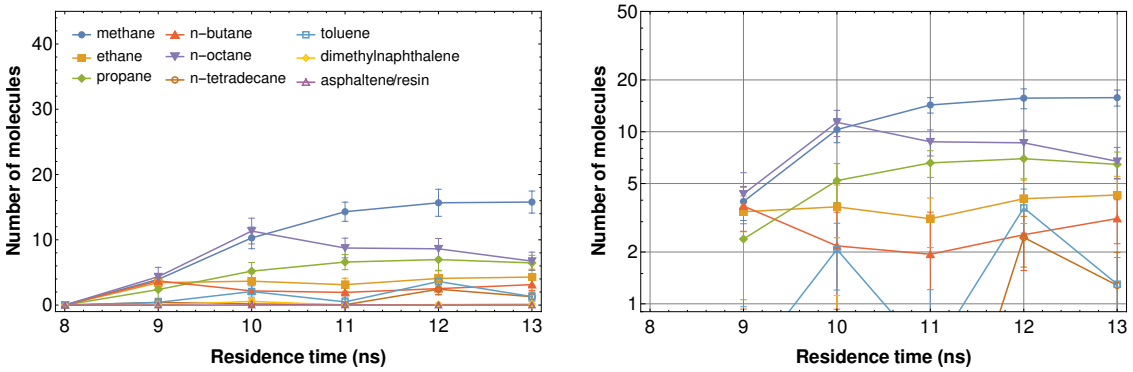
In this part of the study, the first soak cycle of 5 ns is followed with an initial production time of 1 ns, 2 ns, 3 ns, or 4 ns, then each case is followed up by a soak time of 5 ns (see Figure 6.3). The number and type of molecules that are dissolved in the microfracture are recorded as a function of time for all cases and the results are summarized in Figure 6.17. As with the previous cases, an increased production time in cycle 1 decreases the amount of methane produced in cycle 2. The molecular weight of the oil mixture in the microfracture is displayed in Figure 6.18, distinguishing the mixtures according to the production time during cycle 1. The different cases yield mixtures with a similar molecular weight, corresponding to a volatile oil, or perhaps even a black oil, as in the long/short case. Compared to the molecular weights of the fluids in the microfracture during the soak period in cycle 1, heavier mixtures are obtained in cycle 2.



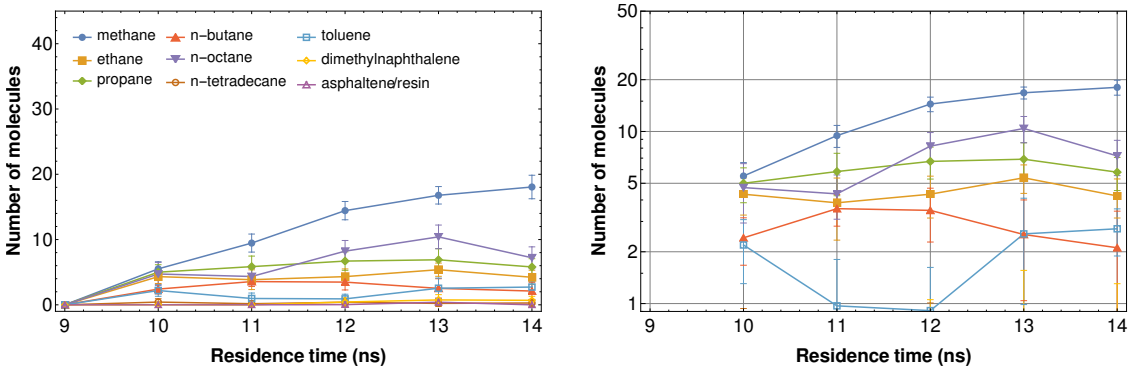
(a) After production for 1 ns



(b) After production for 2 ns



(c) After production for 3 ns



(d) After production for 4 ns

Figure 6.17: Number of molecules of in-situ fluids in the microfracture during the soaking period in cycle 2 (for the long/long case). The figures on the left are linear, the ones on the right present the same information in semilogarithmic scale. The range in the vertical axes are the same throughout the figures to allow direct comparisons.

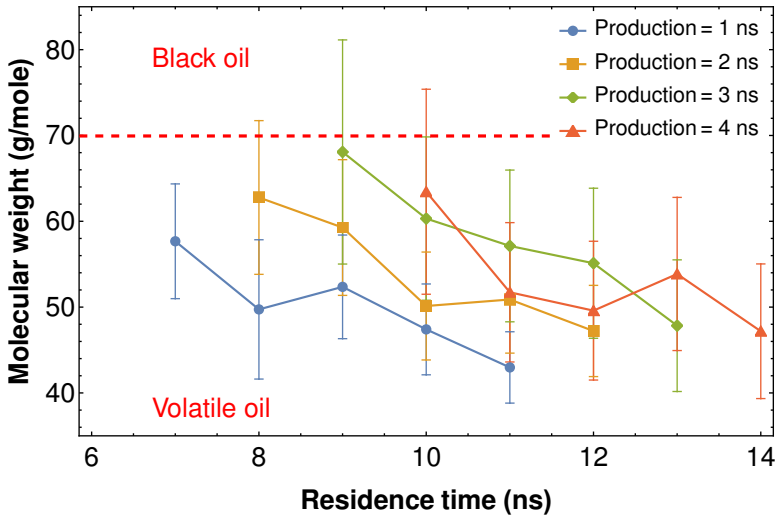


Figure 6.18: Molecular weight of the oil mixture in the microfracture during the soaking period in cycle 2 (for the long/long case). The colors denote different production times in cycle 1.

Soaking time (summary)

Although soaking is not a stage of oil production, the solvent and the in-situ fluids interact so that as the solvent penetrates into kerogen, there is a counterflow of fluid molecules from organic pores to the microfracture. This dissolved oil mixture is heavier than the oil produced during the first cycle and its molecular weight corresponds to a volatile oil or a black oil (see Figures 6.14, 6.16 and 6.18). The oil mixture that remains in solution in the microfracture will be recovered once the system is put back on production again. Therefore, it can be said that the solvent does target, mix and dissolve intermediate molecular species that were not produced during the first huff-n-puff cycle. This level of detail during the soaking stage, however, is not currently achieved experimentally, thus molecular simulations provide microscopic insights into solvent-oil interactions in shales (at least in the organic portion) during soaking.

Production time (short/short – 3 ns/3 ns)

In this part of the analysis, the effect of production time is evaluated across all cycle 1 cases and cycle 2 soak times. First, consider the case of 3 ns soak in cycle 1 and 3 ns soak in cycle 2. Recall that production times in cycle 1 were 1 ns, 2 ns, 3 ns, or 4 ns and production in times cycle 2 are also 1 ns, 2 ns, 3 ns, or 4 ns. The total oil recovery is shown

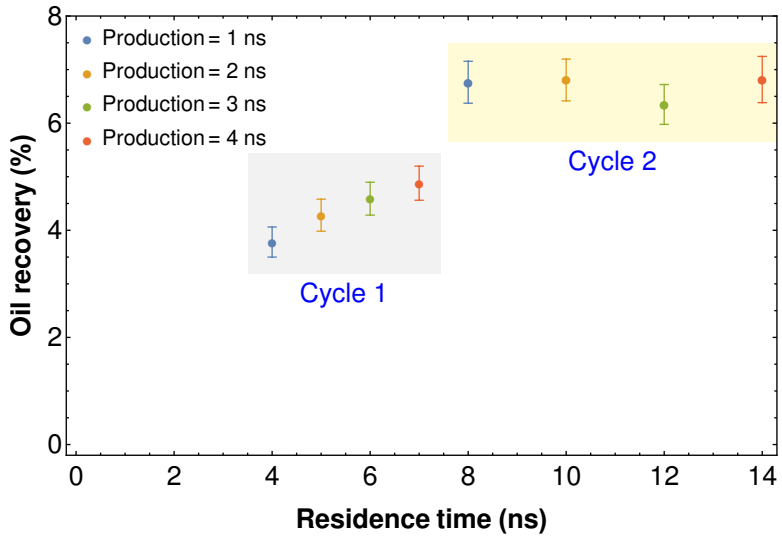


Figure 6.19: Total oil recovery after the second cycle in a huff-n-puff process (for the short/short case). The colors denote different production times. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

in Figure 6.19 as a function of residence time. The important observation here is that a second huff-n-puff cycle provides additional recovery; however, increased production time does not lead to a higher recovery.

The molecular weight of the produced fluid as a function of residence time presented in Figure 6.20 shows that most of the produced fluids can be classified as a rich gas condensate or a light volatile oil. Again, no additional benefit from an increased production time is seen either in terms of oil recovery or in terms of the composition of the produced fluids reflected by the molecular weight. The slight decreasing trend in molecular weight is due to the dilution of the produced oil with lighter species that migrate to the microfracture from deeper within the kerogen as production times in cycle 2 increases.

Production time (long/short – 5 ns/3 ns)

Similarly, this part of the study evaluates the effect of production time across all cycle 1 cases and cycle 2 soak times. Consider now the case of 5 ns soak in cycle 1 and 3 ns soak in cycle 2. Recall that production times in cycle 1 were 1 ns, 2 ns, 3 ns, or 4 ns and production in times cycle 2 are also 1 ns, 2 ns, 3 ns, or 4 ns. The total oil recovery is shown in Figure 6.21 as a function of residence time. As in the previous case, the important remark here is that a second huff-n-puff cycle does provide additional

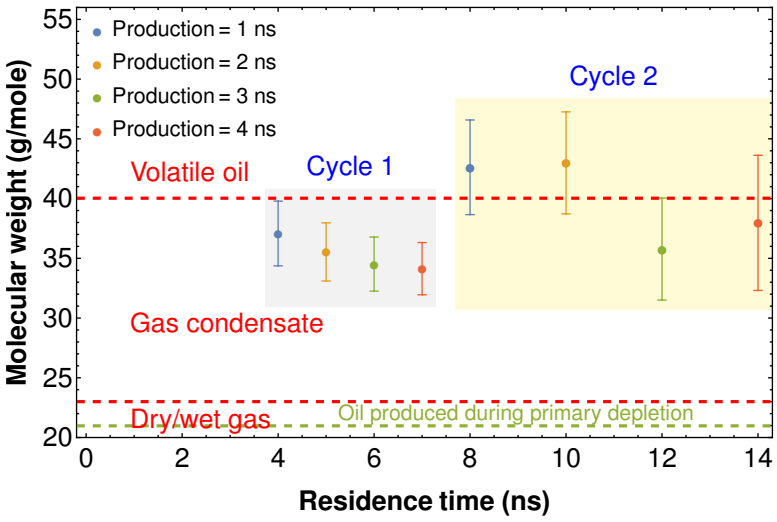


Figure 6.20: Molecular weight of the oil mixtures produced in a huff-n-puff process (for the short/short case). The colors denote different production times. The dashed-green line displays the molecular weight of the oil mixture produced during primary depletion. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

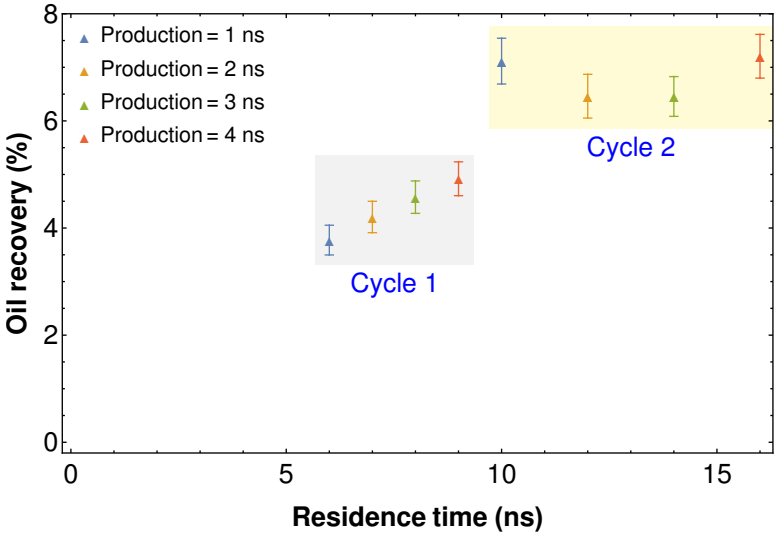


Figure 6.21: Total oil recovery after the second cycle in a huff-n-puff process (for the long/short case). The colors denote different production times. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

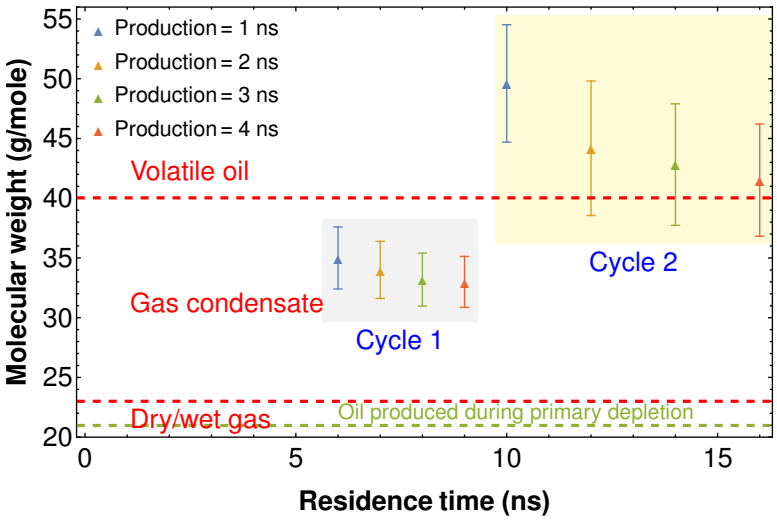


Figure 6.22: Molecular weight of the oil mixtures produced in a huff-n-puff process (for the long/short case). The colors denote different production times. The dashed-green line displays the molecular weight of the oil mixture produced during primary depletion. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

recovery, although increased production time does not lead to a higher recovery. The total recovery in this case is statistically similar to that achieved in the short/short case.

The molecular weight of the produced fluid as a function of residence time presented in Figure 6.22 shows that most of the produced fluids can be classified as a volatile oil or a rich gas condensate. Again, no additional benefit from an increased production time is seen either in terms of oil recovery or in terms of the composition of the produced fluids reflected by the molecular weight. However it is interesting to notice that the produced fluid in cycle 2 is largely a volatile oil regardless of the production time length. This is important and shows that an initial long soak cycle followed by a shorter soak cycle can mobilize heavier mixtures in cycle 2. The early longer soak time in cycle 1, counterintuitively, led to recovery of more of the lighter species during the production step in cycle 1. This is seen by comparing the molecular weight of the produced fluids in cycle 1 in Figure 6.11.

Production time (long/long – 5 ns/5 ns)

Finally, this part of the analysis evaluates the effect of production time across all cycle 1 cases and cycle 2 soak times. Consider now the case of 5 ns soak in cycle 1 and 5 ns soak in cycle 2. Recall that production times in cycle 1 were 1 ns, 2 ns, 3 ns, or 4 ns

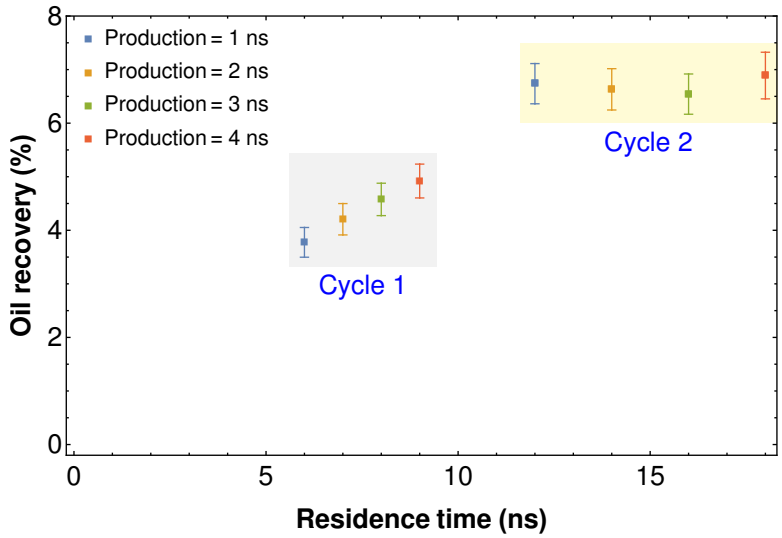


Figure 6.23: Total oil recovery after the second cycle in a huff-n-puff process (for the long/long case). The colors denote different production times. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

and production in times cycle 2 are also 1 ns, 2 ns, 3 ns, or 4 ns. The total oil recovery is shown in Figure 6.23 as a function of residence time. As in the previous cases, the important remark here is that a second huff-n-puff cycle does provide additional recovery, although increased production time does not lead to a higher recovery. The total recovery in this case is statistically similar to those achieved in the previous cases.

The molecular weight of the produced fluid as a function of residence time presented in Figure 6.24 shows that most of the produced fluids can be classified as a rich gas condensate or a light volatile oil. Again, no additional benefit from an increased production time is seen either in terms of oil recovery or in terms of the composition of the produced fluids reflected by the molecular weight. The slight decreasing trend in molecular weight is due to the dilution of the produced oil with lighter species that migrate to the microfracture from deeper within the kerogen as production times in cycle 2 increases.

Production time (summary)

Based on the discussions presented, it is seen that the final oil recovery does not seem to depend on the total time the system interacts with the solvent. Similar oil recoveries are seen in Figures 6.19, 6.21 and 6.23. However, the molecular weight of the recovered mixtures are different. Figures 6.20, 6.22 and 6.24 indicate that heavier mixtures are

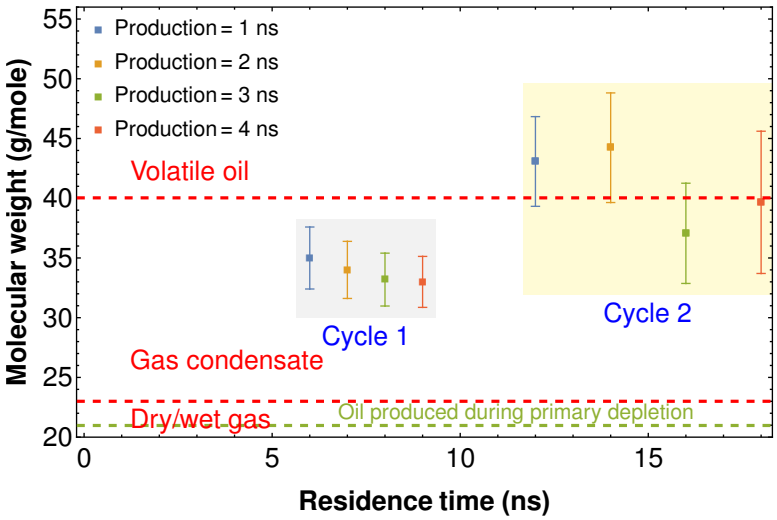


Figure 6.24: Molecular weight of the oil mixtures produced in a huff-n-puff process (for the long/long case). The colors denote different production times. The dashed-green line displays the molecular weight of the oil mixture produced during primary depletion. The light-gray and light-yellow boxes indicate cycles 1 and 2, respectively.

produced during the second cycle when longer soaking and production periods are used in the first cycle. Moreover, the use of a short soaking time in the second cycle can sustain recovery of heavier oil mixtures for an extended production period, whereas a longer soaking time in the second cycle does not seem to perform as well. Therefore, there must be a time after which longer soaking in later cycles becomes detrimental to the weight of the mixture produced.

The optimum huff-n-puff design is discussed in the next section, using the evidences recollected from molecular simulations.

6.5 The optimum huff-n-puff design from molecular simulations

To compare the performance of the different huff-n-puff design strategies, the oil recovered and the respective molecular weight for each case are plotted together in Figures 6.25 and 6.26, respectively, as a function of the residence time. Notice in Figure 6.25 that the recovery during the first cycle increases as residence time increases, but longer production times (i.e. residence times) during the second cycle do not provide additional

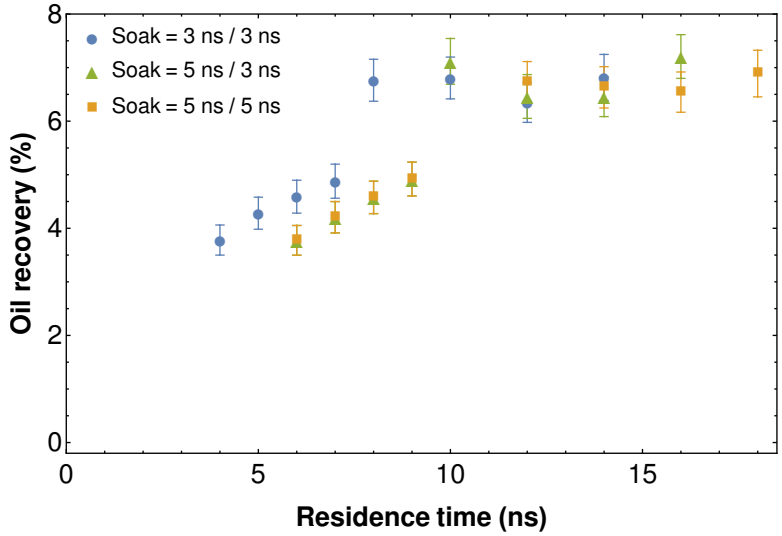


Figure 6.25: Total oil recovery after the second cycle in a huff-n-puff process.

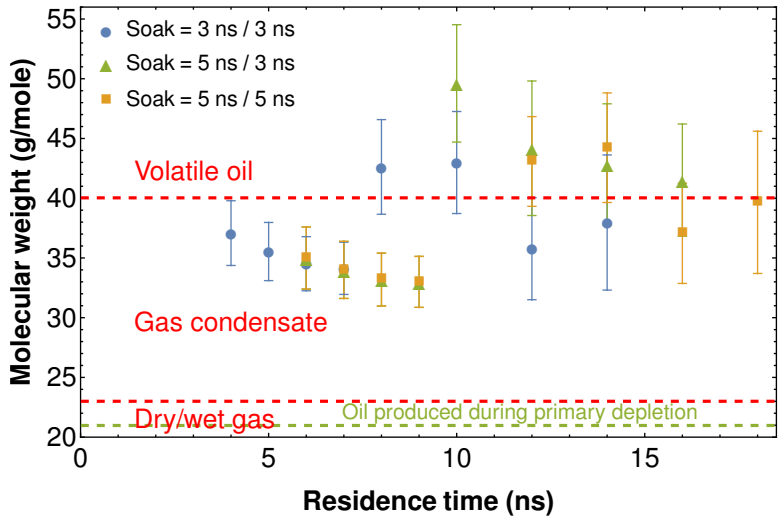


Figure 6.26: Molecular weight of the oil mixtures produced in a huff-n-puff process. The dashed-green line displays the molecular weight of the oil mixture produced during primary depletion.

benefits. The final oil recovery flattens out after some time. Interestingly, the recovery for the different cycle designs all follow one general trend when plotted in terms of residence time, showing therefore a strong dependence on the residence time. This same trend has been also reported experimentally using preserved Eagle Ford crushed samples (7-8 mm) and horizontal plugs (1 in long, 1.5 in diameter) and the same injection gas (Mamoudou et al., 2020), which proves that the molecular system created in this work captures a good portion of the physical phenomenology that occurs in shale rocks despite its small size and the assumptions made.

The mere use of final oil recovery is not enough to decide what the optimum huff-

n-puff design should be like. It seems clear that after some value of residence time oil recovery does not improve, which means the prolonged exposure of the rock to the solvent does not yield more oil. The use of the time evolution of the weight of the hydrocarbon mixture recovered gives insights into the proper break down of the cycles into optimum soaking and production times.

Figure 6.26 indicates that a short soaking time and short production times during the first cycle yield heavier oil mixtures than a long soaking time and long production times. However, this strategy does not perform well during the second cycle because the oil becomes light rather quickly which might be undesirable in terms of economics. Allowing a long exposure of the rock to the solvent and long production times at early cycles help produce the small species that can migrate more easily from deeper in the rock to recharge the depleted zone created in primary depletion (or in previous huff-n-puff cycles), so that during later cycles the solvent more efficiently targets intermediate and heavier species still adsorbed on the pore surfaces in such depleted region, interacts with the remaining fluids in the vicinity of the fractures, and desorbs and dissolves a fraction of those molecular species. This is precisely what is seen during the second cycle in Figure 6.26. Taking also into account that the total oil recovery does increase with residence time, the use of long soaking intervals is not recommended in later cycles to avoid interactions between the solvent and the light species that may desorb and migrate to recharge the invaded zone, probably saturating the solvent and preventing heavier molecules from remaining in solution. This is the reason for the existence of a time after which rock soaking becomes detrimental to the molecular weight of the oil recovered.

Thus, it is concluded that long soaking and production times at early cycles followed by short soaking periods and long production times in later cycles is the optimum manner to design huff-n-puff projects. This conclusion has been also reported using pilot wells in the Eagle Ford (Hoffman, 2018). The molecular system and huff-n-puff modeling done in this work provide a physical ground to the optimum field choices derived from pilot wells.

6.6 Final remarks

The implementation of cyclic injection of an organic solvent improves oil recovery. This has been reported experimentally and in field operations. The simulations carried out in this chapter also prove that point with the additional advantage that they provide a level of detail achieved neither in experiments nor in field operations. The ability of the simulations to reproduce the trends observed experimentally and to infer the overall conclusions derived using pilot wells indicate that although the model created and methodology followed are simplistic, the system captures an important portion of the physics that occurs in shale rocks (or at least in the organic portion of them). This underscores the potential of molecular simulations to provide insights into the phenomenology of fluid flow through organic nanopores, crucial to develop better practices in oil exploitation from unconventional reservoirs.

The discussions in this chapter claim the ability to model huff-n-puff enhanced oil recovery using molecular simulations without showing the molecular picture of the process. Now that it has been argued that the results follow the trends observed experimentally, the next chapter shows the molecular signatures of huff-n-puff. In other words, the next chapter will explore what happens at the molecular scale.

Chapter 7

MOLECULAR SIGNATURES OF HUFF-N-PUFF ENHANCED OIL RECOVERY

7.1 Introduction

In the previous chapter, two huff-n-puff cycles were modeled using molecular dynamics simulations and the results show a good agreement with experimental observations. The simulation design thus captures an important portion of the physical phenomenology that occurs in shales. This chapter explores what happens at the molecular scale when the solvent is injected and interacts with the oil molecules the pores. Additionally, this chapter tracks the penetration depth of the solvent into kerogen during a long soak period to assess the predominant zone of solvent-oil interactions. The collective transport diffusion coefficient (i.e. the molecules of methane and ethane that make up the solvent are considered as one molecular species) is also estimated.

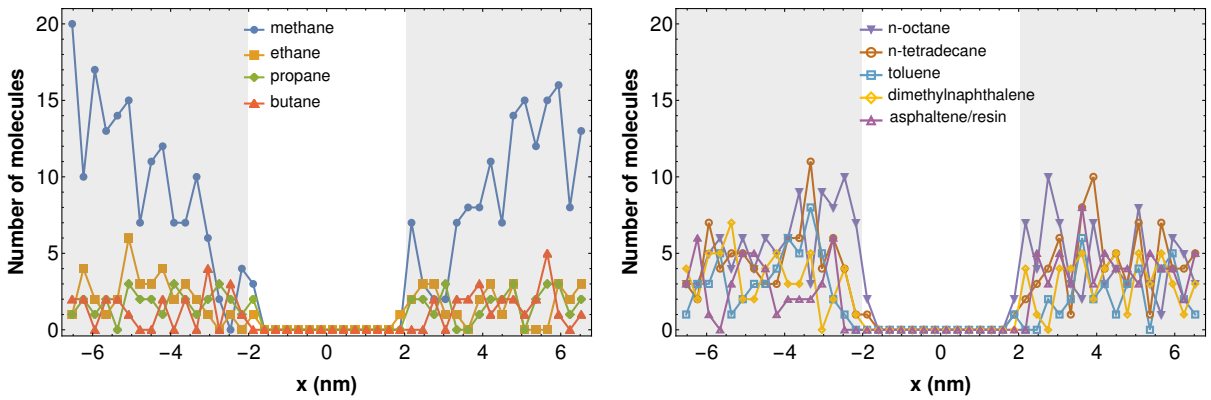
The materials presented in this chapter conclude the study of confined hydrocarbons in organic nanopores and the feasibility of cyclic injection to improve oil recovery from kerogen from a molecular point of view.

7.2 Where are we after primary depletion?

During primary production, the produced fluid comes from the free fluid phase in the vicinity of the microfracture as well as some of the absorbed molecules that desorb due to thermal fluctuations and a drop in pore pressure. Thus the near fracture zone is depleted of the lighter and intermediate molecules while the heavier molecules remain adsorbed. Pores deeper in kerogen largely remain filled with fluid molecules.

This is illustrated in Figure 7.1, which shows the profiles of the different species that remain in the system after 4 ns of primary recovery. This is the initial condition for the huff-n-puff simulations described in Chapter 6. To create the profiles, the simulation box is divided into 60 bins in the x-direction and the number of molecules in each bin is determined. The location of every molecule is given by its center of mass. Figure 7.1(a) depicts the profiles of methane, ethane, propane and n-butane (the smaller species in the system that can be recovered during primary depletion) and Figure 7.1(b) presents the heavier ones which were largely not recovered. Notice that the zones immediately adjacent to the fracture are depleted of methane compared to deeper within kerogen. These zones also show a high concentration of heavier molecules that migrated via surface diffusion (see Section 5.4). These heavy molecules can potentially be a target for enhanced oil recovery. When the solvent is injected at high pressure, it expands into the depleted zone and contacts first these heavier molecules that remain adsorbed.

Figure 7.2 portrays the trajectories described by the center of mass of the unproduced molecules in the simulation box after primary recovery for 4 ns. The conclusions made in Chapter 5 are also true here, namely the unproduced molecules move across the simulation cell, an important fraction of the light species remain unproduced after 4 ns, and the heavier species in the oil model move along the pore surface. In particular, n-octane and n-tetradecane molecules slide along the pore surfaces to migrate to the microfracture, where they remain largely adsorbed on the kerogen-microfracture interface or travel back to the porous space in kerogen.



(a) (b)

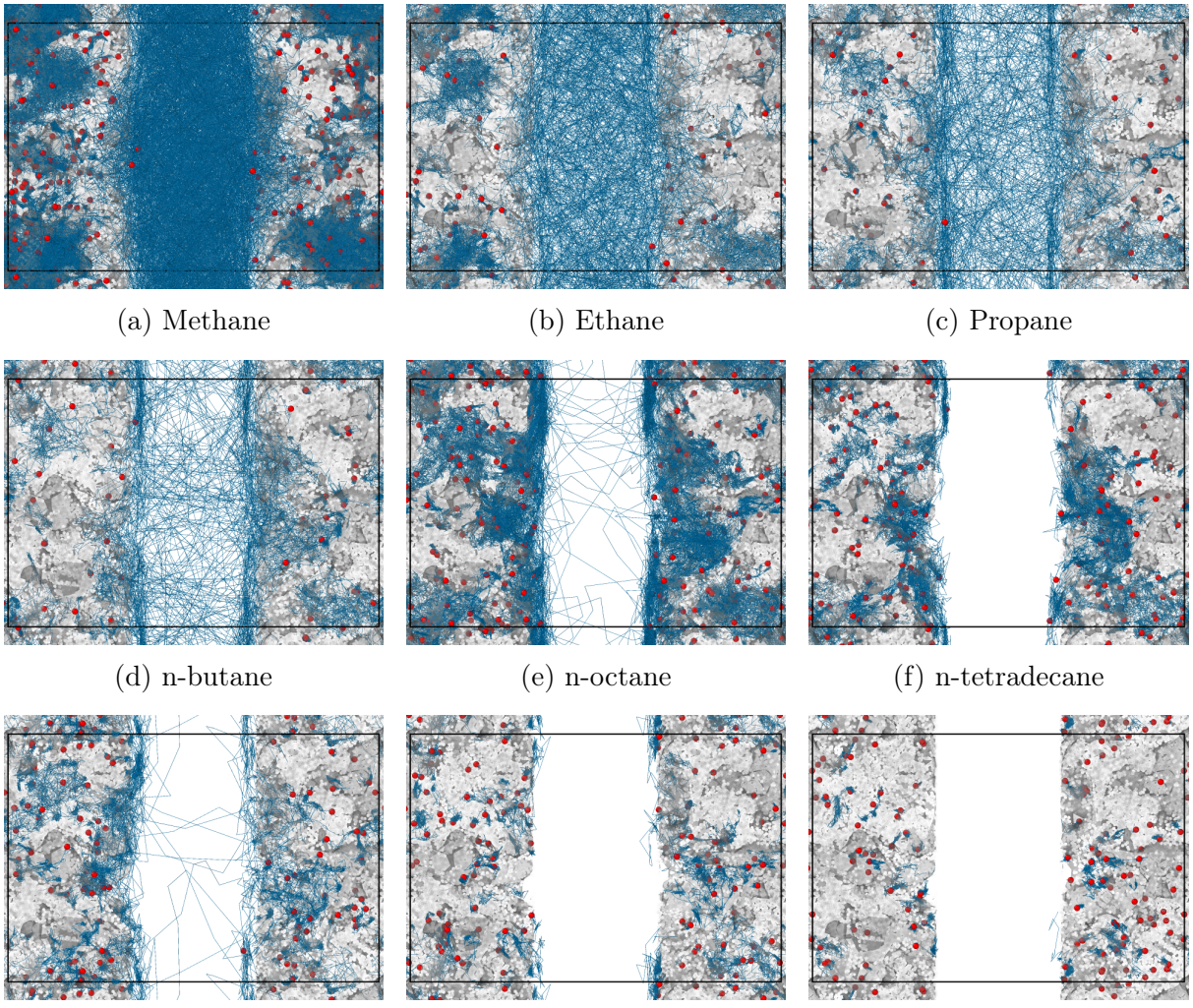
Figure 7.1: Profiles of the different species that remain unproduced after primary depletion for 4 ns. The light-gray regions denote kerogen, the middle zone is the microfracture. This is the initial state of the in-situ fluids in huff-n-puff modeling. (a) Smaller molecules in the system (methane, ethane, propane, n-butane); (b) larger molecules in the system (n-octane, n-tetradecane, toluene, dimethylnaphthalene, asphaltene/resin).

7.3 The role of the solvent during soaking

Following primary depletion, the solvent is injected above the minimum miscibility pressure for a soaking period. This section addresses two questions: (1) How far does the solvent penetrate into kerogen during the soaking period? (2) What is the behavior of the remaining hydrocarbon molecules in the presence of the solvent? The subsections below discuss the insights from molecular simulations in each case.

7.3.1 Solvent penetration into kerogen

In a huff-n-puff cycle, soaking time allows for mixing of oil and solvent via advection or diffusion. The common belief is that longer soaking times promote a deeper solvent penetration, enabling better mixing with larger volumes of oil and therefore, more recovery. It turns out that it is possible to assess this from molecular simulations by quantifying the solvent penetration into kerogen as a function of time. To that end, the trajectory described by the center of mass of each molecule of methane and ethane in the solvent is tracked, considering those species as just one chemical entity, collectively referred to as the solvent. The simulation cell is divided into 50 bins in the x-direction and the number of solvent molecules in each spatial bin is determined at different elapsed times during



(a) Methane (b) Ethane (c) Propane

(d) n-butane (e) n-octane (f) n-tetradecane

(g) Toluene (h) Dimethylnaphthalene (i) Asphaltene/resin

Figure 7.2: Trajectory (blue lines) of the center of mass (red circles) of molecules that are not produced after 4 ns of primary depletion for the black oil case. Black box: simulation box. Light-gray: atoms in kerogen. The dark-gray surface is a render of the pore surface.

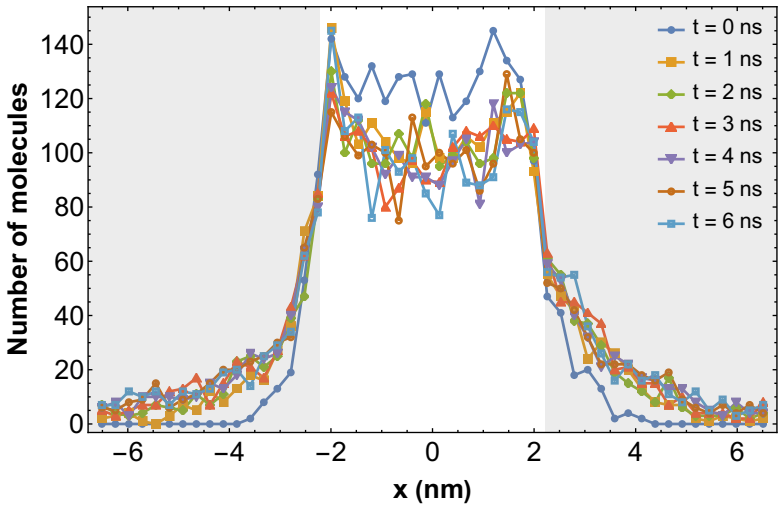
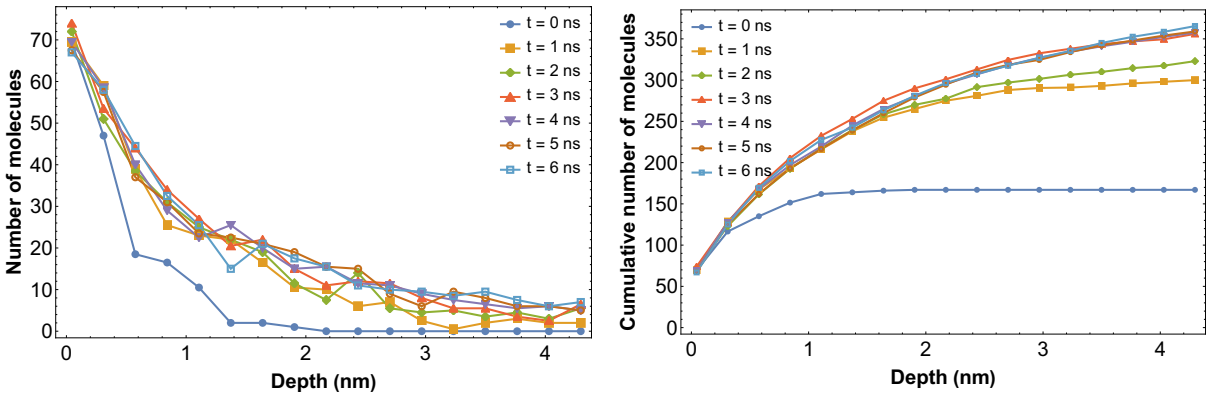


Figure 7.3: Solvent penetration into kerogen during the soaking period. The light-gray regions denote kerogen, the white zone in the middle is the microfracture.

soaking.

Since the solvent is injected at a high pressure of 300 atm (4400 psi) following primary depletion, the pore pressure at the beginning of the soaking period is lower than the injection pressure and the initial migration of the solvent is assisted by a pressure gradient between the microfracture and the interior of kerogen. A longer soak time of 6 ns is considered in this part of the study to capture both pressure and concentration gradient driven transport. Figure 7.3 presents snapshots of the solvent profiles as a function of position at successive simulation times.

Figure 7.4 shows the averaged solvent profiles from both sides of the microfracture. Figure 7.4(a) presents the instantaneous number profile and Figure 7.4(b) presents the cumulative number of solvent molecules that invade kerogen. It is seen that the solvent invasion occurs largely in the depleted zone created during primary recovery even after extended soaking times. There is a big jump between the initial and 1 ns profiles due to the rapid expansion of the solvent. The number of solvent molecules that penetrate kerogen increases up to 3 ns of soaking and then the subsequent profiles are quite similar. Note that solvent penetration is restricted to 2.8 nm at 1 ns (filled-square yellow line) and there are moderate increases beyond that depth for later times. These modest increases suggest that solvent penetration is occurring beyond 3 nm but it is limited and driven by the counterflow of hydrocarbon molecules as will be described in Section 7.3.2. In any



(a) (b)

Figure 7.4: Average penetration of the solvent into kerogen at different elapsed times during soaking. The depth is measured from the interface kerogen-microfracture. (a) Differential profiles; (b) cumulative profiles.

case, Figure 7.4(a) shows that the number of solvent molecules that invade kerogen drops by a factor of 3 to 4 within a depth of 2 nm and by a factor of 7 to 8 within a depth of 3 nm, demonstrating that the solvent does not penetrate into kerogen beyond the depleted zone in significant amounts.

Collective transport diffusion coefficient of the solvent

If there is no pressure gradient between the microfracture and the organic pores, the transport of solvent is driven solely by a concentration gradient. Under this condition, the physical process is described by the Fick's second law given by equation (3.15) (see Section 3.2.1), which in one dimension is

$$\frac{\partial}{\partial t}c(x,t) = D\frac{\partial^2}{\partial x^2}c(x,t), \tag{7.1}$$

where $c(x,t)$ is the concentration (in this case it represents the number of molecules) at a depth into kerogen x and at an elapsed soaking time t, and D is the transport diffusion coefficient.

In order to find an analytic solution to equation (7.1), the following assumptions are made:

1. The kerogen matrix is rigid and does not change volume or density in response to solvent penetration.

2. The diffusion coefficient D is constant throughout the process.

3. The initial condition is expressed as

$$c(x, t = 0) = 0. \tag{7.2}$$

4. The interface boundary condition considers the number of molecules at the kerogen-microfracture interface constant and given by the number of adsorbed solvent molecules to the surface c_0,

$$c(x = 0, t) = c_0. \tag{7.3}$$

5. The no-flow boundary condition given by

$$\frac{\partial}{\partial x} c(x = h, t) = 0, \tag{7.4}$$

where h is the maximum penetration depth, applies in this case.

Under those assumptions, the analytic solution to Equation (7.1) is given by (Crank, 1975)

$$\frac{c(x,t)}{c_0} = 1 - \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} \sin\left[\frac{(2n-1)\pi}{2h} x\right] \exp\left[-\frac{(2n-1)^2 \pi^2 D}{4h^2} t\right]. \tag{7.5}$$

This expression is then fitted to the average invasion profiles shown in Figure 7.4(a). For each elapsed time, a first fit is attempted in which the only fitting parameter is the diffusivity D while c_0 is kept constant and equal to the average number of solvent molecules in the kerogen-microfracture interface during the entire soaking period. This strategy performs very poorly because among the list of assumptions to find the closed expression (7.5), those given by items 2 and 4 are hardly met at early times, precisely the time interval in which advection is still dominant. With that in mind, a second fit is attempted in which both the diffusivity D and the concentration at the interface c_0 are the fitting parameters. This strategy outperforms the previous one, and it is expected that the estimations at later times (say after 3 ns) will be more representative of the

Table 7.1: Fitting parameters to Fick's second law at different elapsed soaking times.

t (ns)	Number of terms	c_0	D ($\times 10^{-10}$ m^2/s)
1	4	68 ± 3	8.2 ± 1.0
2	6	64 ± 3	5.2 ± 0.7
3	4	66 ± 3	4.2 ± 0.6
4	4	62 ± 4	3.6 ± 0.6
5	4	59 ± 4	3.3 ± 0.5
6	3	62 ± 4	2.5 ± 0.4

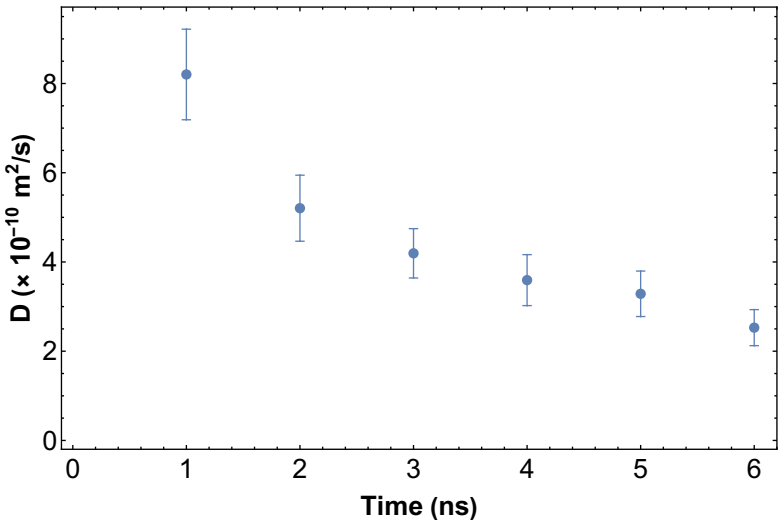


Figure 7.5: Collective transport diffusion coefficient of the solvent at different elapsed soaking times.

diffusive process.

The series in equation (7.5) has an infinite number of terms but it quickly converges within the first 15 terms. Moreover, the best fit for each profile is found considering the first 6 terms of the sum or even fewer. Table 7.1 summarizes the number of terms that give the best fit and the fitting parameters obtained at each elapsed time, and Figure 7.5 presents the values of the collective transport diffusion coefficient of the solvent at different elapsed soaking times.

Column 3 in Table 7.1 shows that the number of molecules at the interface is not constant but may vary without an observed trend. Furthermore, the diffusion coefficients in column 4 and Figure 7.5 decrease as the elapsed time increases due to the overestimation at early times caused by the dominant action of the pressure gradient. Notice however that after an elapsed time of 3 ns the intervals are smaller and the trend levels

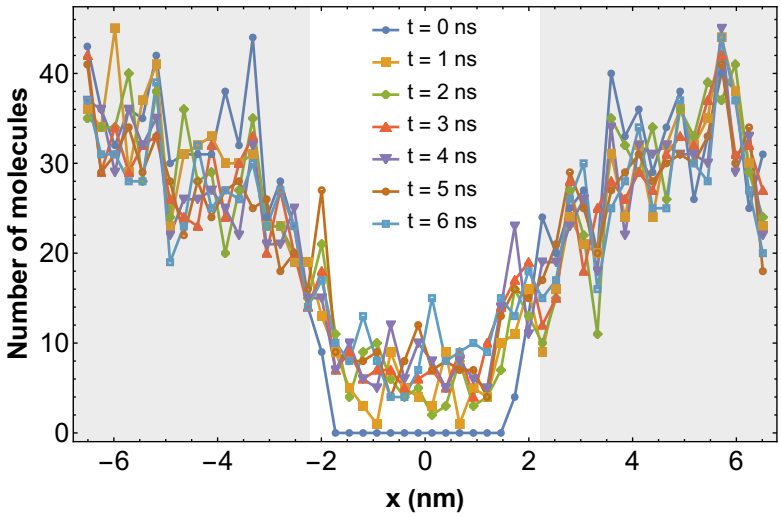


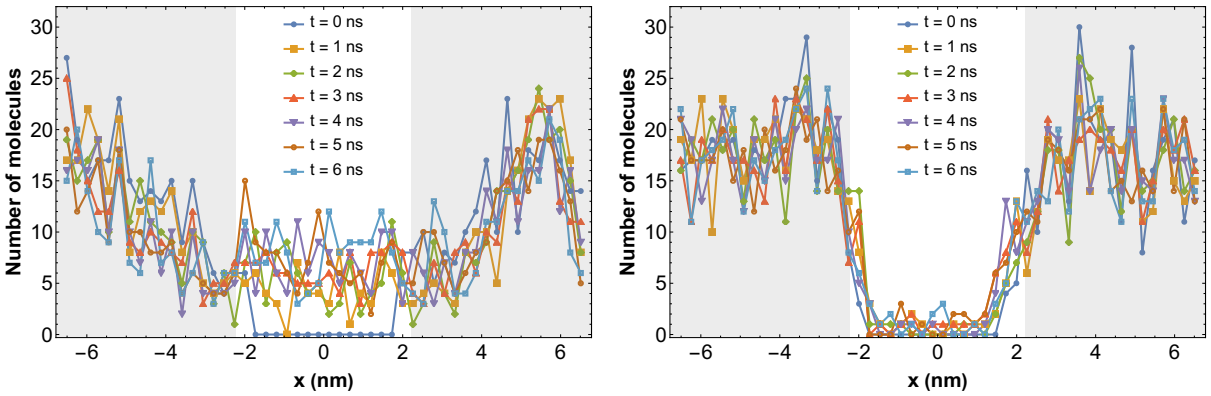
Figure 7.6: Profiles of the reservoir fluids in the system during the soaking period. The light-gray regions denote kerogen, the white zone in the middle is the microfracture.

off, indicating that those are better estimations of the collective transport diffusion coefficient and hence its true value is expected not to exceed 5×10^{-10} m^2/s. All in all, the values of the transport diffusion coefficient are on the order of 10^{-10} m^2/s and following an initial rapid decrease, diminish more gradually as reported experimentally by Dang (2019).

7.3.2 Reservoir fluids response to solvent exposure

It is clear from Chapter 6 that solvent injection promotes additional recovery and enables recovery of species not produced during primary depletion. Figure 7.6 shows the profiles at different elapsed soaking times of the reservoir fluids. The profiles inside kerogen suggest that the solvent helps mobilize oil molecules in the proximity of the microfracture while the number of molecules of the lighter species remains high deeper in the model. As mentioned earlier, a few of the heavier species migrate towards the microfracture as well. It is interesting to note, as mentioned Chapter 6, the occurence of a counterflow of hydrocarbons as shown by the non-zero concentration of the hydrocarbons in the microfracture.

To investigate the composition of the oil mixture in the microfracture, Figure 7.7 redraws the data presented in Figure 7.6 by species with Figure 7.7(a) for the lighter ones produced during primary depletion, and Figure 7.7(b) for the heavier species. The



(a) Lighter species

(b) Heavier species

Figure 7.7: Profiles of the (a) lighter (methane, ethane, propane and n-butane) and (b) heavier (n-octane, n-tetradecane, toluene, dimethylnaphthalene, and the asphlatene/resin fraction) molecular species in the system during the soaking period. The vertical scales are the same to allow direct comparisons. The light-gray region denotes the kerogen, the white region in the middle is the microfracture.

mixture in the microfracture contains solvent molecules (not shown in Figure 7.7) and oil molecules, predominantly of the lighter species. The given solvent promotes desorption of the heavier species but does not keep them in solution. There is, in fact, a continuous competition between desorption/dissolution and re-adsorption of these species.

It was hypothesized in Section 6.5 that long soaking intervals during late cycles in huff-n-puff projects might allow desorption and migration of light species from deep in the formation to the zone invaded by solvent, which would result in a solvent saturated with light species and unable to dissolve heavier molecules that are adsorbed on the pore surfaces. Based on the fluid profiles shown in Figure 7.7(a), the results from the simulations carried out in this work are inconclusive in this regard. The model appears to be too small to investigate this. That is, however, a valid hypothesis.

7.4 Solvent/no solvent: the molecular picture

Figure 7.2 shows the trajectories described by the center of mass of the residual hydrocarbon molecules in the system after primary depletion for 4 ns. It was discussed in Section 5.4 that although those molecules are not produced, they do move across the simulation box to varying extents. It is now interesting to see how these molecules move

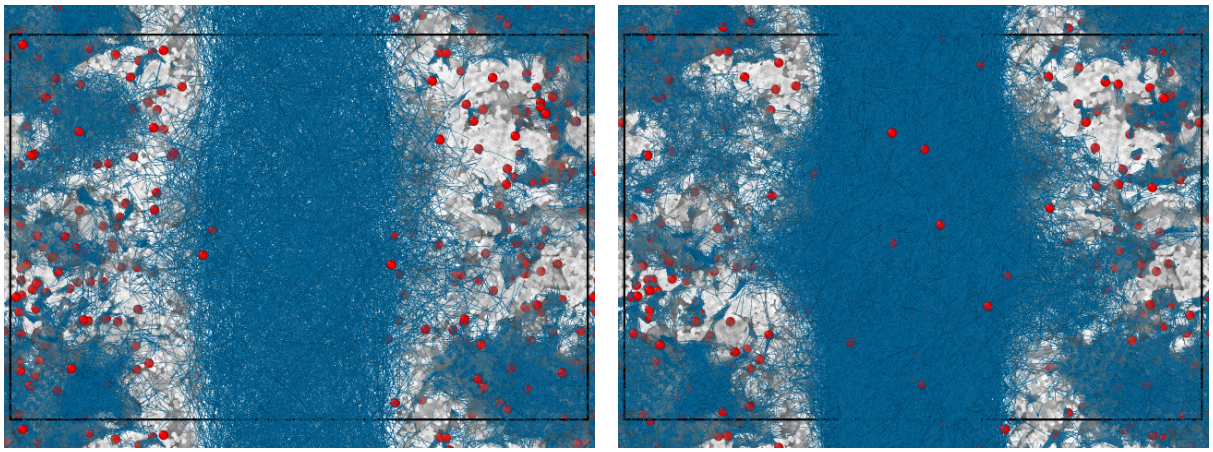
under solvent exposure during soaking. The comparison between the situations of no solvent (primary depletion period) and solvent (soak period) in the system is presented in Figure 7.8. The left column shows what happens during primary depletion and the right column illustrates how every species behaves in the presence of the solvent. In general, the solvent enhances the ability of all species to move around the simulation box. Notice that the trajectory density increases for all species, expressed by the more crowded blue lines both within the pores and in the microfracture (except for the asphaltene/resin fraction that remains inside the pores). All species experience an increased mobility either along the pore surfaces or detach from them and work their way to the microfracture. Three general observations are:

- The shorter the alkane chain, the more efficiently this specific solvent enhances its mobility. This is the trend for all linear alkane molecules in the fluid model.

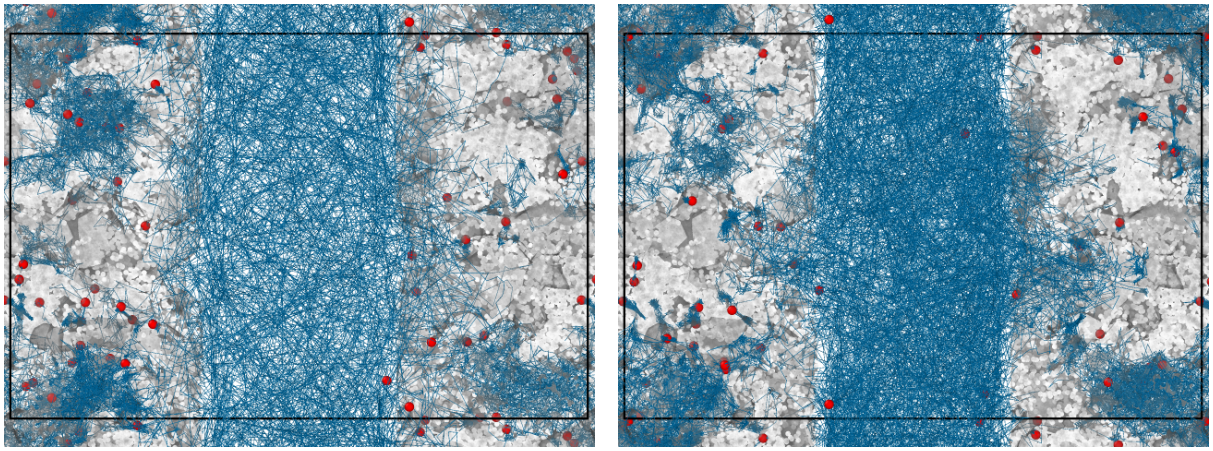
- The geometric organization of the atoms in a molecule determines how it interacts with other neighboring molecules. Dimethylnaphthalene (aromatic) and n-tetradecane (linear chain) species have both 14 carbon atoms but their behavior during primary recovery and response to exposure to the solvent are significantly different.

- Furthermore, the smaller the molecular surface area, the more effectively this particular solvent helps mobilize the species. This is seen in the responses of toluene and dimethylnaphthalene (aromatic species).

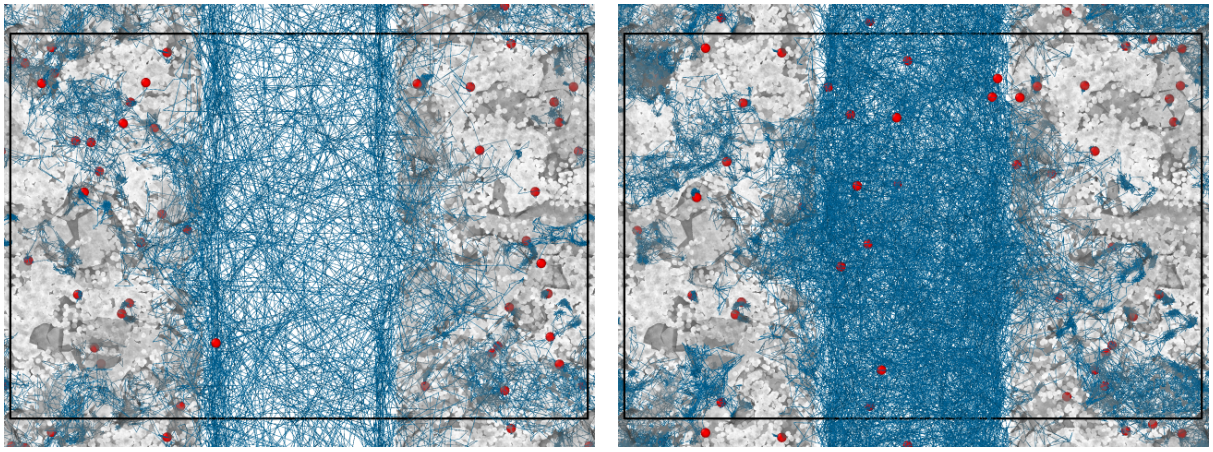
Overall, examining the differences between solvent/no solvent in the system one can see that the solvent is able to help desorb molecules of the intermediate species that were left behind during primary production. Of particular interest is the response of n-octane that, although able to move around and migrate to the microfracture, it hardly desorbs from the pore surface or the kerogen-microfracture interface. In the presence of the solvent n-octane now desorbs and fills the microfracture for an extended time, increasing its potential to be recovered once the system is put back on production. A similar comment can be made about n-tetradecane although to a smaller extent. These



(a) Methane

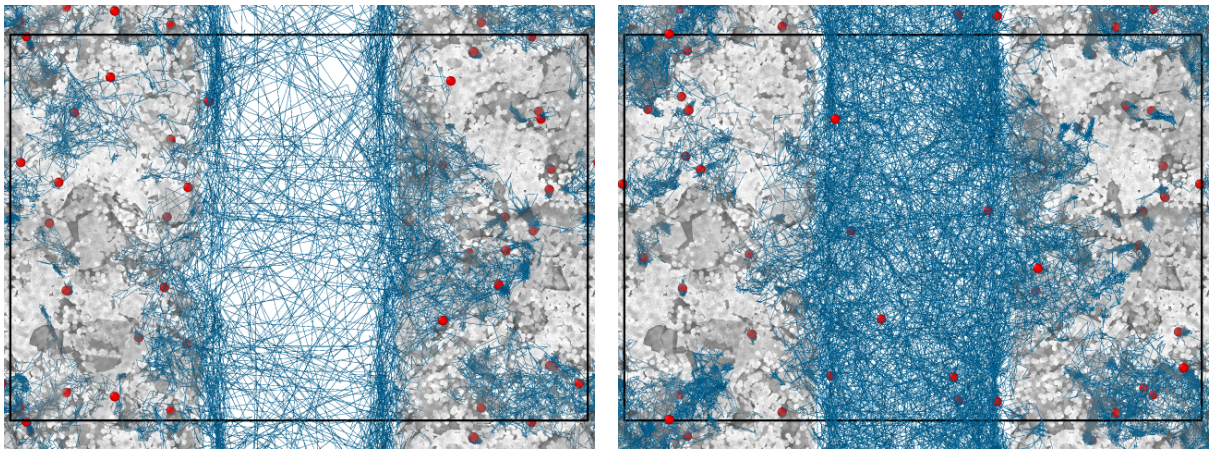


(b) Ethane

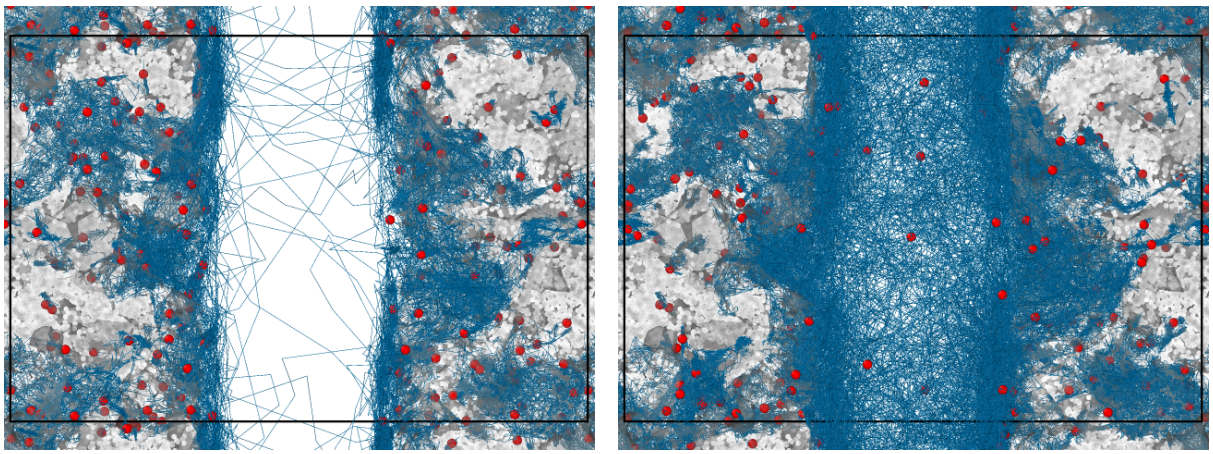


(c) Propane

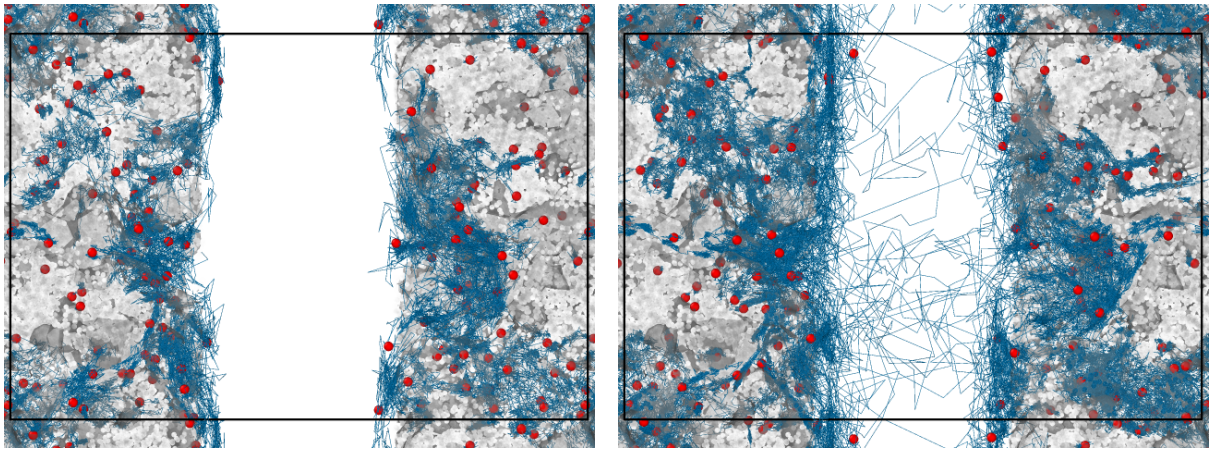
Figure 7.8: Comparison of the trajectories (blue lines) described by the center of mass (red circles) of the unrecovered molecules during 4 ns of primary depletion (left column) and in the presence of an organic solvent during 6 ns of soak period (right column). Black box: simulation box. Light-gray: atoms in kerogen. The dark-gray surface is a render of the pore surface. (To be continued)



(d) n-butane

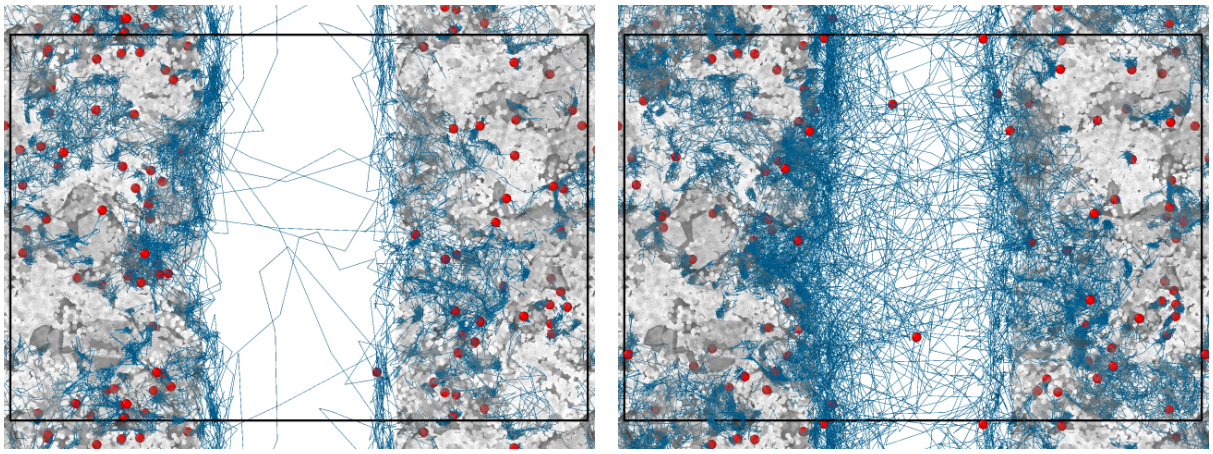


(e) n-octane

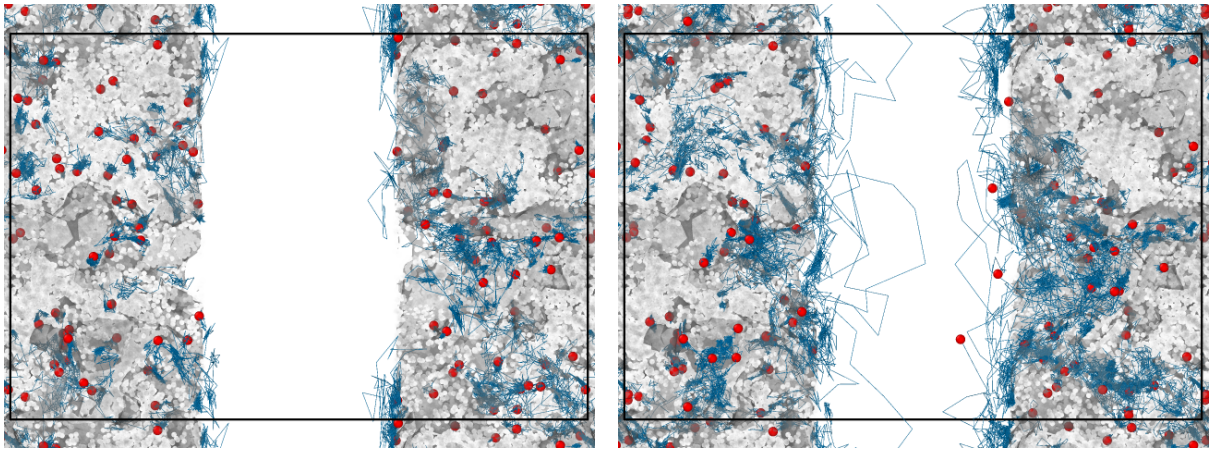


(f) n-tetradecane

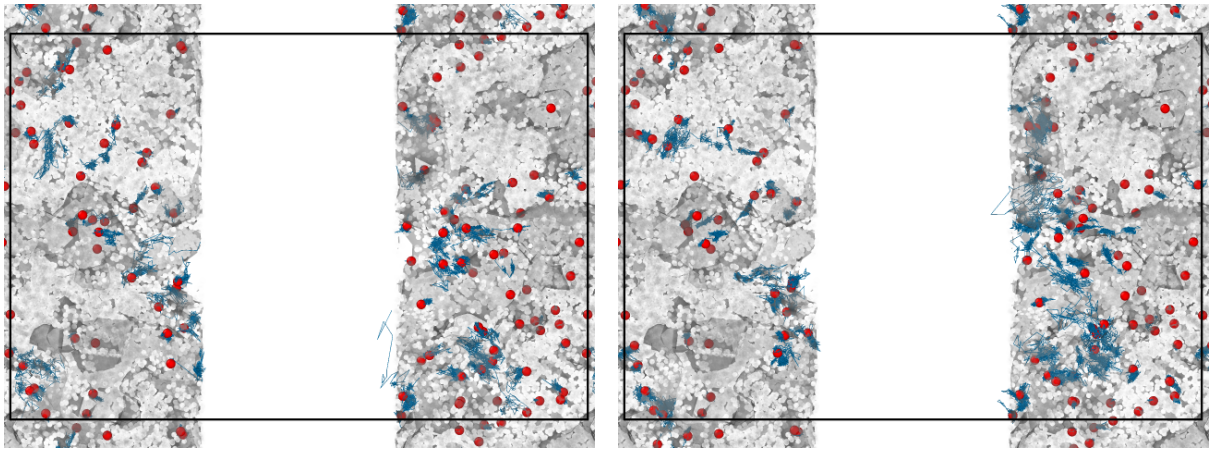
Figure 7.8: continued.



(g) Toluene



(h) Dimethylnaphthalene



(i) Asphaltene/resin

Figure 7.8: continued.

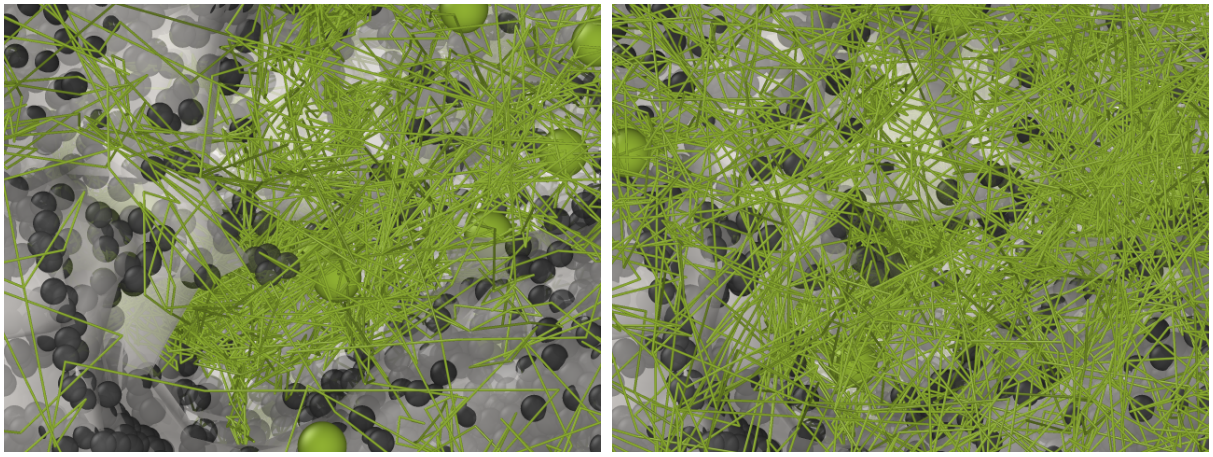
molecular pictures ultimately show why the produced oil during cyclic injection is heavier than that during primary recovery.

7.5 Final remarks

All fluid molecules inside organic pores move around to certain extent. Primary recovery relies on the ability of every species to work its way to the microfracture in response to pressure and concentration gradients. The final recovery is somewhere between 5 and 10%, necessitating some form of improved recovery to produce part of the molecules left behind. Gas cyclic injection has proven to enhance oil recovery in rock cuttings, rock plugs (Mamoudou et al., 2020) and reservoir pilot tests (Hoffman, 2018). However, the underlying mechanisms by which the injected gas operates have remained elusive because of the technological challenges to routine laboratory measurements. Molecular simulations provide the means to explore the phenomenology behind an improved oil recovery strategy.

The common belief is that the longer the rock is exposed to the solvent, the deeper the solvent penetration and the better it mixes with larger volumes of oil leading to more oil recovery. Nevertheless, the simulations carried out show that the solvent does not invade in significant amounts zones deeper than the region depleted during primary recovery. At molecular scale, this is a range between 2 and 3 nm away from the microfracture in the kerogen model; scaling up this range to reservoir scale shows that solvent penetration rarely exceeds a foot or two from the fracture faces (Perez and Devegowda, 2020a). This observation signals that the creation of surface area during fracturing is extremely important as this will determine the volume of reservoir fluids that can be contacted by the solvent. Aside from that, transport diffusion coefficients of a solvent into reservoir fluids on the order of 10^{-10} m^2/s further indicates that a deep penetration of solvent into the rock would take unpractical amounts of time.

Moreover, limited solvent penetration turns out in the existence of three regions inside the rock, one where the concentration of solvent is higher than that of reservoir



(a) (b)

Figure 7.9: Trajectories of the center of mass of n-octane molecules during (a) primary recovery and (b) soak periods. (the observation location is the same in both cases). The gray surface is a render of the pore surface.

fluids, one where the concentrations of solvent and reservoir fluids are approximately the same, and one where the concentration of reservoir fluids is higher than that of solvent. In the shallow zone the solvent dissolves the oil whereas in the deeper one the oil would dissolve the injected gas if its solubility is high enough. Therefore, it is expected that the ways in which the solvent interacts with reservoir fluids are different in every region.

To give a final evidence of the benefits of solvent injection and what it does when it interacts with reservoir fluids in the shallow zone, Figure 7.9 presents the trajectories described by the center of mass of n-octane molecules during primary recovery and during the soaking period at the exact same location to compare the different behavior of this species without and with solvent, respectively. It is evident that n-octane gains mobility when the solvent is present.

Chapter 8

WHAT HAVE WE LEARNED? LESSONS FROM MOLECULAR SIMULATIONS

"All models are wrong, but some are useful."

George Box, Science and Statistics

This works attempts to provide the reader with a comprehensive study from a molecular standpoint of fluid behavior and transport in organic nanopores in shales. It has been long known that the final primary oil recovery from such reservoirs is no greater than 10%, therefore there are still significant amounts of valuable resources to be recovered. With the intrinsic volatility of oil price and at its current value, it seems just cautious to use the available resources more wisely and carefully. A thorough understanding of the fundamental phenomenology that occurs inside such small pores will eventually lead to best practices in exploitation of unconventional reservoirs.

Every chapter focuses on a different aspect of confined fluids in organic nanopores. Chapter 2 considers the question of how the fluids are distributed spatially inside the pores. It is observed that there exist two fluid phases in the pores, one free phase in the middle of the pores where the influence of the pore surfaces is not as strong, and an adsorbed phase onto the pore surfaces due to the enhanced interactions between atoms in the pore walls and atoms in the fluid molecules. The free fluid phase is made up mostly of methane and a few of other small linear molecules, and the tendency to be

adsorbed increases as the size of the molecules increase. When water and carbon dioxide are present in organic matter, water molecules form droplets that may be in contact with the pore surface but do not wet it, meaning the contact angle is less than 90°. Water droplets cover part of the middle region in the pores because of the hydrophobic environment in which water is immersed. Carbon dioxide molecules, on the other hand, are located closer to the pore surface than other hydrocarbon species, meaning this species preferentially adsorbs onto kerogen over hydrocarbons. Moreover, the thickness of the adsorbed phase is between 4 and 5 Å, thick enough to accommodate approximately 80% of the total number of hydrocarbons and carbon dioxide molecules in the pore space for the composition utilized in this work.

Since the different compounds that make up an oil mixture are segregated within organic pores, there exist local concentration gradients that might promote molecular diffusion to produce a uniform mixture. Chapter 3 explores the self diffusion of the hydrocarbon species inside organic pores, i.e. the random motion of the molecules to produce a uniform mixture. Interestingly, the enhanced interactions between pore walls and fluid molecules compete with the natural tendency of matter to eliminate spacial concentration gradients, and the molecules do move within organic pores at different extents as a result of their differential interactions with the pore walls. Methane, the smallest hydrocarbon species in the system, has the highest ability to move around. As the molecular size and molecular surface area increase, the molecular species tends to adsorb more strongly to the pore surface, but it does not imply the freezing of the molecules at specific locations. Thus, even adsorbed hydrocarbon molecules do move inside organic pores via surface diffusion.

Chapter 4 studies how porous organic matter saturated with an oil mixture interacts with a microfracture next to it. In other words, the confined fluids in organic nanopores have now the possibility to escape to a microfracture. The system is allowed to relax until it reaches a state of (quasi-)equilibrium, one in which no considerable exchange of molecules occurs between the small pores and the microfracture. Although the molecular simulations carried out by no means represent reservoir scales over geologic times, some

interesting insights are obtained from short (1 ns) and long (12 ns) simulations. The most remarkable is that the composition of the oil in the microfracture is lighter than that in small pores. The reason is twofold: (1) the smaller the pore, the stronger the interactions between the pore walls and molecules with long lengths and large surface areas; (2) the complexity of the pore throats that connect one pore with another reduces the ability of big molecules to flow. Arguably, simulations that span 1 ns are several orders of magnitude below the time scales considered even in laboratory experiments. However, the fact that no significant exchange of molecules happens between the two regions after 12 ns indicates that the results are not expected to change even if the simulations are run for longer times. Thus, if the results are trustworthy and scalable, large pores host light oils whereas small pores host heavy oils, and if the pore network is connected to a crack in a large fracture network, the first pores to be depleted are the large ones and the oil that will be recovered more easily is lighter in composition than whatever fluid remains behind in the reservoir.

If the microfracture created in Chapter 4 is now part of a large fracture network that connects with a well, oil recovery can be then simulated by periodically removing the fluid molecules that work their way to the microfracture. Although this approach is a discrete representation of a continuous process, molecular dynamics simulations in Chapter 5 show that there is a variation in the composition of the oil mixture produced during primary depletion that is also observed in field operations. The early mixtures are lighter but the molecular weight increases with production time. Based on the lessons learned so far, the free fluid phase is produced more easily first, and later on some heavier molecules desorb and migrate to the microfracture to be produced. Nevertheless, the recovered oils are largely composed of methane, ethane, propane and n-butane, leaving behind most of the heavier species in the pores. Furthermore, the level of detail provided by molecular simulations permits identification of a thin depleted zone neighboring the microfracture. Such a zone is still rich in heavier molecules that cannot be produced due to the strong adsorption to either the pore surface or the microfracture faces.

The ultimate challenge is to push up the recovery of the species that stay in the

pores after primary depletion. The use of cyclic gas injection has proven to increase oil recovery experimentally and in field operations, but the underlying mechanisms by which it operates remain elusive. In Chapter 6, an organic solvent is injected into the microfracture and the subsequent mixing with reservoir fluids is simulated. The composition of the solvent, the conditions of pressure and temperature, and the way simulations proceed correspond with the choices made in the experimental work carried out at the University of Oklahoma. First, the kerogen model is soaked for some time, then the system is put back on production following the strategy used in Chapter 5. This is one cycle in a huff-n-puff enhanced oil recovery process. Then the cycle is repeated all over to end up simulating two cycles. Remarkably, this rather simplistic modeling work reproduces the trends observed experimentally and enables extensions to larger scales. This is evidence that the model captures important physics that occurs in the organic portion of shales. Based on the trends for oil recovery and molecular weight of the hydrocarbon mixture produced for different residence times (a concept defined by equation (6.1)), the optimum huff-n-puff design should comprise long times of solvent exposure at early cycles followed by shorter ones at later cycles. The unsettled question is why.

Chapter 7 explores the footprints of the process at molecular scale. The zone neighboring the microfracture is stripped of light species but still rich in adsorbed heavy molecules. The deeper region in kerogen contains higher concentrations of reservoir fluids, both free fluid and adsorbed phases. The solvent expands rapidly once it is injected and invades the depleted zone before the free fluid phase in deeper pores flows to recharge such zone. This allows the solvent to interact with the species that need external aid to be recovered. If the exposure time to solvent is too long, eventually the free fluid phase works its way to the zone of solvent invasion. This might be undesirable because the solvent will dissolve more easily the lighter species, and once it is saturated there will be no room for heavier ones. The rationale then is to use long residence times at early stages of recovery with a light solvent to deplete the vicinity of the fractures of the most mobile species to promote the future interaction of a more aggressive solvent with the species that do need help to be recovered during later cycles, when exposure to the solvent

should be for shorter periods of time to avoid the arrival of lighter species to saturate the solvent in undesirable ways. Moreover, even after extended periods of time, the solvent does not penetrate deep into the kerogen model, rather it is limited to the vicinity of the microfracture. In fact, the estimated transport diffusion coefficients are on the order of 10^{-10} m^2/s. This is indication that ultimately the success of gas injection depends on the available surface area created either naturally or during hydraulic fracturing.

This concludes the study of fluid behavior and transport in organic nanopores in shales from a molecular point of view. Although the spatial and time scales accessible to molecular simulations are several orders of magnitude below those in experimental and field scales, the fundamental premise is that all observed macroscopic properties in a system arise at the most fundamental level from the interactions and dynamics of individual components (Krapf and Metzer, 2019). The careful design and use of molecular models may provide a means of rationalizing macroscopic observations in terms of the detailed motion of the fundamental constituents of a system. In the end, a physical model is only as good as its ability to reproduce observed responses and correctly predict future outcomes under different conditions. The simplistic models created in this work seem to do this.

REFERENCES

Alfarge, D., M. Wei, and B. Bai (2017). IOR methods in unconventional reservoirs of North America: Comprehensive review. In *SPE Western Regional Meeting, 23-27 April, Bakersfield, California*. Society of Petroleum Engineers. SPE-185640-MS.

Alharthy, N., T. Teklu, H. Kazemi, R. Graves, S. Hawthorne, J. Braunberger, and B. Kurtogle (2018). Enhanced oil recovery in liquid-rich shale reservoirs: Laboratory ot field. *SPE Reservoir Evaluation & Engineering 21*(1), 137–159. SPE-175034-PA.

Allen, M. P. and D. J. Tildesley (2017). *Computer Simulation of Liquids* (Second ed.). Oxford University Press.

Ambrose, R. J., R. C. Hartman, M. Diaz-Campos, I. Y. Akkutlu, and C. H. Sondergeld (2012). Shale gas-in-place calculations part I: New pore-scale considerations. *SPE Journal*, 219–229. SPE-131772-PA.

Aminzadeh, F. and S. N. Dasgupta (2013). Geophysics for Unconventional Resources. In J. Cubitt (Ed.), *Developments in Petroleum Science*, Volume 60, Chapter 9. Elsevier.

Andersen, H. C. (1980). Molecular dynamics at constant temperature and/or pressure. *Journal of Chemical Physics 72*(4), 2384–2393.

Arfken, G. (1985). *Mathematical Methods for Physicists* (Third ed.). Academic Press, Inc. San Diego, CA.

Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry 91*, 6269–3271.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. R. Haak (1984). Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics 81*(8), 3684–3690.

Berendsen, H. J. C. and W. F. van Gunsteren (1986). Practical algorithms for dynamic simulations. In G. Ciccotti and W. G. Hoover (Eds.), *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, pp. 43–65. Proceedings of the International School of Physics "Enrico Fermi": North-Holland Physics Publishing.

Bhattacharya, S. and K. E. Gubbins (2006). Fast method for computing pore size distributions of model materials. *Langmuir 22*, 7726–7731.

Billinge, S. J. L. and I. Levin (2007). The problem with determining atomic structure at the nanoscale. *Science 316*, 561–565.

Born, M. and R. Oppenheimer (1927). Zur quantentheorie der molekeln. *Annalen der Physik 389*(20), 457–484.

Bousige, C., C. M. Ghimbeu, C. Vix-Guterl, A. E. Pomerantz, A. Suleimenova, G. Vaughan, G. Garbarino, M. Feygenson, C. Wildgruber, F.-J. Ulm, et al. (2016). Realistic molecular model of kerogen's nanostructure. *Nature Materials 15*(5), 576–582.

Bransden, B. and C. Joachain (1983). *Physics of Atoms and Molecules*. Longman Scientific & Technical. Copublished in the United States with John Wiley & Sons, Inc. New York.

Byron, Jr., F. W. and R. W. Fuller (1992). *Mathematics of Classical and Quantum Physics*. Dover Publications, Inc. New York.

Callen, H. (1985). *Thermodynamics and An Introduction to Thermostatistics* (Second ed.). Wiley. New Delhi, India.

Calvo-Muñoz, E. M., M. E. Selvan, R. Xiong, M. Ojha, D. J. Keffer, D. M. Nicholson, and T. Egami (2011). Applications of a general random-walk theory for confined diffusion. *Physical Review E 83*, 011120.

Campos, D. (1997). *Fundamentos de Física Atómica y Molecular*. Editorial Universidad Nacional. Bogota, Colombia.

Clark, A. J. (2009). Determination of recovery factor in the Bakken Formation, Mountrail County, ND. In *SPE Annual Technical Conference and Exhibition, 4-7 October, New Orleans, Louisiana*. Society of Petroleum Engineers. SPE-133719-STU.

Collell, J., G. Galliero, F. Gouth, F. Montel, M. Pujol, P. Ungerer, and M. Yiannourakou (2014). Molecular simulation and modelisation of methane/ethane mixtures adsorption onto microporous molecular model of kerogen under typical reservoir conditions. *Microporous and Mesoporous Materials 197*, 194–203.

Collell, J., G. Galliero, R. Vermorel, P. Ungerer, M. Yiannourakou, F. Montel, and M. Pujol (2015). Transport of multicomponent hydrocarbon mixtures in shale organic matter by molecular simulations. *Journal of Physical Chemistry C 119*, 22587–22595.

Collell, J., P. Ungerer, G. Galliero, M. Yiannourakou, F. Montel, and M. Pujol (2014). Molecular simulation of bulk organic matter in type II shales in the middle of the oil formation window. *Energy & Fuels 28*, 7457–7466.

Crank, J. (1975). *The Mathematics of Diffusion* (Second ed.). Oxford University Press. London.

Cristancho, D., I. Y. Akkutlu, L. J. Criscenti, and Y. Wang (2016). Gas storage in model kerogen pores with surface heterogeneities. In *SPE Europec featured at 78th EAGE Conference and Exhibition, 30 May-2 June, Vienna, Austria*. Society of Petroleum Engineers. SPE-180142-MS.

Cygan, R. T., V. N. Romanov, and E. M. Myshakin (2012). Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field. *Journal of Physical Chemistry C 116*, 13079–13091.

Dang, S. T. (2019). *Understanding the fundamental drive mechanisms for huff-n-puff enhanced oil recovery in tight formations*. Ph.D. dissertation, University of Oklahoma.

Dang, S. T., C. H. Sondergeld, and C. S. Rai (2016). A new approach to measuring organic density. *Petrophysics 57*(2), 112–120. SPWLA-2016-v57n2a3.

Dauber-Osguthorpe, P., V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler (1988). Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethorprim, a drug-receptor system. *Proteins: Structure, Function, and Genetics 4*, 31–47.

de Groot, S. R. and P. Mazur (1984). *Non-Equilibrium Thermodynamics*. Dover Publications, Inc. New York.

Devegowda, D., K. Sapmanee, F. Civan, and R. F. Sigal (2012). Phase behavior of gas condensates in shales due to pore proximity effects: Implications for transport, reserves and well productivity. In *SPE Annual Technical Conference and Exhibition, 8-10 October, San Antonio, Texas, USA*. Society of Petroleum Engineers. SPE-160099-MS.

Dodda, L. S., I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen (2017). LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research 45*(W1), W331–W336.

Durand, B. (1980). *Kerogen: Insoluble Organic Matter from Sedimentary Rocks*. Editions TECHNIP. Paris.

EIA (2019). U.S. Crude Oil and Natural Gas Proved Reserves, Year-End 2018. Technical report, U.S. Energy Information Administration, Office of Energy Analysis. U.S. Department of Energy, Washington, DC 20585. December.

Erpenbeck, J. J. and W. W. Wood (1977). Molecular dynamics techniques for hard-core systems. In B. J. Berne (Ed.), *Statistical Mechanics. Part B: Time-Dependent Processes*, Volume 6 of *Modern Theoretical Chemistry*, pp. 1–40. Plenum Press. New York.

Falk, K., B. Coasne, R. Pellenq, F.-J. Ulm, and L. Bocquet (2015). Subcontinuum mass transport of condensed hydrocarbons in nanoporous media. *Nature Communications 6*, 6949.

Freeman, C., G. J. Moridis, and T. A. Blasingame (2013). Modeling and performance interpretation of flowing gas composition changes in shale gas wells with complex fractures. In *International Petroleum Technology Conference, 26-28 March, Beijing, China*. Society of Petroleum Engineers. IPTC-17075-MS.

Freeman, C., G. J. Moridis, G. E. Michael, and T. A. Blasingame (2012). Measurement, modeling, and diagnostics of flowing gas composition changes in shale gas wells. In *SPE Latin America and Caribbean Petroleum Engineering Conference, 16-18 April, Mexico City, Mexico*. Society of Petroleum Engineers. SPE-153391-MS.

Frenkel, D. and B. Smith (2002). *Understanding Molecular Simulation. From Algorithms to Applications*. Academic Press. San Diego, CA.

Gasparik, M., P. Bertier, Y. Gensterblum, A. Ghanizahed, B. M. Kross, and R. Littke (2014). Geological controls on the methane storage capacity in organic-rich shales. *International Journal of Coal Geology 123*, 34–51.

Gasparik, M., A. Ghanizahed, P. Bertier, Y. Gensterblum, S. Bouw, and B. M. Kross (2012). High-pressure methane sorption isotherms of Black Shales from The Netherlands. *Energy & Fuels 26*, 4995–5004.

Goldstein, H. (1980). *Classical Mechanics* (Second ed.). Addison-Wesley.

Haile, J. M. (1992). *Molecular Dynamics Simulation Elementary Methods*. John Wiley & Sons, Inc. New York.

Haile, J. M. and S. Gupta (1983). Extensions of the molecular dynamics simulation method. II. Isothermal systems. *Journal of Chemical Physics 79*(6), 3067–3076.

Heller, R. and M. Zoback (2014). Adsorption of methane and carbon dioxide on gas shale and pure mineral samples. *Journal of Unconventional Oil and Gas Resources 8*, 14–24.

Herdes, C., C. Petit, A. Mejía, and E. Müller (2018). Combined experimental, theoretical, and molecular simulation approach for the description of the fluid-phase behavior of hydrocarbon mixtures within shale rocks. *Energy & Fuels 32*, 5750–5762.

Ho, T. A., L. J. Criscenti, and Y. Wang (2016). Nanostructural control of methane release in kerogen and its applications to wellbore production decline. *Scientific Reports 6*, 28053.

Ho, T. A., Y. Wang, and L. J. Criscenti (2018). Chemo-mechanical coupling in kerogen gas adsorption/desorption. *Physical Chemistry Chemical Physics 20*, 12390–12395.

Hockney, R. W. and J. W. Eastwood (1988). *Computer simulation using particles*. CRC Press.

Hoffman, B. T. (2018). Huff-n-puff gas injection pilot projects in the Eagle Ford. In *SPE Canada Unconventional Resources Conference, 13-14 March, Calgary, Alberta, Canada*. Society of Petroleum Engineers. SPE-189816-MS.

Hoffman, B. T. and J. G. Evans (2016). Improved oil recovery IOR pilot projects in the Bakken formation. In *SPE Low Perm Symposium, 5-6 May, Denver, Colorado, USA*. Society of Petroleum Engineers. SPE-180270-MS.

Hoffman, B. T. and D. Reichhardt (2019). Quantitative evaluation of recovery mechanisms for huff-n-puff gas injection in unconventional reservoirs. In *SPE/AAPG/SEG Unconventional Resources Technology Conference, 22-24 July, Denver, Colorado*. Society of Petroleum Engineers. URTeC-2019-147-MS.

Hoffman, B. T. and J. M. Rutledge (2019). Mechanisms for huff-n-puff cyclic gas injection into unconventional reservoirs. In *SPE Oklahoma City Oil and Gas Symposium, 9-10 April, Oklahoma City, Oklahoma*. Society of Petroleum Engineers. SPE-195223-MS.

Holstein, E. D. (Ed.) (2007). *Reservoir Engineering and Petrophysics*, Volume V of *Petroleum Engineering Handbook*. Society of Petroleum Engineers. Larry W. Lake, Editor-in-Chief.

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

Hoover, W. G. (1986). Constant pressure equations of motion. *Physical Review A 34*(3), 2499–2500.

Hoover, W. G., A. J. C. Ladd, R. B. Hickman, and B. L. Holian (1980). Bulk viscosity via nonequilibrium and equilibrium molecular dynamics. *Physical Review A 21*(5), 1756–1760.

Huang, K. (1987). *Statistical Mechanics* (Second ed.). John Wiley & Sons, Inc. New York.

Ismail, A. F., K. Khulbe, and T. Matsuura (2015). *Gas Separation Membranes. Polymeric and Inorganic*. Springer International Publishing.

Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society 118*, 11225–11236.

Jung, J., W. Nishima, M. Daniels, G. Bascom, C. Kobayashi, A. Adedoyin, M. Wall, A. Lappala, D. Phillips, W. Fische, C.-S. Tung, T. Schlick, Y. Sugita, and K. Y. Sanbonmatsu (2019). Scaling molecular dynamics beyond 100000 processor cores for large-scale biophysical simulations. *Journal of Computational Chemistry 40*(21), 1919–1930.

Kärger, J., D. M. Ruthven, and D. N. Theodorou (2012). *Diffusion in Nanoporous Materials*, Volume 1. Wiley-VCH Verlag & Co. KGaA. Boschstr. 12, 69469 Weinheim, Germany.

Kazemi, M., A. Takbiri-Borujeni, J. Hansel, and M. Valera (2019). Enhanced oil recovery of shale oil: A molecular simulation study. In *SPE/AAPG/SEG Unconventional Resources Technology Conference, 22-24 July, Denver, CO*. Society of Petroleum Engineers. URTeC-2019-937-MS.

Keffer, D. J. (2002). The working man's guide to obtaining self diffusion coefficients from molecular dynamics simulations. Course website: Modeling and Simulation in Materials Science and Engineering: Classical Mechanics. http://utkstair.org/clausius/docs/mse614/text/notes.html. University of Tennessee, Knoxville.

Kelemen, S. R., M. Afeworki, M. L. Gorbaty, M. Sansone, P. J. Kwiatek, C. C. Walters, H. Freund, M. Siskin, A. E. Bence, D. J. Curry, M. Solum, R. J. Pugmire, M. Vandenbroucke, M. Leblond, and F. Behar (2007). Direct characterization of kerogen by X-ray and solid-state ^{13}C nuclear magnetic resonance. *Energy & Fuels 21*, 1548–1561.

King, G. E. (2014). Maximizing recovery factors. Improving recovery factors in liquids-rich resource plays requires new approaches. The American Oil & Gas Reporter.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983). Optimization by simulated annealing. *Science 220*(4598), 671–680.

Kondepudi, D. and I. Prigogine (2015). *Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Second ed.). Wiley. New Delhi, India.

Krapf, D. and R. Metzer (2019). Strange interfacial molecular dynamics. *Physics Today 72*(9), 48–54.

Lagache, M. H., P. Ungerer, and A. Boutin (2004). Prediction of thermodynamic derivative properties of natural condensate gases at high pressure by Monte Carlo simulation. *Fluid Phase Equilibria 220*(2), 211–223.

Landau, L. and E. Lifshitz (1976). *Course of Theoretical Physics* (Third ed.), Volume 1: Mechanics. Pergamon Press. New York.

Landau, L. and E. Lifshitz (1977). *Course of Theoretical Physics* (Third ed.), Volume 3: Quantum Mechanics–Non-relativistic Theory. Pergamon Press. New York.

Lee, T., L. Bocquet, and B. Coasne (2016). Activated desorption at heterogeneous interfaces and long-time kinetics of hydrocarbon recovery from nanoporous media. *Nature Communications 7*, 11890.

Mamoudou, S., F. Perez, A. Tinni, S. Dang, C. H. Sondergeld, C. S. Rai, and D. Devegowda (2020). Evaluation of huff-n-puff in shale using experiments and molecular simulations. In *Unconventional Resources Technology Conference, 20-22 July, Austin, TX*. Society of Petroleum Engineers. URTeC: 2923.

Marbach, S., D. S. Dean, and L. Bocquet (2018). Transport and dispersion across wiggling nanopores. *Nature Physics 14*, 1108–1113.

Martínez, L., R. Andrade, E. G. Birgin, and J. M. Martínez (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry 30*(13), 2157–2164.

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

McCain, W. D. (1990). *The Properties of Petroleum Fluids* (second ed.). PennWell Books.

McQuarrie, D. A. (1976). *Statistical Mechanics*. Harper & Row Publishers. New York.

Meng, X., J. J. Sheng, and Y. Yu (2017). Experimental and numerical study of enhanced condensate recovery by gas injection in shale gas-condensate reservoirs. *SPE Reservoir Evaluation & Engineering 20*(2), 471–477. SPE-183645-PA.

Michalec, L. and M. Lísal (2017). Molecular simulation of shale gas adsorption onto overmature type II model kerogen with control microporosity. *Molecular Physics 115*(9–12), 1086–1103.

Mitchell, D. L. and J. G. Speight (1973). The solubility of asphaltenes in hydrocarbon solvents. *Fuel 52*(2), 149–152.

Mosher, K., J. He, Y. Liu, E. Rupp, and J. Wilcox (2013). Molecular simulation of methane adsorption in micro- and mesoporous carbons with applications to coal and gas shale systems. *International Journal of Coal Geology 109-110*, 36–44.

Müller, E. A., F. R. Hung, and K. E. Gubbins (2000). Adsorption of water vapor-methane mixtures on activated carbons. *Langmuir 16*(12), 5418–5424.

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

Nosé, S. (1984b). A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics 81*(1), 511–519.

Nosé, S. (1991). Constant temperature molecular dynamics methods. *Progress of Theoretical Physics Supplements 103*, 1–46.

Obliger, A., R. Pellenq, F.-J. Ulm, and B. Coasne (2016). Free volume theory of hydrocarbon mixture transport in nanoporous materials. *Journal of Physical Chemistry Letters 7*, 3712–3717.

Okiongbo, K. S., A. C. Aplin, and S. R. Larter (2005). Changes in type II kerogen density as a function of maturity: Evidence from the Kimmeridge Clay Formation. *Energy & Fuels 19*, 2495–2499.

Parrinello, M. and A. Rahman (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics 52*(12), 7182–7190.

Pathak, M., H. Kweon, M. Deo, and H. Huang (2017). Kerogen swelling and confinement: Its implication on fluid thermodynamic properties in shales. *Scientific Reports 7*, 12530.

Pathak, M., P. Panja, H. Huang, and M. Deo (2016). Enhanced recovery in shales: Molecular investigation of CO_2 energized fluid for re-fracturing shale formations. In *SPE/AAPG/SEG Unconventional Resources Technology Conference, 1-3 August, San Antonio, TX*. Society of Petroleum Engineers. URTeC-2461982-MS.

Perez, F. and D. Devegowda (2017). Estimation of adsorbed-phase density of methane in overmature kerogen models using molecular simulations for accurate gas in place calculations. *Journal of Natural Gas Science and Engineering 46*, 865–872.

Perez, F. and D. Devegowda (2019). Spatial distribution of reservoir fluids in mature kerogen using molecular simulations. *Fuel 235*, 448–459.

Perez, F. and D. Devegowda (2020a). How far does EOR solvent penetrate kerogen in shales? In *Unconventional Resources Technology Conference held in Austin, Texas, USA, 20-22 July 2020*. Society of Petroleum Engineers. URTeC: 2913.

Perez, F. and D. Devegowda (2020b). A molecular dynamics study of primary production from shale organic pores. *SPE Journal Preprint*. SPE-201198-PA.

Perez, F. and D. Devegowda (2020c). A molecular dynamics study of soaking during enhanced oil recovery in shale organic pores. *SPE Journal 25*(02). SPE-199879-PA.

Perez, F. and D. Devegowda (2020d). Molecular signatures of solvent injection in shale organic nanopores. In *SPE Improved Oil Recovery Conference*. Society of Petroleum Engineers. SPE-200307-MS.

Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics 117*, 1–19.

Present, R. D. (1958). *Kinetic Theory of Gases*. McGrow-Hill Book Company, Inc. York.

Price, M. L. P., D. Ostrovsky, and W. L. Jorgensen (2001). Gas-phase and liquid-state properties od esters, nitriles, and nitro compounds with the OPLS-AA force field. *Journal of Computational Chemistry 22*(13), 1340–1352.

Rexer, T. F., M. J. Benham, A. C. Aplin, and K. M. Thomas (2013). Methane adsorption on shale under simulated geological temperature and pressure conditions. *Energy & Fuels 27*(6), 3099–3109.

Rexer, T. F., E. J. Mathia, A. C. Aplin, and K. M. Thomas (2014). High-pressure methane adsorption and characterization of pores in Posidonia shales and isolated kerogens. *Energy & Fuels 28*(5), 2886–2901.

Ryckaert, J.-P., G. Ciccotti, and J. J. Berendsen (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics 23*, 327–341.

Schroeder, D. V. (2000). *An Introduction to Thermal Physics*. Addison Wesley Longman. San Francisco, CA.

Shoaib, S. and B. Hoffman (2009). CO_2 flooding the Elm Coulee field. In *SPE Rocky Mountain Petroleum Technology Conference, 14-16 April 2009, Denver, Colorado*. Society of Petroleum Engineers. SPE-123176-MS.

Sinha, A., S. Dang, C. Sondergeld, and C. Rai (2017). Impact of solvent extraction on surface area measurements in organic-rich shales using nitrogen adsorption. In *Unconventional Resources Technology Conference, Austin, Texas, 24-26 July 2017*, pp. 580–600. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers. URTeC-2668849-MS.

Stukowski, A. (2010). Visualization and analysis of atomistic simulation data with OVITO–The Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering 18*(1), 015012.

Stukowski, A. (2014). Computational analysis methods in atomistic modeling of crystals. *The Journal of The Minerals, Metals & Materials Society 66*(3), 399–407.

Sun, H. (1998). COMPASS: An ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *Journal of Physical Chemistry B 102*(38), 7338–7364.

Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson (1982). A computer simulation method for the calculation of equilibrium-constants for the formation of

physical clusters of molecules – Application to small water clusters. *Journal of Chemical Physics 76*(1), 637–649.

Takbiri-Borujeni, A., M. Kazemi, S. Liu, and Z. Zhong (2019). Molecular simulation of enhanced oil recovery in shale. In J. Yan, H.-X. Yang, H. Li, and X. Chen (Eds.), *Innovative Solutions for Energy Transitions*, Volume 158 of *Energy Procedia*, pp. 6067–6072. Elsevier.

Tarbuck, E., F. Lutgens, and D. Tasa (2006). *Earth Science*. Pearson/Prentice Hall.

Teklu, T. W., N. Alharthy, H. Kazemi, X. Yin, R. M. Graves, and A. M. AlSumaiti (2014). Phase behavior and minimum miscibility pressure in nanopores. *SPE Reservoir Evaluation & Engineering 17*(3), 396–403. SPE-168865-PA.

Thijssen, J. M. (2007). *Computational Physics*. Cambridge University Press. Cambridge.

Tinni, A., C. Sondergeld, and C. S. Rai (2018). New perspectives on the effects of gas adsorption on storage and production of natural gas from shale formations. *Petrophysics 59*, 99–104. SPWLA-2018-v59n1a9.

Tissot, B. P. and D. H. Welte (1984). *Petroleum Formation and Occurrence* (Second ed.). Springer-Verlag Berlin Heidelberg GmbH.

Tovar, F. D., M. A. Barrufet, and D. S. Schechter (2018). Gas injection for EOR in organic rich shales. Part II: Mechanisms of recovery. In *SPE/AAPG/SEG Unconventional Resources Technology Conference, 23-25 July, Houston, TX*. Society of Petroleum Engineers. URTeC-2903026-MS.

Tuckerman, M. E. (2010). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press. New York.

Ungerer, P., J. Collell, and M. Yiannourakou (2015). Molecular modeling of the volumetric and thermodynamic properties of kerogen: Influence of organic type and maturity. *Energy & Fuels 29*, 91–105.

Verlet, L. (1967). Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review 159*(1), 98–103.

Wan, T., J. J. Sheng, and M. Y. Soliman (2013). Evaluation of the EOR potential in shale oil reservoirs by cyclic gas injection. In *SPWLA 54th Annual Logging Symposium*. Society of Petroleum Engineers. SPWLA-2013-MM.

Wang, L., Y. Tian, X. Yu, C. Wang, B. Yao, S. Wang, P. H. Winterfeld, X. Wang, Z. Yang, Y. Wang, J. Cui, and Y.-S. Wu (2017). Advances in improved/enhanced oil recovery technologies for tight and shale reservoirs. *Fuel 210*, 425–445.

Wang, T., S. Tian, G. Li, M. Sheng, W. Ren, Q. Liu, and S. Zhang (2018). Molecular simulation of CO_2/CH_4 competitive adsorption on shale kerogen for CO_2 sequestration and enhanced gas recovery. *Journal of Physical Chemistry C 122*, 17009–17018.

Wongkoblap, A. and D. D. Do (2008). Adsorption of polar and nonpolar fluids in finite-length carbon slit pore: A Monte Carlo simulation study. *Chemical Engineering Communications 195*(11), 1382–1395.

Yiannourakou, M., P. Ungerer, B. Leblanc, X. Rozanska, P. Saxe, S. Vidal-Gilbert, F. Gouth, and F. Montel (2013). Molecular simulation of adsorption in microporous materials. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles 68*(6), 977–994.

Zhang, T., G. S. Ellis, S. C. Ruppel, K. Milliken, and R. Yang (2012). Effect of organic-matter type and thermal maturity on methane adsorption in shale-gas systems. *Organic Geochemistry 47*, 120–131.

Zwanzig, R. (1965). Time-correlation functions and transport coefficients in statistical mechanics. *Annual Review of Physical Chemistry 16*(1), 67–102.

Appendix A

STATISTICAL MECHANICS AND MOLECULAR DYNAMICS SIMULATIONS

This appendix does not pretend to give a formal and exhaustive review of statistical mechanics, thermodynamics and molecular dynamics simulations. Rather, it attempts to connect the dots between them in a somewhat simplistic fashion for the reader that lacks a formal training in these subjects.

A.1 Introduction

Statistical mechanics, in a broad sense, is a theoretical framework that aims to predict the macroscopically observable static and dynamic properties of a many-body system starting from its microscopic constituents and their interactions (Tuckerman, 2010). The machinery of statistical mechanics seeks to establish a bridge between the microscopic and macroscopic realms, providing a means of rationalizing observed properties of a system in terms of the detailed motion of its fundamental constituents. The basic premise then is that the observed macroscopic properties arise at the most fundamental level from the interactions and dynamics of individual components (Krapf and Metzer, 2019). Statistical mechanics thus deals with systems with a very large number of degrees of freedom. The degrees of freedom are normally the positions $\{\mathbf{r_i}\} \equiv \{\mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_N}\}$ and

momenta $\{\mathbf{p_i}\} \equiv \{\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_N}\}$ of the N particles that make up the system. The $6N$-dimensional space spanned by the degrees of freedom is called phase space, and every point $\mathbf{\Gamma}(t) \equiv \{\mathbf{r_1}(t), \mathbf{r_2}(t), \ldots, \mathbf{r_N}(t), \mathbf{p_1}(t), \mathbf{p_2}(t), \ldots, \mathbf{p_N}(t)\}$ in phase space represents a particular configuration of the system.

In molecular simulations, the microscopic interaction rules of a system are implemented numerically on a computer, allowing the extraction of the underlaying mechanisms by which the system evolves in time, and the formalism of statistical mechanics provides the means to generate predictions of macroscopic observables. Therefore, as stated by Tuckerman (2010), *molecular simulations are the computational realization of statistical mechanics*, as long as there exists a rule governing the behavior of the basic constituents of the system.

A.2 How representative are molecular systems?

With the progression of time, the system follows a path in phase space that is determined by the equations of motion of the particles. Nonetheless, the knowledge of the position and momentum of every single particle is of little practical relevance; more interesting is the knowledge of time averages of observable physical quantities such as pressure and temperature[1]. The reason, as stated by Callen (1985), is that macroscopic measurements are tremendously slow on the atomic scale of time and tremendously coarse on the atomic scale of distance. While a macroscopic measurement is made, the constituents of a system move very rapidly and the measurement cannot respond to the incredibly large number of atomic coordinates that vary in time with typical atomic periods (on the order of femtoseconds, i.e. 10^{-15} s), effectively sensing only a coarse spatial average of atomic coordinates. Therefore, only the smaller set of combinations of atomic coordinates that are *time independent* are macroscopically observable. Such a set corresponds to *equilibrium states*, and the macroscopic theory that concerns these equilibrium states

[1]There are two types of thermodynamic variables: *extensive variables* are those that scale with the system size, for instance the volume V and the number of particles N; *intensive variables* are those that specify a local property and are independent of the system size, for instance the temperature T and the pressure P (Kondepudi and Prigogine, 2015).

is thermodynamics.

When a system is said to be in equilibrium, it goes from one state to another within the set of equilibrium states. If an instantaneous observation of some physical quantity could somehow be performed, the result would be very close to the time average of such quantity as a result of the law of very large numbers which states that if a quantity is composed of N uncorrelated contributions, the magnitude of its fluctuations is of order $1/\sqrt{N}$. Thus, for typical macroscopic systems for which $N \sim 10^{24}$, the fluctuations are on the order of 10^{-12} if the particles are not correlated. If there are correlations that extend over 100 particles, the number of uncorrelated contributions to the physical quantity is $10^{24}/10^2 = 10^{22}$ and fluctuations remain very small.

Nonetheless, even for the most powerful supercomputers, computer simulations are limited to sampling a small number of degrees of freedom. In 2019, researchers at Los Alamos National Laboratory created the largest molecular simulation to this date of an entire gene of DNA comprising one billion atoms using the Trinity supercomputer (Jung et al., 2019), the sixth fastest in the world[2], thus the size of the largest system modeled up to this day is on the order of 10^9 particles. Hence, system sizes in simulations are much smaller than those of experimental (macroscopic) systems. Furthermore, small systems are dominated by surface effects because the interactions of atoms with the container walls easily extend 4 to 10 atomic diameters from each wall (Haile, 1992). For instance, in a system with $10^3 = 1000$ particles arranged in a cubic box, $8^3 = 512$ particles lie in the interior and 488 (49% of the total) lie on the surfaces; if the system contains 100^3 particles, 94% lie in the interior while 6% lie on the surfaces. The particles at and close to the surface will experience quite different forces from those in bulk (Allen and Tildesley, 2017).

The use of periodic boundary conditions removes unwanted surface effects. The idea is to consider the system of N particles confined to a volume V be only a small portion of the bulk material, which will be composed of an infinite number of exact replicas of the volume V, called primary cell, in all spatial directions. The replicas are called image

[2]As of April 22, 2019.

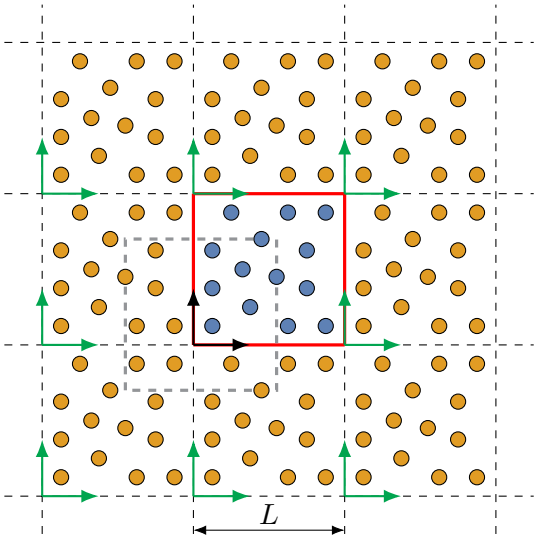


Figure A.1: Periodic boundary conditions in two dimensions. The red box represents the primary cell. The yellow balls are image particles in image cells. Each cell has a reference frame located in one of the corners. The primary frame (frame in the primary cell) is black and all others are image frames in green. The dashed gray box could have been the primary cell just as well, illustrating that the choice is purely arbitrary.

cells, all have the same size and shape as the primary cell and each contains N atoms that are images of the atoms in the primary cell, as illustrated in Figure A.1. Cells are separated by open boundaries (i.e. boundaries that do not exist physically and therefore their locations are entirely arbitrary), thus particles and images can freely enter or leave any cell, but the number of particles in each cell is always N because when particle i leaves the primary cell through a particular face, an image of i simultaneously reenters the primary cell at the same point of the face directly opposite. The question is whether the properties of a small, infinitely periodic system and the macroscopic system being represented are the same. This ultimately depends both on the range of the intermolecular potential $v(r)$ (discussed in Section A.5) and the phenomenon under investigation (Allen and Tildesley, 2017).

In addition to spatial considerations, the time average of a physical quantity θ is given by

$$\bar{\theta} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt\, \theta(t) \tag{A.1}$$

where t_0 is the starting time and τ is the duration of the measurement, but simulations cannot run for an infinite amount of time. For instance, in molecular dynamics simulations

(which will be discussed in Section A.5) the typical simulation time is on the order of nanoseconds (i.e. 10^{-9} s), microseconds (i.e. 10^{-6} s) at most.

Despite of these limitations, the outcomes from molecular dynamics are representative as long as and only if the spatial correlations extend over ranges smaller than the system size and the correlation time of the system is shorter than the simulation time (Thijssen, 2007). In the case of long-range potentials (i.e. $v(r) \sim r^{-\nu}$ where ν is less than the dimensionality of the space), there will be a substantial interaction between a particle and its own images in image cells and consequently the symmetry of the primary cell is imposed on a system that might be in reality isotropic (Allen and Tildesley, 2017). In the case of correlation time, it can be defined as the time scale over which a physical quantity θ relaxes towards its equilibrium value $\bar{\theta}$, in other words it is the time scale over which a system looses memory of initial condition and the trajectories in phase space generated by the simulation become essentially indistinguishable from any other trajectory (Tuckerman, 2010). It is of the highest important because for systems in thermodynamic equilibrium the average $\bar{\theta}$ must be independent of the location of the starting time t_0, i.e. of the initial conditions. Furthermore, the standard deviation of $\bar{\theta}$ calculated over M samples generated by the simulation depends on the number of samples that are effectively independent, and this quantity is the total number of samples divided by the correlation time (measured in simulation steps) (Thijssen, 2007). If the simulation time is less than or equal to the correlation time, there are no uncorrelated samples and the average value $\bar{\theta}$ is rendered meaningless.

A.3 The N-body problem

The motion of a point particle with mass m subject to a force $\mathbf{F}$ that depends on the particle position $\mathbf{r}(t)$ and perhaps on its velocity $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$ and on time t is determined by the Newton's equation of motion

$$\mathbf{F}(\mathbf{r}(t), \dot{\mathbf{r}}(t), t) = m\ddot{\mathbf{r}}(t). \tag{A.2}$$

The dot over a variable denotes a time derivative, thus one dot corresponds to d/dt, two dots to d^2/dt^2, and so on. An analytic solution can be found for each set of initial conditions $\mathbf{r}(t_0) \equiv \mathbf{r_0}$ and $\dot{\mathbf{r}}(t_0) \equiv \dot{\mathbf{r}}_\mathbf{0}$ given at some time t_0.

In the case of two point particles interacting through a central force, i.e. a force that lies along the line that connects the two particles, there also exists an analytical solution (Landau and Lifshitz, 1976; Goldstein, 1980). However, for the case of three particles, analytical solutions exist for particular configurations but the general problem can only be solved numerically (Thijssen, 2007). In general, the N-body problem requires numerical techniques to approximate solutions to the equations of motion

$$\mathbf{F_i}(\mathbf{\Gamma}(t)) = m_i \ddot{\mathbf{r}}_\mathbf{i}(t), \quad i = 1, 2, \ldots, N \tag{A.3}$$

of the system.

A.3.1 Atomic and molecular systems

Accurate simulations of atomic and molecular systems require the use of the formalism of quantum mechanics, although its techniques are currently computationally demanding and are efficiently applicable only to small systems containing a few tens of atoms or small molecules. Nonetheless, and in the same spirit of the discussion above, even if these simulations were possible to carry out, much of the information generated would be of little practical interest from a macroscopic point of view. Consequently, the details of electronic motion, for instance, are not particularly interesting and are lost in the averaging process. This idea suggests the possibility to separate the motion of electrons and nuclei in atomic systems.

A.3.2 The Born-Oppenheimer approximation

In quantum mechanics, a system is characterized by its degrees of freedom and by the Hamiltonian $\mathcal{H}$ which assigns an energy to every possible state of the system. By state of the system it is understood a specific set of values of the degrees of freedom, i.e.

a specific point in phase space. The stationary states of the system are determined by the time-independent Schrödinger equation (Landau and Lifshitz, 1977)

$$\mathcal{H}\Psi(\mathbf{r},\mathbf{x}) = \mathcal{E}\Psi(\mathbf{r},\mathbf{x}) \tag{A.4}$$

where Ψ is its wavefunction and $\mathcal{E}$ is its energy. Generally, Ψ is a function of the coordinates of the N nuclei $\mathbf{r} \equiv \{\mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_N}\}$ and of the N' electrons $\mathbf{x} \equiv \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_{N'}}\}$.

Born and Oppenheimer (1927) decoupled the motion of electrons and nuclei based on the large difference between electric and nuclear masses and therefore the much faster motion of electrons as compared to that of nuclei, and gave two separate equations. The first one describes the electronic motion

$$\mathcal{H}\Psi(\mathbf{x};\mathbf{r}) = \mathcal{E}(\mathbf{r})\Psi(\mathbf{x};\mathbf{r}) \tag{A.5}$$

which depends parametrically on the positions of the nuclei $\mathbf{r}$, and $\mathcal{E}(\mathbf{r})$ is called the potential energy surface. The solution of (A.5) is the so-called *ab initio* calculation. The second equation describes the motion of the nuclei on such potential energy surface

$$\mathcal{H}\Phi(\mathbf{r}) = \mathcal{E}(\mathbf{r})\Phi(\mathbf{r}) \tag{A.6}$$

where Φ is the the wavefunction of the nuclei.

The solution of (A.6) provides the means to compute the structure and evolution of a system. In principle, (A.5) can be solved for the potential energy $\mathcal{E}(\mathbf{r})$, assuming that the nuclei locations are fixed and the fast motion of the electrons yields an electronic distribution and an average potential energy, next (A.6) is solved by releasing the locations of nuclei, allowing them to move under the effect of the average potential energy, leading to molecular vibrations and rotations[3] (Bransden and Joachain, 1983; Campos, 1997).

Finding the solution to (A.5) requires a tremendous computational effort. Instead, empirical fits to the potential energy surface $\mathcal{E}(\mathbf{r})$ using ab initio calculations or exper-

[3]This is called *adiabatic approximation* in molecular physics.

imental data, the so-called *force fields*, can be utilized. Furthermore, the heavy mass of the nuclei in the system often times renders quantum mechanical effects insignificant, thus (A.6) can be replaced by the Newton's equation of motion (A.3) and the problem can be treated as a classic N-body problem.

A.4 Statistical ensembles

In the course of an experiment there are certain macroscopic parameters that are kept fixed or whose average values are controlled externally. Different experimental conditions correspond to different parameters being kept fixed. In the theory of statistical mechanics these cases are associated with different ensembles. To explain the concept of statistical ensemble, consider again the problem of solving the equations of motion (A.3) for a system of 10^{23} particles with the complex, nonlinear interactions that govern the behavior of real systems. Even if it could be somehow done, the vast amount of detailed microscopic information is not necessary to describe macroscopic observable quantities.

Although all that information is certainly sufficient to predict any macroscopic observable, there are many microscopic configurations of a system that yield the same macroscopic observable quantity. For example, the temperature of a system is related to the average kinetic energy of each of its constituents (see Section A.5.2), and there are many ways to assign the velocities of the particles consistent with a specified total energy so that the same total kinetic energy and, therefore, the same measure of temperature is obtained. Now, each assignment corresponds to a different point in phase space, i.e. a different and unique microscopic state. The trajectories described if the system is launched from each one of these assignments are each unique in phase space, but in the long time limit all these trajectories should lead to the same macroscopic dynamical observables. Therefore, macroscopic measurable quantities of a system are not sensitive to the precise microscopic details.

Originally introduced by J. Willard Gibbs in 1902, an ensemble is a collection of systems described by the same set of microscopic interactions and sharing a common set

of macroscopic properties (Tuckerman, 2010). Each system evolves uniquely in phase space subject to the microscopic laws of motion from a different initial condition. Once an ensemble is defined, macroscopic observable quantities are computed by carrying out averages over the systems in the ensemble, the so-called ensemble averages.

There exist essentially two numerical techniques for determining statistical averages over a finite and rather restricted set of states in an ensemble: the molecular dynamics and Monte Carlo methods[4]. In molecular dynamics, the system moves in phase space along its physical trajectory as determined by its equations of motion as discussed in Section A.5, while in Monte Carlo the system follows a directed random walk. Adapting the simulation techniques to the particular experimental situations is a nontrivial problem (Thijssen, 2007), as will be briefly discussed in Section A.5.2.

A.4.1 The microcanonical ensemble

The fundamental assumption of statistical mechanics, the ergodic theorem, states that in a system with fixed internal energy[5] E, volume V, and number of particles N, all states accessible to the system are equally likely to be visited in a sufficiently long time, or equivalently, the locus $\mathbf{\Gamma}$ will cover the entire accessible phase space (Huang, 1987). Thus, there is a correspondence between the time average of a physical quantity θ given in (A.1) with a uniform average over all accessible states (ensemble average) denoted as $\langle\theta\rangle$. Let $\{\mathbf{\Gamma}_{\mathrm{NVE}}\}$ be the set of all accessible states. Such a set is called microcanonical ensemble. The ensemble average is given by

$$\langle\theta\rangle = \frac{\sum_{\{\mathbf{\Gamma}_{\mathrm{NVE}}\}} \theta(\mathbf{\Gamma})}{\sum_{\{\mathbf{\Gamma}_{\mathrm{NVE}}\}}} = \frac{\sum_{\mathbf{\Gamma}} \theta(\mathbf{\Gamma})\delta[\mathcal{H}(\mathbf{\Gamma}) - E]}{\sum_{\mathbf{\Gamma}} \delta[\mathcal{H}(\mathbf{\Gamma}) - E]} = \bar{\theta} \tag{A.7}$$

where the denominator counts the number of states with the prescribed internal energy E, is called degeneracy, and is denoted by $\Omega(N, V, E)$. The function δ represents the Dirac delta function, whose value is zero unless its argument is zero. For readers not

[4]Molecular dynamics and Monte Carlo methods sample (or generate) a set of states in the ensemble, i.e. the states are consistent with the macroscopic conditions of interest.

[5]The internal energy of an isolated closed system is all energy not associated with the center of mass degrees of freedom (such as translation of the constituents).

familiar with the Dirac delta function, excellent reviews are provided in Arfken (1985) and Byron, Jr. and Fuller (1992).

In quantum mechanics, identical particles are indistinguishable as a matter of principle and therefore configurations that are permutations of each other should be counted only once, which implies that the sum in the denominator of (A.7) should be divided by $N!$ (or $N_1!N_2!\ldots N_k!$ in the case of a mixture of k components). However, although classical simulations deal with distinguishable particles, it has been shown from first principles that the factor $N!$ should be present for a system of N very similar but distinguishable particles, in order to obtain extensive entropy and Helmholtz free energy (discussed below) functions, and hence it is not intimately connected with quantum mechanics (Allen and Tildesley, 2017). Thus

$$\Omega(N, V, E) = \frac{1}{N!} \sum_{\mathbf{\Gamma}} \delta[\mathcal{H}(\mathbf{\Gamma}) - E]. \tag{A.8}$$

Consider now a system with total energy E that consists of two weakly interacting subsystems[6]. There are many ways in which the total energy E can be distributed over the two subsystems such that $E_1 + E_2 = E$. For a choice of E_1, the total number of states is $\Omega_1(E_1)\Omega_2(E_2)$, which is not the sum but the product of the number of states in the individual subsystems. It is more convenient to have a measure of the degeneracy of the subsystems that is additive. Taking the natural logarithm of the total number of states,

$$\ln \Omega(E_1, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E - E_1). \tag{A.9}$$

This choice has one more advantage. The degeneracy of a system is a very rapid growing function of its arguments, and the numbers become very large very quickly. The property of logarithms $\ln a^b = b \ln a$ makes the numbers more tractable from an operative point of view.

[6]Weakly interacting in this context means that the subsystems can exchange energy but the total energy of the whole system can be written as the sum of the energies E_1 and E_2 of the subsystems.

The entropy is defined in terms of the degeneracy $\Omega(N, V, E)$ as

$$S(N, V, E) = k_B \ln \Omega(N, V, E) \tag{A.10}$$

where $k_B = 1.381 \times 10^{-23}$ J/K is the Boltzmann constant. The entropy is an extensive function (i.e. it also scales with the system size) of the extensive parameters N, V, E which, in equilibrium conditions and in the absence of internal constraints, assume values that maximize the entropy over the set of equilibrium states (Callen, 1985). The relation that gives the entropy as a function of the extensive parameters is regarded as a fundamental thermodynamic relation[7] because it encloses all conceivable thermodynamic information about the system under consideration (Callen, 1985).

In experimental processes, the extensive parameters N, V and E change and therefore so does the entropy S too. Thus

$$dS = \left(\frac{\partial S}{\partial E}\right)_{V,N} dE + \left(\frac{\partial S}{\partial V}\right)_{E,N} dV + \left(\frac{\partial S}{\partial N}\right)_{E,V} dN. \tag{A.11}$$

The partial derivatives appearing in (A.11) define the intensive variables temperature T, pressure P, and chemical potential μ, respectively:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{V,N}, \tag{A.12a}$$

$$\frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{E,N}, \tag{A.12b}$$

$$-\frac{\mu}{T} = \left(\frac{\partial S}{\partial N}\right)_{E,V}. \tag{A.12c}$$

The definition of entropy given in (A.10) provides a bridge between the microscopic description and macroscopic thermodynamic observables defined in (A.12).

The fundamental thermodynamic relation can be also written in the energy repre-

[7] The relation $S = S(N, V, E)$ is called the fundamental thermodynamic relation in the entropy representation.

sentation as $E = E(S, V, N)$, thus variations in the internal energy are

$$dE = \left(\frac{\partial E}{\partial S}\right)_{V,N} dS + \left(\frac{\partial E}{\partial V}\right)_{S,N} dV + \left(\frac{\partial E}{\partial N}\right)_{S,V} dN. \tag{A.13}$$

and the partial derivatives define the intensive parameters

$$T = \left(\frac{\partial E}{\partial S}\right)_{V,N}, \tag{A.14a}$$

$$-P = \left(\frac{\partial E}{\partial V}\right)_{S,N}, \tag{A.14b}$$

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{S,V}. \tag{A.14c}$$

With the definitions in (A.14), the first law of thermodynamics is readily recovered (Callen, 1985; Kondepudi and Prigogine, 2015):

$$dE = TdS - PdV + \mu dN. \tag{A.15}$$

The temperature, pressure, and chemical potential are partial derivatives of functions of S, V, N and are therefore also functions of S, V, N. The set of relationships $T = T(S, V, N)$, $P = P(S, V, N)$, and $\mu = \mu(S, V, N)$ that express intensive parameters in terms of the independent extensive parameters are called *equations of state*[8] (Callen, 1985).

A.4.2 The canonical ensemble

Often times, the condition controlled experimentally is the temperature rather than the energy. To achieve this, the system is coupled to a heat bath, which is a much larger system with which it is allowed to exchange heat. In this case, the fixed parameters are temperature T, volume V, and number of particles N. Let $\{\mathbf{\Gamma}_{\mathrm{NVT}}\}$ be the set of all accessible states consistent with the constraints aforementioned. Such a set is called canonical ensemble.

[8]This definition differs from the common idea that relations such as the ideal gas equation $PV = Nk_BT$ is an equation of state.

The time average of a physical quantity θ is equal to a weighted average over states with fixed volume V and number of particles N (the energy E is now allowed to vary), being the weighting factor the so-called Boltzmann factor $\mathrm{e}^{-\mathcal{H}(\mathbf{\Gamma})/k_BT}$. Writing $\beta \equiv 1/k_BT$,

$$\langle\theta\rangle_{NVT} = \frac{1}{N!Z}\sum_{\mathbf{\Gamma}}\theta(\mathbf{\Gamma})\,\mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} \tag{A.16}$$

where

$$Z(N,V,T) = \frac{1}{N!}\sum_{\mathbf{\Gamma}}\mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} \tag{A.17}$$

is called the partition function, which is related to the Helmholtz free energy F by

$$F(N,V,T) = -k_BT\ln Z(N,V,T) \tag{A.18}$$

much in the same way as the entropy S is related to the degeneracy $\Omega(N,V,E)$ in (A.10). The average in (A.16) is called the canonical ensemble average.

The partition function can be also written as a sum over sets of states with fixed energy:

$$\begin{aligned} Z(N,V,T) = \frac{1}{N!}\sum_{\mathbf{\Gamma}}\mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} &= \frac{1}{N!}\sum_{\mathbf{\Gamma}}\sum_{E}\mathrm{e}^{-\beta E}\,\delta[\mathcal{H}(\mathbf{\Gamma})-E] \\ &= \sum_{E}\mathrm{e}^{-\beta E}\,\frac{1}{N!}\sum_{\mathbf{\Gamma}}\delta[\mathcal{H}(\mathbf{\Gamma})-E] = \sum_{E}\mathrm{e}^{-\beta E}\,\Omega(N,V,E). \end{aligned} \tag{A.19}$$

The degeneracy $\Omega(N,V.E)$ increases very rapidly with the energy E, whereas the Boltzmann factor is a rapidly decreasing function of E. The product of the two terms peaks sharply at some value $\bar{E}$ and the energy of the system fluctuates in time about this value. This observation indicates that there is not much difference between the canonical and the microcanonical system whose energy is fixed at $\bar{E}$, and is a manifestation of the ensemble equivalence which states that by virtue of the law of large numbers, the fluctuations in measurable physical quantities are small and fixing them to their average values leaves the system unchanged. For finite systems as the ones considered in molecular simulations, however, the differences between ensembles increase with decreasing system

size (Thijssen, 2007).

Digression on Legendre transformations

It was mentioned in Section A.4.1 that the internal energy E, as a function of the extensive parameters S, V, N, is a fundamental relation as it encloses all thermodynamic information about the system. The same is true for the entropy S as a function of the extensive variables N, V, E. However, for practical considerations, it is desirable to recast the mathematical formalism in such a way that intensive parameters replace extensive ones as independent variables without sacrificing any of the informational content of the given fundamental relation.

It is not a trivial problem whose solution is provided by the mathematical technique of Legendre transformations. Intuitively, consider a fundamental relation $y = f(x)$ that depends on only one independent variable x. Geometrically, the fundamental relation is represented by a curve in the xy-plane, as shown in Figure A.2(a). The derivative $m(x) = dy/dx = g(x)$ is the slope of the curve at any given abscissa x. If now m is to be considered as the independent variable in place of x, the naive way to attempt this is to eliminate x so that $y = f(g^{-1}(m))$. This approach, however, sacrifices some mathematical information because, from a geometrical point of view, the knowledge of y as a function of the slope m does not permit the unequivocal reconstruction of the curve $y = f(x)$, as illustrated by the displaced curves in Figure A.2(b). From an analytical point of view, $y = y(m)$ is a first-order differential equation whose integration gives $y = y(x)$ only to within an arbitrary integration constant.

The solution is supplied by the Pluecker line geometry, whose essential concept is that a given curve can be represented either as the envelope of a family of tangent lines (yellow lines in Figure A.3) or as the locus of points that satisfy the relation $y = f(x)$ (the blue curve in Figure A.3) (Callen, 1985).

Therefore, the mathematical expression to construct the family of tangent lines determines the curve just as well as the relation $y = f(x)$, in other words all mathematical information is preserved. Thus, just as every point in the plane xy is described by the

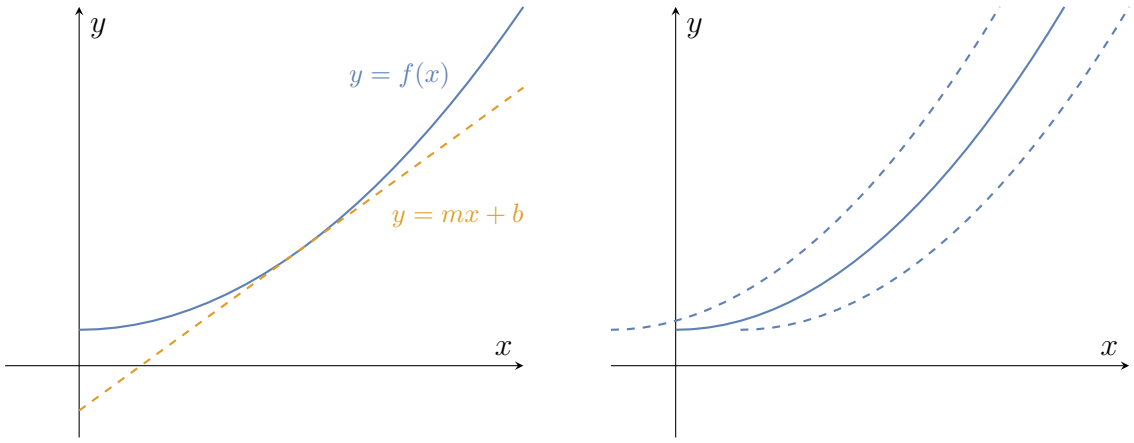


Figure A.2: Geometrical interpretation of Legendre transformations.

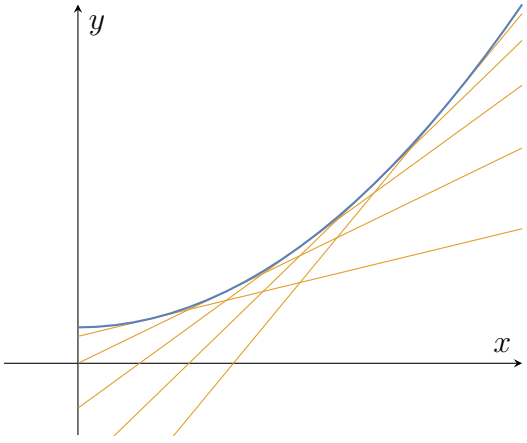


Figure A.3: Pluecker line geometry.

tuple (x, y), so every straight line in the plane can be described by its slope m and its intercept b with the y-axis, and just as the relation $y = f(x)$ selects a subset of all possible points (x, y), so the relation $b = b(m)$ selects a subset of all possible straight lines. This relation carries the same information as the fundamental relation $y = f(x)$.

The equation of a straight line is $y = mx + b$, hence

$$b = y - mx. \tag{A.20}$$

Assuming the fundamental relation

$$y = f(x) \tag{A.21}$$

is known, then the slope m can be calculated as

$$m = \frac{dy}{dx}, \tag{A.22}$$

x and y can be eliminated among (A.20), (A.21), and (A.22), and the relation $b = b(m)$ is obtained. (A.20) is the analytical definition of the Legendre transformation, and b is referred to as a Legendre transform of y. ■

The digression above allows to find the fundamental relation in the case of fixed N, V, T parameters, one of which (the temperature T) is intensive. According to (A.14a), $T = (\partial E/\partial S)_{V,N}$, thus the (partial) Legendre transform of E with respect to S is

$$F(N, V, T) = E - S\left(\frac{\partial E}{\partial S}\right)_{N,V} = (E - TS)_{S \to S(N,V,T)}. \tag{A.23}$$

In equilibrium conditions, the Helmholtz free energy assumes its minimum value (the entropy assumes its maximum value) under fixed volume, temperature, and number of particles, while energy fluctuations are non-zero. Moreover, the Helmholtz free energy is not just the internal energy E expressed in terms of the variables N, V, T but the fundamental relation under such conditions, and its significance is that it describes implicitly the role played by the heat reservoir, allowing to focus the attention only on

the (sub)system of interest.

The thermodynamic quantities entropy S, pressure P, and chemical potential μ can be related to the Helmholtz free energy. Since $F = F(N, V, T)$,

$$dF = \left(\frac{\partial F}{\partial T}\right)_{V,N} dT + \left(\frac{\partial F}{\partial V}\right)_{N,T} dV + \left(\frac{\partial F}{\partial N}\right)_{V,T} dN. \tag{A.24}$$

From (A.23) and the first law of thermodynamics (A.15),

$$dF = dE - TdS - SdT = -SdT - PdV + \mu dN. \tag{A.25}$$

Comparing (A.24) and (A.25),

$$-S = \left(\frac{\partial F}{\partial T}\right)_{N,V}, \tag{A.26a}$$

$$-P = \left(\frac{\partial F}{\partial V}\right)_{N,T}, \tag{A.26b}$$

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{V,T}. \tag{A.26c}$$

A.4.3 The isothermal-isobaric ensemble

In addition to controlling the temperature using a heat bath, the pressure can be also kept constant by allowing changes in the volume, for instance using a cylinder closed by a movable piston. Let $\{\mathbf{\Gamma}_{\mathrm{NPT}}\}$ be the set of all accessible states consistent with the constraints aforementioned. Such a set is called isothermal-isobaric ensemble.

The time average of a physical quantity θ is given by

$$\langle\theta\rangle_{NPT} = \frac{1}{N!Q} \int dV \ \mathrm{e}^{-\beta PV} \sum_{\mathbf{\Gamma}} \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})}\, \theta(\mathbf{\Gamma}) \tag{A.27}$$

where

$$Q(N, P, T) = \int dV \ \mathrm{e}^{-\beta PV}\, \frac{1}{N!} \sum_{\mathbf{\Gamma}} \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} = \int dV \ \mathrm{e}^{-\beta PV}\, Z(N, V, T) \tag{A.28}$$

is the partition function, which is related in the microscopic realm to the canonical partition function $Z(N,V,T)$ in much the same way as $Z(N,V,T)$ is related to the microcanonical degeneracy $\Omega(N,V,E)$ in (A.19), and to the Gibbs free energy (also known as Gibbs potential) in the macroscopic realm by

$$G(N,P,T) = -k_B T \ln Q(N,P,T), \tag{A.29}$$

similar to the bridges presented in (A.10) and (A.18) between the microscopic description and macroscopic thermodynamic observables. The average in (A.27) is called the isothermal-isobaric ensemble average.

The fundamental relation in this case is found using the (partial) Legendre transformation of E in terms of S and V, such that

$$G(N,P,T) = E - S\left(\frac{\partial E}{\partial S}\right)_{N,V} - V\left(\frac{\partial E}{\partial V}\right)_{N,S} = (E - TS + PV)_{\substack{S\to S(N,P,T)\\ V\to V(N,P,T)}}. \tag{A.30}$$

which assumes its minimum value when the system reaches equilibrium under fixed temperature, pressure, and number of particles.

The thermodynamic quantities entropy S, volume V, and chemical potential μ can be related to the Gibbs potential. Since $G = G(N,P,T)$,

$$dG = \left(\frac{\partial G}{\partial T}\right)_{P,N} dT + \left(\frac{\partial G}{\partial P}\right)_{N,T} dP + \left(\frac{\partial G}{\partial N}\right)_{P,T} dN. \tag{A.31}$$

From (A.30) and the first law of thermodynamics (A.15),

$$dG = dE - TdS - SdT + PdV + VdP = -SdT + VdP + \mu dN. \tag{A.32}$$

Comparing (A.31) and (A.32),

$$-S = \left(\frac{\partial G}{\partial T}\right)_{N,P}, \tag{A.33a}$$

$$V = \left(\frac{\partial G}{\partial P}\right)_{N,T}, \tag{A.33b}$$

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{P,T}. \tag{A.33c}$$

A.4.4 The grand canonical ensemble

Sometimes, the volume and temperature of a system are kept constant while the number of particles is allowed to vary. In addition to coupling the system to a heat bath, particles can flow now across the system boundaries. Once the system reaches equilibrium, neither flow of heat nor of particles occurs across the system boundaries. The temperature T in the bath and the system is equalized, so the heat flux ceases. Similarly, the chemical potential μ in the reservoir and the system is equalized and the particle flux ceases. Let $\{\mathbf{\Gamma}_{\mu \mathrm{VT}}\}$ be the set of all accessible states consistent with the constraints aforementioned. Such a set is called grand canonical ensemble.

The time average of a physical quantity θ is given by

$$\langle\theta\rangle_{\mu VT} = \frac{1}{\mathcal{Z}} \sum_N \mathrm{e}^{\beta\mu N} \frac{1}{N!} \sum_{\mathbf{\Gamma}} \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})}\, \theta(\mathbf{\Gamma}) \tag{A.34}$$

where

$$\mathcal{Z}(\mu, V, T) = \sum_N \mathrm{e}^{\beta\mu N} \frac{1}{N!} \sum_{\mathbf{\Gamma}} \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} = \sum_N \mathrm{e}^{\beta\mu N}\, Z(N, V, T) \tag{A.35}$$

is the grand canonical partition function. As in the previous cases, it is related to the canonical partition function $Z(N, V, T)$ in a similar fashion as $Z(N, V, T)$ is related to the microcanonical degeneracy $\Omega(N, V, E)$ in (A.19). Furthermore, it defines the grand canonical potential Ξ

$$\Xi(\mu, V, T) = -k_B T \ln \mathcal{Z}(\mu, V, T), \tag{A.36}$$

analogous to the expressions in (A.10), (A.18), and (A.29). The average in (A.34) is

called the grand canonical ensemble average.

The fundamental relation in this case is found using the (partial) Legendre transformation of E in terms of S and N, such that

$$\Xi(\mu, V, T) = E - S\left(\frac{\partial E}{\partial S}\right)_{N,V} - N\left(\frac{\partial E}{\partial N}\right)_{S,V} = (E - TS - \mu N)_{\substack{S\to S(\mu,V,T)\\ N\to N(\mu,V,T)}}. \tag{A.37}$$

which assumes its minimum value when the system reaches equilibrium under fixed temperature, volume, and chemical potential.

The thermodynamic quantities entropy S, pressure P, and number of particles N can be related to the grand canonical potential. Since $\Xi = \Xi(\mu, V, T)$,

$$d\Xi = \left(\frac{\partial \Xi}{\partial T}\right)_{\mu,V} dT + \left(\frac{\partial \Xi}{\partial V}\right)_{\mu,T} dV + \left(\frac{\partial \Xi}{\partial \mu}\right)_{V,T} d\mu. \tag{A.38}$$

From (A.37) and the first law of thermodynamics (A.15),

$$d\Xi = dE - TdS - SdT - \mu dN - Nd\mu = -SdT - PdV - Nd\mu. \tag{A.39}$$

Comparing (A.38) and (A.39),

$$-S = \left(\frac{\partial \Xi}{\partial T}\right)_{\mu,V}, \tag{A.40a}$$

$$-P = \left(\frac{\partial \Xi}{\partial V}\right)_{\mu,T}, \tag{A.40b}$$

$$-N = \left(\frac{\partial \Xi}{\partial \mu}\right)_{V,T}. \tag{A.40c}$$

A.5 Molecular dynamics simulations

The bottom line in the previous section is that experimental values of physical observables in many-particle systems can be found as ensemble averages. However, experimental systems are so large that it is impossible to compute ensemble averages by summing over *all* accessible states $\mathbf{\Gamma}$ in a computer. Rather, Monte Carlo and molecular

dynamics simulations generate a *sample* consisting of a large number of configurations that are consistent with the conditions of the ensemble of interest.

Molecular dynamics is a popular method for studying classical many-particle systems, which essentially integrates numerically the equations of motion (A.3) of the particles in the system. It can be regarded as a simulation of the temporal evolution of the system. The great advantage of molecular dynamics is that it provides not only the means to evaluate expectation values of static quantities but also to study dynamical phenomena.

Consider a system whose microscopic state may be specified in terms of the positions and momenta of a constituent set of particles, i.e. the atoms and molecules. As discussed in Section A.3.2, within the Born-Oppenheimer approximation it is possible to average out the fast motion of the electrons and write the Hamiltonian $\mathcal{H}$ of the system in terms of nuclear variables. With the additional assumption that the classical description is adequate, the Hamiltonian of a system of N particles[9] may be written as a sum of the kinetic and potential energy functions, $\mathcal{K}$ and $\mathcal{V}$, respectively, of the coordinates $\{\mathbf{r_i}\}$ and momenta $\{\mathbf{p_i}\}$. Thus

$$\mathcal{H}(\mathbf{\Gamma}) = \mathcal{K}(\{\mathbf{p_i}\}) + \mathcal{V}(\{\mathbf{r_i}\}) \tag{A.41}$$

where it is explicitly stated that the kinetic energy $\mathcal{K}$ is a function only of the momenta $\{\mathbf{p_i}\}$ whereas the potential energy $\mathcal{V}$ is a function only of the coordinates $\{\mathbf{r_i}\}$.

For a simple atomic system, the kinetic energy $\mathcal{K}$ takes the form

$$\mathcal{K}(\{\mathbf{p_i}\}) = \sum_{i=1}^{N} \sum_{\alpha=1}^{3} \frac{p_{i\alpha}^2}{2m_i} \tag{A.42}$$

where m_i is the mass of particle i and the index α runs over the components (x, y, z). The potential energy $\mathcal{V}$ contains information about interparticle interactions. It may be divided into terms depending on the coordinates of individual particles, pairs, triplets,

[9]For the sake of simplicity in the introduction of the main ideas, it is assumed that the system is composed of atoms or small molecules that can be represented by their center of mass. When the compounds are large molecules, intramolecular interactions such as bonding, bending and torsions must be also included.

and so on as

$$\mathcal{V}(\{\mathbf{r_i}\}) = \sum_i v_1(\mathbf{r_i}) + \sum_i \sum_{j>i} v_2(\mathbf{r_i}, \mathbf{r_j}) + \sum_i \sum_{j>i} \sum_{k>j} v_3(\mathbf{r_i}, \mathbf{r_j}, \mathbf{r_k}) + \dots \tag{A.43}$$

where the sums avoid double counting of any pairs, triplets, etc. The first term in (A.43) represents the effect of an external field on the system (including the walls of the container), and the remaining terms describe particle interactions. The second term is called pair potential and depends only on the magnitude of the separation $r_{ij} = |\mathbf{r_{ij}}| \equiv |\mathbf{r_i} - \mathbf{r_j}|$ between the pair of particles. The third term involving triplets of particles is significant at liquid densities, and terms of higher level of particle interactions are expected to be small compared to v_2 and v_3 (Allen and Tildesley, 2017).

In most cases, the pairwise approximation yields notably good descriptions of liquid properties given that the average three-body effects can be partially included by defining an effective pair potential v_2^{eff} such that (A.43) can be recast as

$$\mathcal{V}(\mathbf{r}) \approx \sum_i v_1(\mathbf{r_i}) + \sum_i \sum_{j>i} v_2^{\text{eff}}(r_{ij}). \tag{A.44}$$

Perhaps the simplest and most common pair potential used in computer simulations is the Lennard-Jones 12-6 potential

$$v^{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]. \tag{A.45}$$

This form of the potential has a long-range attractive tail of the form $1/r^6$, a well of depth ϵ, and a steeply rising repulsive wall at distances $r < \sigma$, as schematically shown in Figure A.4. If the particles are electrically charged, the Coulomb long-range interaction

$$v^{\text{C}}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r} \tag{A.46}$$

is also included, where q_i, q_j are the charges of the particles and ε_0 is the permitivity of free space.

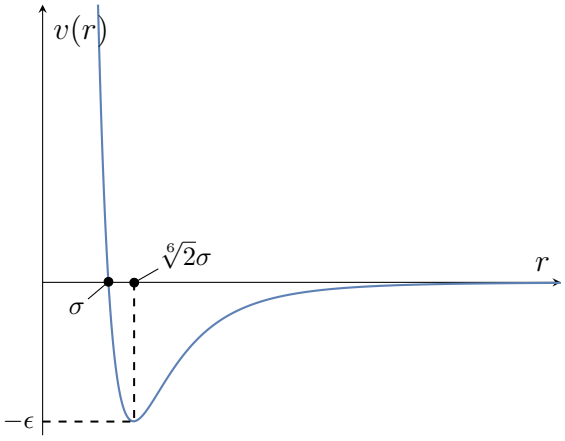


Figure A.4: Lennard-Jones 12-6 potential.

A.5.1 Integrating the equations of motion

There are three fundamental aspects to a molecular dynamics simulation: (1) the model that describes the interparticle interactions, (2) the accurate and efficient computation of energies and forces from the model, and (3) the algorithm used to integrate the equations of motion (Tuckerman, 2010). The election of the model to describe the interparticle interactions (i.e. the force field) is crucial at the moment of designing a simulation since the quality of the predictions will entirely depend on the adequacy of the force field to represent the particles in the system. The calculation of the force acting on every particle in a system with pairwise additive interactions requires to consider the contribution to the force on particle i due to all its neighbors, which for a system of N particles corresponds to evaluating $N(N-1)/2$ pair distances, thus the calculation scales as N^2 and is the most time consuming step in the simulation, limiting its ability to sample a sufficiently large number of microstates to obtain reliable averages. Fortunately, there exist efficient techniques for speeding up the evaluation of both short-range (such as the Lennard-Jones potential) and long-range (such as Coulombic) interactions so that the computing time scales as $N^{3/2}$ rather than N^2. Excellent reviews are presented in Tuckerman (2010) and Frenkel and Smith (2002).

The simplest way to obtain a numerical integration scheme is to use a Taylor series. The position of particle i at time $t + \Delta t$ is expressed in terms of the its position, velocity

and acceleration at time t as

$$\mathbf{r_i}(t+\Delta t) = \mathbf{r_i}(t) + \dot{\mathbf{r}}_\mathbf{i}(t)\Delta t + \frac{1}{2}\frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{A.47}$$

where (A.3) was used to write the acceleration $\ddot{\mathbf{r}}_\mathbf{i}(t)$ in terms of the force $\mathbf{F_i}(t)$ acting on particle i. A velocity-independent scheme is obtained by writing a similar expansion for $\mathbf{r_i}(t-\Delta t)$,

$$\mathbf{r_i}(t-\Delta t) = \mathbf{r_i}(t) - \dot{\mathbf{r}}_\mathbf{i}(t)\Delta t + \frac{1}{2}\frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{A.48}$$

and adding (A.47) and (A.48) to obtain

$$\mathbf{r_i}(t+\Delta t) + \mathbf{r_i}(t-\Delta t) = 2\mathbf{r_i}(t) + \frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4). \tag{A.49}$$

The *Verlet algorithm* (Verlet, 1967) is obtained after rearranging (A.49):

$$\mathbf{r_i}(t+\Delta t) = 2\mathbf{r_i}(t) - \mathbf{r_i}(t-\Delta t) + \frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4). \tag{A.50}$$

Thus, the estimation of the new position $\mathbf{r_i}(t+\Delta t)$ contains an error of order Δt^4, where Δt is the time step in the molecular dynamics scheme. The velocities, if needed, can be found by subtracting (A.48) from (A.47),

$$\mathbf{r_i}(t+\Delta t) - \mathbf{r_i}(t-\Delta t) = 2\dot{\mathbf{r}}_\mathbf{i}(t)\Delta t + \mathcal{O}(\Delta t^3) \tag{A.51}$$

and rearranging to obtain

$$\dot{\mathbf{r}}_\mathbf{i}(t) = \frac{\mathbf{r_i}(t+\Delta t) - \mathbf{r_i}(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{A.52}$$

which is accurate to order Δt^2. Notice, however, that the position is known at time $t+\Delta t$ while the velocity is known at time t, that is one step backward.

A variant of the Verlet algorithm is the *velocity Verlet algorithm* (Swope et al., 1982) which explicitly evolves positions and velocities. Just as starting from $\mathbf{r_i}(t)$ and

evolving forward in time to obtain $\mathbf{r_i}(t+\Delta t)$, it also possible to start from $\mathbf{r_i}(t+\Delta t)$ and $\dot{\mathbf{r}}_\mathbf{i}(t+\Delta t)$, compute $\mathbf{F_i}(t+\Delta t)$, and evolve backwards in time to obtain $\mathbf{r_i}(t)$ according to

$$\mathbf{r_i}(t) = \mathbf{r_i}(t+\Delta t) - \dot{\mathbf{r}}_\mathbf{i}(t+\Delta t)\Delta t + \frac{1}{2}\frac{\mathbf{F_i}(t+\Delta t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^3). \tag{A.53}$$

Substituting $\mathbf{r_i}(t)$ from (A.47) into (A.53) and solving for $\dot{\mathbf{r}}_\mathbf{i}(t+\Delta t)$ leads to

$$\dot{\mathbf{r}}_\mathbf{i}(t+\Delta t) = \dot{\mathbf{r}}_\mathbf{i}(t) + \frac{\Delta t}{2m_i}\left[\mathbf{F_i}(t) + \mathbf{F_i}(t+\Delta t)\right] + \mathcal{O}(\Delta t^2). \tag{A.54}$$

The velocity Verlet algorithm uses both (A.47) and (A.54) to evolve positions and velocities simultaneously and is accurate to order Δt^2.

The second-order Verlet and velocity Verlet algorithms satisfy two very important properties that are crucial for the long-time stability of numerical solvers: (1) They are time-reversible, which means that if one starts from $\{\mathbf{r_i}(t+\Delta t), \dot{\mathbf{r}}_\mathbf{i}(t+\Delta t)\}$ $(i = 1, 2, \ldots, N)$ and uses a time step $-\Delta t$, the state $\{\mathbf{r_i}(t), \dot{\mathbf{r}}_\mathbf{i}(t)\}$ will be obtained, and (2) they are symplectic algorithms, meaning that although in general numerical solvers do not exactly conserve the Hamiltonian $\mathcal{H}(\mathbf{\Gamma})$ due to errors resulting from (a) numerical integration schemes and (b) finite precision arithmetic inherent to computers, there exists a Hamiltonian $\tilde{\mathcal{H}}(\mathbf{\Gamma}, \Delta t)$ such that, along a trajectory in phase space, $\tilde{\mathcal{H}}(\mathbf{\Gamma}, \Delta t)$ remains close to the true Hamiltonian (i.e. $\tilde{\mathcal{H}}(\mathbf{\Gamma}, \Delta t)$ approaches the true Hamiltonian $\mathcal{H}(\mathbf{\Gamma})$ as $\Delta t \to 0$) and is exactly conserved by the algorithms. In essence, it can be said that the velocity Verlet algorithm is numerically stable, exactly reversible in time, conserves linear momentum if the forces are all conservative, and has excellent energy-conserving properties even with long time steps (Allen and Tildesley, 2017). Figure A.5 summarizes the algorithm.

There are other classes of integrators that claim to be more accurate than the Verlet and velocity Verlet algorithms such as the Beeman algorithm, which provides a better estimate of the velocity, or the velocity-corrected Verlet algorithm, or the predictor-corrector methods, but those are neither symplectic nor time-reversible and, therefore, may lead to undesirable drifts in the total energy. These methods are, however, reviewed

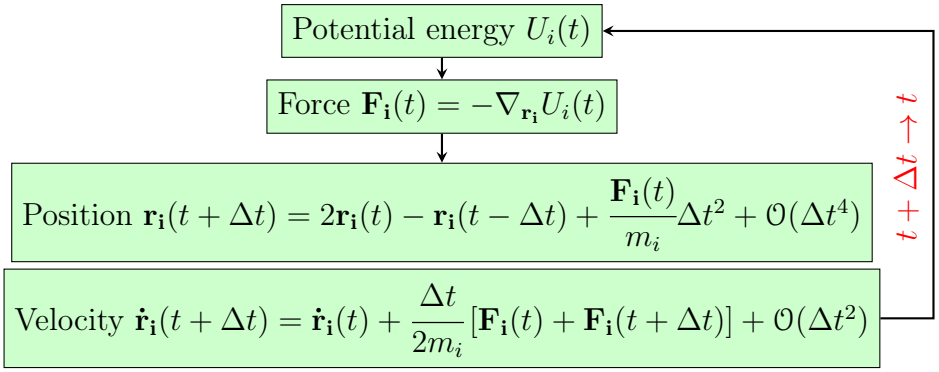


Figure A.5: Velocity Verlet algorithm for the numerical integration of the equations of motion.

in Berendsen and van Gunsteren (1986).

A.5.2 Molecular dynamics for different ensembles

The Verlet and velocity Verlet algorithms conserve the energy of the system, therefore they naturally sample configurations in the microcanonical ensemble. Nonetheless, as it was discussed in Section A.4, different experimental conditions lead to different statistical ensembles. The temperature is often times kept fixed instead of the internal energy, or the pressure instead of the volume.

Constant temperature: Canonical ensemble

In statistical physics, in a system of N particles at temperature T, every form of energy whose mathematical expression is a quadratic function of a coordinate or velocity component contributes with an average energy of $k_B T/2$. This is called the *equipartition theorem* (Schroeder, 2000). In the atomic case, the momenta $\{\mathbf{p_i}\}$ appear as squared terms in the Hamiltonian as shown in (A.42), thus

$$\langle \mathcal{K} \rangle = \left\langle \sum_{i=1}^{N} \sum_{\alpha=1}^{3} \frac{p_{i\alpha}^2}{2m_i} \right\rangle = \frac{3N}{2} k_B T \tag{A.55}$$

and the temperature of the system is proportional to its average kinetic energy. Expression (A.55) allows to define an instantaneous kinetic temperature function

$$\mathcal{T} = \frac{2\mathcal{K}}{3Nk_B} = \frac{1}{3Nk_B}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\frac{p_{i\alpha}^2}{m_i} \tag{A.56}$$

whose average is equal to T. If the system contains rigid molecules that can be described in terms of center of mass positions and velocities along with orientational variables, the angular velocities also contribute to the kinetic energy and appear in the definition of the instantaneous temperature $\mathcal{T}$. For a system of N particles subject to a total number of N_c independent internal molecular constraints (such as fixed bonds and angles), the number of degrees of freedom is reduced to $3N - N_c$ and the instantaneous kinetic temperature is

$$\mathcal{T} = \frac{2\mathcal{K}}{(3N - N_c)k_B} = \frac{1}{(3N - N_c)k_B}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\frac{p_{i\alpha}^2}{m_i}. \tag{A.57}$$

Expressions (A.56) and (A.57) provide a connection between the trajectories in phase space generated by the integration of the equations of motion and the macroscopic observable temperature.

In real macroscopic systems the temperature is controlled by allowing heat exchange across the system boundaries with a thermal bath. In the microscopic picture, heat exchange takes place through collisions of particles in the system with particles of the walls that separate the system from the thermal bath, and thermal relaxation occurs next through collisions between particles in the system. In computer simulations, any particle in the system can be coupled to the thermal bath.

There are several techniques to control the temperature of a simulation system and a few are mentioned here without extensive details, which can be found elsewhere (Frenkel and Smith, 2002; Thijssen, 2007; Tuckerman, 2010; Allen and Tildesley, 2017). Andersen (1980) designed a method in which the system is coupled to a heath bath that is represented by stochastic impulse forces that act every so often on randomly selected particles. The implementation consists in replacing the velocity of the selected particle by a velocity drawn from a Maxwell distribution with the desired temperature. The system

evolves at constant energy between stochastic collisions, which ensure that all accessible constant energy shells in phase space are visited according to their Boltzmann weighting factor (Frenkel and Smith, 2002). In this method, velocity alterations resemble particle collisions with the walls.

Other method is the velocity rescaling every time step rather than occasionally. The idea behind the rescaling is that, for a sufficiently large system, the total kinetic energy in the canonical ensemble evolves towards its equilibrium value $3(N - N_c)k_BT/2$ and the fluctuations around this value are small, thus instead of having a narrow distribution of the kinetic energy, it is replaced by a delta function. The rescaling is given by

$$p_i \rightarrow p_i \sqrt{\frac{\frac{3}{2}(N - N_c)k_BT}{\mathcal{K}}} \tag{A.58}$$

where p_i is the magnitude of the momentum of particle i and T corresponds to the desired temperature.

Yet other methods introduce the coupling with the thermal bath through a frictional force proportional to the velocity of the particles so that the equations of motion look like

$$m_i \ddot{\mathbf{r}}_\mathbf{i} = \mathbf{F}_\mathbf{i}(\{\mathbf{r}_\mathbf{i}\}) - \zeta(\mathbf{\Gamma})\dot{\mathbf{r}}_\mathbf{i} \tag{A.59}$$

where $\zeta(\mathbf{\Gamma})$ is a friction parameter that is the same for all particles. There exist several forms for such parameter (Hoover et al., 1980; Haile and Gupta, 1983; Berendsen et al., 1984; Hoover, 1985).

The method devised by Andersen (1980) is stochastic. Nosé (1984a,b) extended the Lagrangian $\mathcal{L}$ of the system,

$$\mathcal{L}(\mathbf{\Gamma}) = \mathcal{K}(\{\mathbf{p}_\mathbf{i}\}) - \mathcal{V}(\{\mathbf{r}_\mathbf{i}\}), \tag{A.60}$$

to perform deterministic molecular dynamics at constant temperature. It is more common now to use the Nosé scheme in the formulation of Hoover (1985, 1986), which is referred to as the Nosé-Hoover thermostat and explicitly introduces the heat bath into the system

in the form of a single degree of freedom. The details are discussed elsewhere (Frenkel and Smith, 2002; Thijssen, 2007; Tuckerman, 2010; Allen and Tildesley, 2017).

Constant pressure: Isothermal-isobaric ensemble

For a system of N particles, the relation between pressure and molecular quantities is given by the virial theorem

$$-\frac{1}{3}\left\langle \sum_{i=1}^{N} \mathbf{r_i} \cdot \nabla_{\mathbf{r_i}} \mathcal{V} \right\rangle = \frac{1}{3}\left\langle \sum_{i=1}^{N} \mathbf{r_i} \cdot \mathbf{F_i}^{\mathrm{tot}} \right\rangle = -Nk_BT \tag{A.61}$$

from classical dynamics (Allen and Tildesley, 2017). $\mathbf{F_i}^{\mathrm{tot}}$ represents the sum of intermolecular forces and external forces acting on particle i. The derivation of the virial theorem, however, appears to fail for a system with periodic boundaries, which is normally the case in molecular dynamics simulations (Erpenbeck and Wood, 1977). Nonetheless, the interpretation of pressure as a momentum flux leads to an alternative derivation of the virial theorem that is suitable for periodic systems. The details are presented in Haile (1992) and the expression is

$$PV = \frac{2}{3}\langle \mathcal{K} \rangle + \frac{1}{3}\left\langle \sum_{i=1}^{N-1} \sum_{j=i}^{N} \mathbf{r_{ij}} \cdot \mathbf{F_{ij}} \right\rangle \tag{A.62}$$

where $\mathbf{F_{ij}}$ is the intermolecular force between particles i and j and $\mathbf{r_{ij}} \equiv \mathbf{r_i} - \mathbf{r_j}$ is their relative position. Depending on the ensemble in which the average kinetic energy $\langle \mathcal{K} \rangle$ is computed, the fluctuations are different in general but give the same pressure P when averaged.

Expression (A.62) provides a connection between quantities computed along the trajectories in phase space generated by the integration of the equations of motion and the macroscopic observable pressure.

Experimental conditions may require to keep constant the pressure instead of the volume, thus the boundaries of the system are allowed to move. The technique used to simulate this class of experiments is similar to the design of the Nosé-Hoover thermostat in which the Lagrangian of the system is extended to include new degrees of freedom. In

this case, the volume is considered as a dynamical variable; the other new variable is called the canonical momentum and can be thought of as the momentum of a piston closing the system (Thijssen, 2007). This approach was first proposed by Andersen (1980) and later modified by Parrinello and Rahman (1981) to allow for anisotropic volume changes, and Martyna et al. (1994) to generate the correct distribution in the isothermal-isobaric ensemble (Frenkel and Smith, 2002; Allen and Tildesley, 2017).

A.6 Final remarks

Statistical mechanics concerns systems with a very large number of particles. Under some assumptions, the classical equations of motion can be solved to describe the evolution of the system. No analytic solution is possible for systems with more than three particles, thus a numerical algorithm is in place. Molecular dynamics is essentially such numerical algorithm for the integration of the equations of motion and generates trajectories in the phase space. The information produced by molecular dynamics corresponds to positions and velocities of every particle in the system. On the other hand, macroscopically thermodynamic variables such as temperature and pressure are more interesting than the knowledge of such detailed microscopic information. The premise is that the observed macroscopic properties are the average effect of the interactions and dynamics of individual components. Statistical mechanics provides the means to perform such averages.

Experimental arrangements impose some constraints in the variable that can vary and those that are kept fixed in a system. The set of all possible microscopic states that are compatible with the constraints is called statistical ensemble. Molecular dynamics and Monte Carlo simulations sample states in an ensemble. However, due to the still limited computational capabilities, this sampling is rather small. Accurate and efficient techniques are needed to speed up the evaluation of both long-range and short-range interactions, so that a sufficiently large sampling is achieved in a reasonable amount of time to produce meaningful ensemble averages. Fortunately, this is an ongoing effort that

has made good progress over the last few decades.

More importantly, however, is the existence of highly accurate force fields that describe the particles in a particular system of interest. The quality of the calculations from molecular simulations will be, at most, as good as the force field utilized. The study of a determined phenomenon is restricted to the ability to feed the simulation with the parameters that describe the interactions between the fundamental constituents of the system.

Appendix B

FORCE FIELD PARAMETERS

This appendix summarizes the parameters utilized in LAMMPS throughout this work. The original set of force field parameters were converted into LAMMPS *real* units, where necessary. The units[1] are listed below.

Table B.1: Real units in LAMMPS.

Quantity	Units
mass	gram/mole (g/mole)
distance	Angstrom (Å)
time	femtosecond (fs)
energy	kcal/mole (kcal/mole)
velocity	Angstrom/femtosecond (Å/fs)
force	kcal/mole×Angstrom (kcal/mole·Å)
torque	kcal/mole (kcal/mole)
temperature	Kelvin (K)
pressure	atmosphere (atm)
dynamic viscosity	Poise (P)
charge	multiple of electron charge (e)
dipole	charge×Angstrom (e·Å)
electric field	volt/Angstrom (V/Å)
density	gram/cm$^{\text{dimension}}$ (g/cm^d)

The cvff force field parameters are listed in Dauber-Osguthorpe et al. (1988); the original OPLS-AA force field parameters are given in Jorgensen et al. (1996) and the refinement to the torsional parameters is described in Price et al. (2001); the CLAYFF force field parameters for carbon dioxide are published in Cygan et al. (2012); the SPC/E water model is described in (Berendsen et al., 1987).

[1] https://lammps.sandia.gov/doc/units.html

B.1 Non-bonded interactions

The non-bonded interactions in the force fields considered in this work are described by the standard Lennard-Jones 12-6 potential (A.45)

$$E_{\mathrm{LJ},ij} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \quad r < r_c, \tag{A.45}$$

where r_c is the cutoff radius, and the Coulombic pairwise interaction given by (A.46)

$$E_{\mathrm{C},ij} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}, \tag{A.46}$$

where q_i and q_j are the charges on the two atoms and ε_0 is the dielectric constant. In the cvff force field, however, the Lennard-Jones potential is cast as

$$E_{\mathrm{LJ},ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}, \quad r < r_c, \tag{B.1}$$

thus the parameters ϵ_{ij} and σ_{ij} were computed using A_{ij} and B_{ij}.

The non-bonded interactions are given by the sum of these two terms,

$$E_{\mathrm{nb},ij} = E_{\mathrm{LJ},ij} + E_{\mathrm{C},ij}. \tag{B.2}$$

Table B.2: Non-bonded parameters.

Atom type	ϵ (kcal/mole)	σ (Å)
	cvff	
cp	0.148	3.617
c5	0.148	3.617
nh	0.167	3.501
hn	0.000	0.000
h	0.038	2.450
c2	0.039	3.875
c1	0.039	3.875
np	0.167	3.501
o	0.228	2.860
c'	0.148	3.617
s	0.043	3.368
o'	0.228	2.860

(To be continued)

Atom type	ϵ (kcal/mole)	σ (Å)
c3	0.039	3.875
c=	0.148	3.617
cs	0.148	3.617
sp	0.164	3.368
ho	0.000	0.000
OPLS-AA		
CT	0.066	3.500
HC	0.030	2.500
CA	0.070	3.550
HA	0.030	2.420
OS	0.140	2.900
CLAYFF		
C	0.056	2.800
O	0.160	3.028
SPC		
O	0.155	3.166
H	0.000	0.000

B.1.1 Combining rules

Interactions between unlike atoms in different carbon dioxide molecules, which are described by the CLAYFF force field, are computed using the Lorentz-Berthelot combining rules (Allen and Tildesley, 2017)

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \tag{B.3a}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j). \tag{B.3b}$$

Interactions between unlike atoms in different molecules described by the force fields cvff and OPLS-AA are computed using the geometric mixing rules

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \tag{B.4a}$$

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j}. \tag{B.4b}$$

Interactions between atoms in molecules that are described by different force fields are computed using the geometric mixing rules.

B.2 Bonded interactions

The bonded interactions include the potential energy terms bond, angle, dihedral, and improper,

$$E_{\mathrm{b}} = E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{dihedral}} + E_{\mathrm{improper}}. \tag{B.5}$$

B.2.1 Bond energy

In the cvff force field, the bond energy is given by the Morse potential

$$E_{\mathrm{bond}} = D\left[1 - \mathrm{e}^{-\alpha(r-r_0)}\right]^2 \tag{B.6}$$

where r_0 is the equilibrium bond distance, α is a stiffness parameter, and D determines the depth of the potential well.

In the force fields OPLS-AA and CLAYFF, the bond energy is given by the harmonic potential

$$E_{\mathrm{bond}} = K(r - r_0)^2 \tag{B.7}$$

where r_0 is the equilibrium bond distance and K is a prefactor. Usually there is a factor $1/2$ in the harmonic potential that is included in K in LAMMPS.

The SPE water model considers the molecule as a rigid body, thus there is neither bond nor angle energies associated with it.

Table B.3: Bond parameters for the cvff force field.

Bond type	D (kcal/mole)	α (Å^{-1})	r_0 (Å)
cp-cp	120.000	2.000	1.340
cp-h	116.000	1.770	1.080
cp-c5	70.000	2.000	1.340
cp-c1	76.000	1.930	1.510
c5-c5	70.000	2.000	1.390
cp-c2	76.000	1.930	1.510
c5-nh	80.000	2.000	1.380
c5-h	116.000	1.770	1.080
nh-hn	93.000	2.280	1.026
c2-c2	88.000	1.915	1.526
h-c2	108.600	1.771	1.105

(To be continued)

Bond type	D (kcal/mole)	α (Å^{-1})	r_0 (Å)
c2-c1	88.000	1.915	1.526
h-c1	108.600	1.771	1.105
c1-o	68.300	2.000	1.425
cp-c3	76.000	1.930	1.510
cp-np	140.000	2.000	1.260
cp-o	96.000	2.000	1.370
c2-o	68.300	2.000	1.425
c2-c'	76.000	1.930	1.520
c'-o'	145.000	2.060	1.230
c2-s	57.000	2.000	1.800
c1-c'	76.000	1.930	1.520
c5-c2	76.000	1.930	1.510
c1-c1	88.000	1.915	1.526
c2-c3	88.000	1.915	1.526
h-c3	108.600	1.771	1.105
cp-c=	80.700	2.000	1.500
c=-c=	163.800	2.000	1.330
c2-c=	80.700	2.000	1.500
h-c=	90.400	2.000	1.090
c5-cs	100.000	2.000	1.363
cs-sp	80.000	2.000	1.711
cp-cs	120.000	2.000	1.390
c5-o	105.000	2.000	1.370

Table B.4: Bond parameters for the OPLS-AA and CLAYFF force fields.

Bond type	K (kcal/mole·Å^2)	r_0 (Å)
OPLS-AA		
CT-CT	268.000	1.529
HC-CT	340.000	1.090
CA-CT	317.000	1.510
CA-HA	367.000	1.080
CA-CA	469.000	1.400
CT-OS	320.000	1.410
CLAYFF		
C-O	1008.963	1.162

B.2.2 Angle energy

The angle energy in the force fields utilized in this work are given by the harmonic potential

$$E_{\text{angle}} = K(\theta - \theta_0)^2 \tag{B.8}$$

where θ_0 is the equilibrium value of the angle and K is a prefactor. Usually there is a factor 1/2 in the harmonic potential that is included in K in LAMMPS. θ_0 is specified in degrees, but LAMMPS converts it to radians internally, thus the units of K are kcal/mole·rad^2.

Table B.5: Angle parameters.

Angle type	K (kcal/mole·rad^2)	θ_0 (°)
	cvff	
cp-cp-cp	90.000	120.000
cp-cp-h	37.000	120.000
cp-cp-c5	90.000	120.000
cp-cp-c1	44.200	120.000
cp-c5-c5	90.000	120.000
c5-c5-c5	90.000	120.000
cp-cp-c2	44.200	120.000
c5-c5-nh	90.000	120.000
c5-c5-h	37.000	120.000
nh-c5-h	40.000	120.000
c5-nh-c5	75.100	114.000
c5-nh-hn	27.500	120.000
cp-c2-c2	46.600	110.500
cp-c2-h	44.400	110.000
h-c2-c2	44.400	110.000
h-c2-h	39.500	106.400
c2-c2-c1	46.600	110.500
h-c2-c1	44.400	110.000
cp-c1-c2	46.600	110.500
h-c1-c2	44.400	110.000
c2-c1-o	70.000	109.500
cp-c1-h	44.400	110.000
cp-c1-o	46.000	110.500
h-c1-o	57.000	109.500
c2-c2-c2	46.600	110.500
cp-cp-c3	44.200	120.000
cp-cp-np	90.000	120.000
h-cp-np	50.000	120.000
cp-cp-o	60.000	120.000
cp-np-cp	75.100	114.000
c2-c2-o	70.000	109.500
h-c2-o	57.000	109.500
c2-o-c1	60.000	109.500
c2-c2-c'	46.600	110.500
h-c2-c'	45.000	109.500
c2-c'-o	122.800	110.000
c2-c'-o'	68.000	120.000
o-c'-o'	145.000	123.000
c2-c1-c2	46.600	110.500
cp-c2-c1	46.600	110.500
c1-c2-s	62.000	109.000
h-c2-s	45.000	109.000
c2-c2-s	62.000	109.000
c2-s-c2	58.000	99.000
c2-c1-c'	46.600	110.500
h-c1-c'	45.000	109.500
c2-c'-c1	40.000	120.000
c1-c'-o'	68.000	120.000
c1-c2-c'	46.600	110.500
c5-cp-h	37.000	120.000
c5-c5-c2	44.200	120.000

(To be continued)

Angle type	K (kcal/mole·rad^2)	θ_0 (°)
nh-c5-c2	50.000	120.000
c5-c2-c2	46.600	110.500
c5-c2-h	44.400	110.000
c1-c'-c1	40.000	120.000
c1-c1-c'	46.600	110.500
c2-c1-c1	46.600	110.500
h-c1-c1	44.400	110.000
c1-c2-c1	46.600	110.500
cp-c5-c2	80.000	120.000
c5-c2-c1	46.600	110.500
c1-c2-o	70.000	109.500
c2-c2-c3	46.600	110.500
h-c2-c3	44.400	110.000
h-c3-c2	44.400	110.000
h-c3-h	39.500	106.400
c2-o-c2	60.000	109.500
cp-cp-c=	80.000	120.000
cp-c=-c=	36.200	120.000
cp-c=-c2	36.200	120.000
c2-c=-c=	36.200	122.300
h-c=-c=	33.800	121.200
cp-c=-h	37.500	120.000
h-c2-c=	44.400	110.000
c1-c2-c=	46.000	109.500
cp-c2-c3	46.600	110.500
c5-cp-c1	80.000	120.000
c5-c5-cs	70.000	112.700
cp-c5-cs	80.000	120.000
c5-cs-sp	88.760	111.039
cp-cs-c5	80.000	120.000
cp-cs-sp	89.000	114.000
cs-sp-cs	126.506	92.567
c5-cp-c2	80.000	120.000
cp-c5-nh	90.000	120.000
c1-c2-c3	46.600	110.500
cp-cp-cs	80.000	120.000
c2-cp-cs	80.000	120.000
h-cp-cs	37.000	120.000
c5-c5-o	60.000	120.000
cp-c5-o	60.000	120.000
cp-o-c2	50.000	109.500
c5-o-c1	60.000	109.500
cp-c2-o	46.000	110.500
cp-c3-h	44.400	110.000
cp-c1-c1	46.600	110.500
cp-o-cp	60.000	109.500
c'-o-ho	50.000	112.000
	OPLS-AA*	
HC-CT-HC	33.000	107.800
HC-CT-CT	37.500	110.700
CT-CT-CT	58.350	112.700
HC-CT-CA	35.000	109.500
CT-CA-CA	70.000	120.000
HA-CA-CA	35.000	120.000

(To be continued)

Angle type	K (kcal/mole·rad^2)	θ_0 (°)
CA-CA-CA	63.000	120.000
1	85.000	117.300
2	85.000	134.900
3	70.000	120.000
4	63.000	114.000
5	58.350	112.700
6	70.000	128.600
7	50.000	109.500
8	60.000	109.500
9	35.000	120.000
10	37.500	110.700
11	35.000	109.500
12	33.000	107.800
	CLAYFF	
O-C-O	54.003	180.000

*The numeral angle types were assigned automatically using LigParGen (Dodda et al., 2017).

B.2.3 Dihedral energy

In the cvff force field, the dihedral energy is given by the harmonic potential

$$E_{\text{dihedral}} = K\,[1 + d\cos(n\phi)] \tag{B.9}$$

where K is a prefactor, d takes on the values ± 1, and n is an integer.

In the force field OPLS-AA, the dihedral energy is given by the potential

$$E_{\text{dihedral}} = \frac{1}{2}K_1\,[1 + \cos(\phi)] + \frac{1}{2}K_2\,[1 - \cos(2\phi)] + \frac{1}{2}K_3\,[1 + \cos(3\phi)] + \frac{1}{2}K_4\,[1 - \cos(4\phi)] \tag{B.10}$$

where $K_1,\ K_2,\ K_3,\ K_4$ are constants that do not contain the usual 1/2 factor.

Table B.6: Dihedral parameters for the cvff force field.

Dihedral type	K (kcal/mole)	d	n
cp-cp-cp-cp	3.000	−1	2
cp-cp-cp-c5	3.000	−1	2
cp-cp-cp-h	3.000	−1	2
c5-cp-cp-h	3.000	−1	2
h-cp-cp-h	3.000	−1	2

(To be continued)

Dihedral type	K (kcal/mole)	d	n
cp-cp-c5-c5	3.000	−1	2
cp-cp-cp-c1	3.000	−1	2
cp-cp-cp-c2	3.000	−1	2
h-cp-cp-c1	3.000	−1	2
cp-cp-c1-c2	0.000	0	0
cp-cp-c1-h	0.000	0	0
cp-cp-c1-o	0.000	0	0
cp-c5-c5-cp	3.000	−1	2
cp-c5-c5-c5	3.000	−1	2
c5-c5-c5-c5	3.000	−1	2
cp-c5-c5-nh	3.000	−1	2
cp-c5-c5-h	3.000	−1	2
c5-c5-c5-nh	3.000	−1	2
c5-c5-c5-h	3.000	−1	2
h-cp-cp-c2	3.000	−1	2
cp-cp-c2-c2	0.000	0	0
cp-cp-c2-h	0.000	0	0
c5-c5-nh-c5	1.000	−1	2
c5-c5-nh-hn	1.000	−1	2
h-c5-nh-c5	1.000	−1	2
h-c5-nh-hn	1.000	−1	2
cp-c2-c2-c1	0.158	1	3
cp-c2-c2-h	0.158	1	3
h-c2-c2-c1	0.158	1	3
h-c2-c2-h	0.158	1	3
c2-c2-c1-cp	0.158	1	3
c2-c2-c1-h	0.158	1	3
c2-c2-c1-o	0.158	1	3
h-c2-c1-cp	0.158	1	3
h-c2-c1-h	0.158	1	3
h-c2-c1-o	0.158	1	3
c2-c1-o-c2	0.130	1	3
cp-c1-o-c2	0.130	1	3
h-c1-o-c2	0.130	1	3
cp-c2-c2-c2	0.158	1	3
h-c2-c2-c2	0.158	1	3
c2-c2-c2-c1	0.158	1	3
c2-c2-c1-c2	0.158	1	3
c2-c2-c1-c’	0.158	1	3
h-c2-c1-c2	0.158	1	3
h-c2-c1-c’	0.158	1	3
cp-cp-cp-c3	3.000	−1	2
cp-cp-cp-o	3.000	−1	2
c2-cp-cp-c3	3.000	−1	2
cp-cp-c2-c1	0.000	0	0
cp-cp-c3-h	0.000	0	0
cp-cp-cp-np	3.000	−1	2
cp-cp-np-cp	2.000	−1	2
h-cp-np-cp	2.000	−1	2
c2-cp-cp-np	3.000	−1	2
c2-cp-cp-o	3.000	−1	2
cp-cp-o-cp	0.750	−1	2
c2-cp-cp-c2	3.000	−1	2
cp-c2-c2-o	0.158	1	3
h-c2-c2-o	0.158	1	3

(To be continued)

Dihedral type	K (kcal/mole)	d	n
c2-c2-o-c1	0.130	1	3
h-c2-o-c1	0.130	1	3
h-cp-cp-o	3.000	−1	2
c2-c2-c2-c’	0.158	1	3
h-c2-c2-c’	0.158	1	3
c2-c2-c’-o’	0.000	0	0
h-c2-c’-o’	0.000	0	0
c2-c2-c2-o	0.158	1	3
cp-c2-c1-o	0.158	1	3
cp-c2-c1-h	0.158	1	3
cp-c2-c1-c2	0.158	1	3
c1-c2-c2-s	0.158	1	3
h-c2-c2-s	0.158	1	3
c1-c2-s-c2	0.140	1	3
h-c2-s-c2	0.140	1	3
s-c2-c1-c2	0.158	1	3
s-c2-c1-c’	0.158	1	3
s-c2-c1-h	0.158	1	3
c2-c2-s-c2	0.140	1	3
c2-c1-c’-c2	0.000	0	0
c2-c1-c’-o’	0.000	0	0
h-c1-c’-c2	0.000	0	0
h-c1-c’-o’	0.000	0	0
h-c2-c’-c1	0.000	0	0
c1-c2-c’-c1	0.000	0	0
c1-c2-c’-o’	0.000	0	0
c’-c2-c1-c2	0.158	1	3
c’-c2-c1-h	0.158	1	3
c5-cp-cp-c3	3.000	−1	2
h-cp-c5-c5	3.000	−1	2
c5-c5-c5-c2	3.000	−1	2
cp-c5-c5-c2	3.000	−1	2
cp-cp-c2-o	0.000	0	0
h-cp-cp-c3	3.000	−1	2
c2-c5-nh-c5	1.000	−1	2
c2-c5-nh-hn	1.000	−1	2
c5-c5-c2-c2	0.000	0	0
c5-c5-c2-h	0.000	0	0
nh-c5-c2-c2	0.000	0	0
nh-c5-c2-h	0.000	0	0
c5-c2-c2-c1	0.158	1	3
c5-c2-c2-h	0.158	1	3
c2-c1-c’-c1	0.000	0	0
h-c1-c’-c1	0.000	0	0
c1-c1-c’-c1	0.000	0	0
c1-c1-c’-o’	0.000	0	0
c1-c2-c1-c’	0.158	1	3
c1-c2-c1-h	0.158	1	3
c1-c2-c1-c1	0.158	1	3
h-c2-c1-c1	0.158	1	3
cp-c1-c1-c’	0.158	1	3
h-c1-c1-c’	0.158	1	3
c2-c1-c1-c’	0.158	1	3
cp-c1-c1-c2	0.158	1	3
h-c1-c1-c2	0.158	1	3

(To be continued)

Dihedral type	K (kcal/mole)	d	n
c2-c1-c1-c2	0.158	1	3
cp-c1-c1-h	0.158	1	3
h-c1-c1-h	0.158	1	3
c1-c2-c1-cp	0.158	1	3
c1-c2-c1-c2	0.158	1	3
cp-c2-c1-cp	0.158	1	3
c1-cp-cp-c1	3.000	−1	2
c5-cp-cp-np	3.000	−1	2
cp-cp-c1-c1	0.000	0	0
cp-cp-c5-c2	3.000	−1	2
h-cp-cp-np	3.000	−1	2
c2-c5-c5-c2	3.000	−1	2
cp-c5-c2-c1	0.000	0	0
cp-c5-c2-h	0.000	0	0
c5-c5-c2-c1	0.000	0	0
h-cp-c5-c2	3.000	−1	2
c5-c2-c1-c2	0.158	1	3
c5-c2-c1-h	0.158	1	3
o-c2-c1-c2	0.158	1	3
o-c2-c1-h	0.158	1	3
c1-c2-o-c2	0.130	1	3
h-c2-o-c2	0.130	1	3
cp-c2-c2-c3	0.158	1	3
h-c2-c2-c3	0.158	1	3
c2-c2-c3-h	0.158	1	3
h-c2-c3-h	0.158	1	3
c2-c2-o-c2	0.130	1	3
c1-c2-c2-o	0.158	1	3
cp-cp-cp-c=	3.000	−1	2
c2-cp-cp-c=	3.000	−1	2
cp-cp-c2-c3	0.000	0	0
cp-cp-c=-c=	0.125	−1	2
cp-cp-c=-h	0.000	0	0
cp-cp-c=-c2	0.125	−1	2
cp-c=-c=-cp	4.075	−1	2
cp-c=-c=-h	4.075	−1	2
cp-c=-c=-c2	4.075	−1	2
h-c=-c=-c2	4.075	−1	2
h-c2-c=-cp	0.211	1	3
c1-c2-c=-cp	0.211	1	3
h-c2-c=-c=	0.211	1	3
c1-c2-c=-c=	0.211	1	3
c=-c2-c1-o	0.158	1	3
c=-c2-c1-h	0.158	1	3
c=-c2-c1-c2	0.158	1	3
cp-c2-c3-h	0.158	1	3
c5-cp-c2-c2	0.000	0	0
c5-cp-c2-h	0.000	0	0
cp-cp-c5-cs	3.000	−1	2
c1-cp-c5-c5	3.000	−1	2
c1-cp-c5-cs	3.000	−1	2
c5-cp-c1-c2	0.000	0	0
c5-cp-c1-h	0.000	0	0
c5-c5-cs-sp	6.452	−1	2
c5-c5-cs-cp	1.500	−1	2

(To be continued)

Dihedral type	K (kcal/mole)	d	n
cp-c5-cs-sp	6.452	−1	2
cp-c5-cs-cp	1.500	−1	2
cp-c5-c5-cs	3.000	−1	2
cs-c5-c5-cs	3.000	−1	2
c2-cp-c5-cs	3.000	−1	2
c2-cp-c5-c5	3.000	−1	2
c5-cs-sp-cs	12.000	−1	2
cp-cs-sp-cs	12.000	−1	2
cp-cp-cs-c5	1.500	−1	2
c2-cp-cs-c5	1.500	−1	2
cp-cp-cs-sp	6.452	−1	2
c2-cp-cs-sp	1.500	−1	2
h-cp-cs-c5	1.500	−1	2
h-cp-cs-sp	2.000	−1	2
c5-cp-cp-c2	3.000	−1	2
c5-cp-c2-c1	0.000	0	0
c1-c2-c1-o	0.158	1	3
c2-cp-c5-c2	3.000	−1	2
cp-c5-nh-c5	1.000	−1	2
cp-c5-nh-hn	1.000	−1	2
cp-cp-c5-nh	3.000	−1	2
h-cp-c5-nh	3.000	−1	2
nh-c5-c2-c1	0.000	0	0
c3-c2-c1-c2	0.158	1	3
c3-c2-c1-h	0.158	1	3
c1-c2-c3-h	0.158	1	3
cp-cp-cp-cs	3.000	−1	2
c2-cp-cp-cs	3.000	−1	2
cs-cp-c2-c1	0.000	0	0
cs-cp-c2-h	0.000	0	0
h-cp-cp-cs	3.000	−1	2
cp-cp-c5-o	3.000	−1	2
cp-c5-c5-o	3.000	−1	2
c2-c5-c5-o	3.000	−1	2
c5-c5-o-c1	1.800	−1	2
cp-c5-o-c1	1.800	−1	2
c2-c2-o-cp	0.130	1	3
h-c2-o-cp	0.130	1	3
cp-cp-o-c2	1.800	−1	2
cp-c2-c1-c1	0.158	1	3
c1-c2-c2-c1	0.158	1	3
c1-c2-o-c1	0.130	1	3
c2-c1-o-c5	0.130	1	3
h-c1-o-c5	0.130	1	3
cp-c2-o-c2	0.130	1	3
c5-c2-c1-o	0.158	1	3
c2-c2-c1-c1	0.158	1	3

Table B.7: Dihedral parameters for the OPLS-AA force field.

Dihedral type*	K_1 (kcal/mole)	K_2 (kcal/mole)	K_3 (kcal/mole)	K_4 (kcal/mole)
HC-CT-CT-HC	0.000	0.000	0.300	0.000
HC-CT-CT-CT	0.000	0.000	0.300	0.000
CT-CT-CT-CT	1.300	−0.050	0.200	0.000

(To be continued)

Dihedral type*	K_1 (kcal/mole)	K_2 (kcal/mole)	K_3 (kcal/mole)	K_4 (kcal/mole)
1	0.000	7.250	0.000	0.000
2	0.000	7.000	0.000	0.000
3	1.300	−0.200	0.200	0.000
4	0.650	−0.250	0.670	0.000
5	1.711	−0.500	0.663	0.000
6	0.000	0.000	0.468	0.000
7	0.000	0.000	0.760	0.000
8	0.000	0.000	0.462	0.000
9	0.000	0.000	0.000	0.000

*The numeral dihedral types were assigned automatically using LigParGen (Dodda et al., 2017).

B.2.4 Improper energy

In the cvff force field, the improper energy is given by the potential

$$E_{\text{improper}} = K\left[1 + d\cos(n\phi)\right] \tag{B.11}$$

where K is a prefactor, d takes on the values ± 1, and n is an integer less than or equal to 6.

Table B.8: Improper parameters for the cvff force field.

Improper type	K (kcal/mole)	d	n
cp-cp-cp-h	0.370	−1	2
cp-cp-cp-c5	0.370	−1	2
cp-cp-cp-cp	0.370	−1	2
cp-cp-cp-c1	0.370	−1	2
cp-c5-c5-c5	0.370	−1	2
cp-cp-cp-c2	0.370	−1	2
c5-c5-h-nh	0.370	−1	2
hn-nh-c5-c5	0.370	−1	2
cp-cp-cp-c3	0.370	−1	2
cp-cp-h-np	0.370	−1	2
cp-cp-cp-np	0.370	−1	2
cp-cp-cp-o	0.000	0	0
c2-c'-o-o'	11.600	−1	2
c2-c'-c1-o'	10.000	−1	2
cp-cp-c5-h	0.370	−1	2
c5-c5-nh-c2	0.370	−1	2
c1-c'-c1-o'	10.000	−1	2
cp-c5-c5-c2	0.370	−1	2
cp-cp-cp-c=	0.370	−1	2
cp-c=-c2-c=	11.100	−1	2
cp-c=-h-c=	11.100	−1	2
cp-cp-c5-c1	0.370	−1	2
cp-c5-c5-cs	0.370	−1	2

(To be continued)

Improper type	K (kcal/mole)	d	n
cp-cs-c5-sp	0.370	−1	2
cp-cp-c5-c2	0.370	−1	2
cp-c5-c5-nh	0.370	−1	2
cp-cp-c2-cs	0.370	−1	2
cp-cp-cs-h	0.370	−1	2
cp-c5-c5-o	0.000	0	0