INVESTIGATION OF FLUID PHASE BEHAVIOR IN SHALE RESERVOIRS USING EQUATION OF STATE, MOLECULAR SIMULATION AND MACHINE LEARNING

A Dissertation

by

FANGXUAN CHEN

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hadi Nasrabadi
Committee Members,	Debjyoti Banerjee
	Eduardo Gildin
	Kan Wu
Head of Department,	Akhil Datta-Gupta

May 2023

Major Subject: Petroleum Engineering

ABSTRACT

Fluid phase behavior in shale reservoirs differs significantly from phase behavior in conventional reservoirs due to the strong interactions between fluid and boundary in nanopores. In this study, we applied equation-of-state (EOS) modeling, machine learning (ML) technique and molecular simulation to investigate fluid phase behavior in shale reservoirs.

One common issue observed in liquid-rich shale (LRS) production is that oil recovery of LRS reservoirs is much lower compared to oil recovery from a conventional reservoir with the same drawdown. To understand this phenomenon, EOS modeling is developed to analyze the fluid compositions in the bulk and confined regions. Our simulation results indicate that hydrocarbons distribute heterogeneously with respect to pore size on a nanoscale. The leaner bulk composition leads to the reduction in oil recovery from LRS reservoirs. Although EOS modeling can accurately simulate fluid phase behavior in shale reservoirs, the required simulation time is much longer than that for models of conventional reservoirs. To solve this problem, ML techniques were applied to accelerate the phase-equilibrium calculations in the EOS modeling. In contrast to previous models designed for a specific type of hydrocarbon, we have developed a generalized, ML-assisted phase-equilibrium calculation model that is suitable for shale reservoirs. In total, the average CPU time required for phase-equilibrium calculation using the generalized ML-assisted phase-equilibrium model was reduced by more than two orders of magnitude while maintaining an accuracy of 97%. With the development of shale oil and gas, depleted shale gas reservoirs may be attractive candidates for hydrogen (H_2) storage. Molecular simulation was used to investigate the potential for H_2 storage in depleted shale gas reservoirs. The results of the simulation suggest that a higher proportion of H_2 exists in the bulk region. Because fluid is mainly produced from the bulk region,

the high percentage of H_2 in bulk fluid would lead to high purity of H_2 during the recovery process. This work contributes to the understanding and application of fluid phase behavior in shale reservoirs.

ACKNOWLEDGEMENTS

I would like to thank my committee chair Dr. Nasrabadi, my committee members, Dr. Banerjee, Dr. Gildin, and Dr. Wu for their guidance and support throughout the course of this research.

Thanks to Texas A&M High Performance Research Computing (HPRC) Facility for their support.

Thanks also go to my colleagues in Dr. Nasrabadi's research group for their assistance with my research.

Finally, I appreciate the encouragement, patience, and love of my family.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Hadi Nasrabadi (Douglas Von Gonten Faculty Fellow) of the Department of Petroleum Engineering, Professor Debjyoti Banerjee (James J. Cain ’51 Faculty Fellow | Dean’s Fellow for EnMed) of the Department of Mechanical Engineering, Professor Eduardo Gildin (Ted H. Smith ’75 and Max R. Vordenbaum ’73 DVG Development Professor) of the Department of Petroleum Engineering, and Professor Kan Wu (Chevron Corporation Faculty Fellow) of the Department of Petroleum Engineering.

All simulation work was supported by the Texas A&M High Performance Research Computing (HPRC) Facility.

The analysis depicted in Section 2 was conducted in part by Dr. Sheng Luo who was in Dr. Nasrabadi’s research group and was published in 2021.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

This work is not sponsored by any funding.

NOMENCLATURE

AFM	Atomic force microscopy
AMCSD	American Mineralogist Crystal Structure Database
ANN	Artificial neural network
BP	Backpropagation
CBMC	Configurational-bias Monte Carlo
CGR	Condensate gas ratio
CPU	Central processing unit
CVD	Constant volume depletion
DSC	Differential scanning calorimetry
EOS	Equation-of-state
FIB-SEM	Focused ion beam scanning electron microscopy
GCMC	Grand canonical Monte Carlo
GEMC	Gibbs ensemble Monte Carlo
HC	Hydrocarbon
IUPAC	International Union of Pure and Applied Chemistry
LJ	Lennard Jones
LPA	Low-pressure adsorption
LRS	Liquid-rich shale
LSSVM	Least-squares support vector machine
MAE	Mean absolute error
MC	Monte Carlo
MCCCS	Monte Carlo for Complex Chemical systems

MD	Molecular dynamics
MICP	Mercury injection capillary pressure
ML	Machine learning
MLP	Multi-layer perceptron
MSE	Mean square error
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
NPT-GEMC	Pressure-imposed Gibbs ensemble Monte Carlo
NVT-GEMC	Volume-imposed Gibbs ensemble Monte Carlo
OGIP	Original gas in place
OOIP	Original oil in place
PAH	Polycyclic aromatic hydrocarbons
PBC	Periodic boundary condition
PDE	Partial differential equation
PINN	Physics-informed neural network
PR EOS	Peng-Robinson equation of state
PR-C EOS	Extended Peng-Robinson equation of state
ReLU	Rectified linear unit
RVM	Relevance vector machine
SVM	Support vector machine
TEM	Transmission electron microscopy
TPD	Tangent plane distance
VLE	Vapor-liquid equilibrium

A	Helmholtz free energy
A_c	Cross-sectional area of cylindrical pores
a_p	Confinement-modified energy parameter
$B_{g,i}$	Initial gas formation volume factor
BIP	Binary interaction parameter
b_p	Confinement-modified volume parameter
d	Diameter
E_{conf}	Configurational energy
F_p	Percentage of molecules of component in surface-adsorbed region
k_B	Boltzmann constant
K_i	Equilibrium ratio of component *i*
K_r	Relative permeability
m_{ex}	Excess adsorption capacity
MMscf	Million standard cubic feet
Mscf	Thousand standard cubic feet
MW	Molecular weight
N	Number of molecules
n	Number of moles
N_A	Avogadro constant
N_c	Coordination number
nc	Number of components
ns	Number of samples
P	Pressure

P_c	Critical pressure
P_d	Dew point pressure
q	Intramolecular contribution
r_i	Initial solution oil-gas ratios
r_p	Pore radius
S	Saturation
S_A	Surface area
S_{ij}	Relative selectivity of component *i* and *j*
stb	Standard barrel
T	Temperature
T_a	Pore-size-dependent transmissibility
T_c	Critical temperature
U	Nonbonded interaction energy
V	Volume
V_f	Free volume
V_p	Pore volume
VSP	Volume shift parameter
W	Distance
x	Mole fraction
δ_p	Square-well width
ε_p	Square-well depth
κ	Normalized fluid-pore surface affinity factor
λ	Translational contribution

μ	Chemical potential
ρ_a	Adsorbed phase density
ρ_{max}	Packing density of pure component
σ	Effective diameter
ω	Acentric factor

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Problem Statement

In contrast to conventional reservoirs, which are dominated by macro-scale pores, shale reservoirs have a high proportion of nanopores [3]. According to the criteria suggested by IUPAC [4, 5], pores are categorized as three types: micropores (d < 2 nm), mesopores (2 nm < d < 50 nm) and macropores (d > 50 nm). The wide distribution of nanopores (including micropores and mesopores) leads to inaccurate predictions of hydrocarbon phase behavior using conventional models. Phase behavior has a significant effect on the composition and mobility of extracted oil and gas and their production rates [6]. Therefore, it is of great importance to develop new models to accurately describe fluid phase behavior in shale reservoirs.

Liquid-rich shale (LRS) reservoirs belong to ultra-tight unconventional formations with a significant number of hydrocarbons [7]. In recent years, the widespread application of horizontal wells and multi-stage hydraulic fracturing has greatly enhanced the recovery of LRS [8]. One common issue observed in LRS production is that oil recovery from LRS is much lower compared with oil recovery from a conventional reservoir with the same drawdown [9]. To understand this phenomenon, the phase behavior of hydrocarbons in LRS should be investigated.

The Peng–Robinson equation of state (PR EOS) [10] has been a commonly used equation-of-state (EOS) modeling in petroleum engineering. However, it becomes inaccurate when the pore diameter decreases to below several nanometers. Travalloni et al. [11, 12] have proposed an extended PR EOS (PR-C EOS) which is accurate under both macro-scale and nanoscale conditions. Although the PR-C EOS has demonstrated much higher accuracy at nanoscale condition, the required CPU time increases significantly due to its complexity. Acceleration of

phase-equilibrium calculations using PR-C EOS is required before PR-C EOS can be incorporated in reservoir simulators.

As a low-energy-density fuel [13, 14], hydrogen (H_2) requires extremely large storage volume to meet energy demand compared with conventional fossil fuels. Previous studies [15-18] have shown that geologic H_2 storage sites (including depleted gas reservoirs, salt caverns and saline aquifers) have high storage capacity. With the development of shale oil and gas, depleted shale gas reservoirs may have great potential for H_2 storage, and this possibility needs to be further investigated as well [19, 20].

1.2 Review of Current Techniques

Many studies have investigated fluid phase behavior in shale reservoirs. Commonly used methods have been experimental studies, EOS modeling, molecular simulation and machine learning (ML).

The relevant experimental studies can be categorized as the adsorption-desorption experiments, the differential scanning calorimetry (DSC) experiments, and the lab-on-a-chip experiments. The adsorption and desorption experiments are performed to observe the phase transition by measuring the sharp change of fluid density. The sudden change of fluid density is caused by the accumulation or dispersion of molecules near the boundary surface [6]. Previous adsorption-desorption experiments [21-24] showed that the saturation pressure and critical temperature would decrease under confined conditions compared with bulk (unconfined) conditions. DSC is an important thermal technique that records the exothermic and endothermic rates at different temperatures. In the process of vaporization, the fluid absorbs heat from the surrounding to vaporize and the temperature stays constant. As a result, the heat flow rate increases

sharply at this specific temperature, which can be regarded as the bubble point [25]. Luo et al. [26, 27] conducted a series of experiments using the DSC technique and found that the bubble-point temperature at the nano-scale condition is higher than the bubble-point temperature under bulk condition. The recently well-developed lab-on-a-chip technique is another effective method in the investigation of phase behavior. Several channels with different depths are etched on a chip to simulate the nanopores with various depths. The chip is placed under a high-resolution camera and the phase transition process can be observed and recorded directly [28, 29]. The experimental results show that, as the depth of the nanochannel decreases, the deviation of saturation point becomes more significant [30-32].

EOS modeling plays an important role in phase-equilibrium calculation in reservoir simulations. In petroleum engineering, PR EOS is one of the most widely used EOSs due to its accuracy and convenience. However, the accuracy of PR EOS largely decreases under nano-scale conditions. To incorporate the confinement effect into PR EOS, modifications are implemented in mainly three approaches: 1) considering the capillary pressure calculated by the Young-Laplace equation [33, 34], 2) tuning the critical parameters of pure components using experimental or molecular simulation results [35], and 3) use the generalized van der Waals theory to extend the application of PR EOS [11, 12, 36]. These modified PR EOSs have been applied to simulate fluid phase behavior in unconventional reservoirs and show better prediction results than traditional PR EOSs. Although the modified PR EOSs show higher accuracy in phase-equilibrium calculations, the required CPU time increases due to the complexity of the modified PR EOSs, especially for the PR-C EOS.

Molecular simulation precisely simulates the interactions and motions of molecules at various conditions. The commonly used molecular simulation methods in phase behavior studies

are molecular dynamics (MD) simulation and Monte Carlo (MC) simulation [6, 37-39]. MD simulation analyzes the physical movements of particles based on Newton's equations of motion. It has the advantage of investigating both equilibrium and non-equilibrium conditions [40]. MC simulation is a statistical method that performs different types of MC moves to allow the system to reach equilibrium. Different ensembles are designed to simulate various conditions. The grand canonical ensemble allows the exchange of molecules and energy and thus, the grand canonical Monte Carlo (GCMC) simulation is often used to investigate the fluid adsorption [41, 42]. Panagiotopoulos [43] developed the Gibbs ensemble Monte Carlo (GEMC) simulation by separately simulating the two phases into two boxes. The GEMC method is specifically effective in analyzing fluid density and phase behavior [44]. To investigate the confinement effect in a multi-scale system, a gauge cell with fixed volume is added to an ensemble to simulate the confined region. Vasileiadis et al. [45] proposed the gauge-GEMC method for both single-component and multi-component conditions. For a multi-component system, each component is assigned with a gauge cell which can only transfer particles with the system box. Jin et al. [46-48] proposed the gauge-GCMC method to generate phase diagrams of multi-component systems under confinement. Both of their results showed that, in a multi-scale system, the critical temperature decreases while the critical density increases compared with the bulk condition.

ML technique has been utilized in petroleum engineering to facilitate the production forecasting [49, 50], parameter optimization [51, 52], compositional simulation [53, 54], and image recognition [55, 56]. Specifically, many researchers applied the ML technique in the stabilization and acceleration of vapor-liquid equilibrium (VLE) calculations [57-60]. Gaganis and Varotsis [61] used ML to speed up the compositional reservoir simulation. For the stability test, the Support Vector Machines (SVM) algorithm [62] is used to determine the phase stability.

Instead of collecting data points uniformly, data points around the phase boundary are more likely to be selected to train the ML model while the data points with temperature higher than cricondentherm or pressure higher than cricondenbar are ignored since the fluid under that condition can be directly determined. For the phase split computation, Gaganis and Varotsis utilized feed-forward artificial neural network to estimate the equilibrium ratio, which is an important coefficient in the Rachford-Rice equation [63]. Kashinath et al. [64] improved Gaganis and Varotsis's work by applying the relevance vector machine (RVM) [65] to predict the fluid phase stability. Compared with SVM, RVM sets a discriminant threshold for the classification results. The predicted result is accepted only when the posterior probability is higher than the threshold, therefore reducing the misclassification rate. Groven et al. [53] used the ANN to predict the VLE properties of polycyclic aromatic hydrocarbons (PAHs). The predicted results were consistent with the experimental results. Mesbah et al. [66] applied the Least-Squares Support Vector Machine (LSSVM) [67] algorithm to predict the saturation pressure of CO_2/hydrocarbon binary mixture. The predicted results were in good agreement with the results calculated by EOS. Wang et al. [68] proposed a proxy flash calculator which can predict the phase condition for fluid system and provide the initial guess for phase split computation if the fluid system is in two-phase condition. Different from the previous work, the capillary pressure effect is incorporated in the proxy flash calculator to better predict the fluid phase behavior in unconventional reservoirs.

2. FLUID PHASE BEHAVIOR USING EQUATION-OF-STATE MODELING[*†]

This section describes the application of PR-C EOS to investigate hydrocarbon phase behavior in shale reservoirs during the constant volume depletion (CVD) process. The fluid compositions in the bulk region and nanopores were calculated, and the produced fluid compositions at different pressures were estimated. An in-house reservoir simulator was used for the simulations. The simulation results explain the phenomenon that oil recovery from LRS is much lower compared with oil recovery from a conventional reservoir with the same drawdown.

2.1 Methodology

Travalloni et al. [12] formulated the PR-C EOS through the generalized van der Waals theory by including the molecular descriptions of fluid-fluid and fluid-pore wall interactions. The PR-C EOS calculates the fractions of molecules in the adsorbed region and the central (unadsorbed) region. The fluid–pore wall interaction is described by a square-well potential with the parameters of square-well depth (ε_p) and square-well width (δ_p), as shown in Figure 1. The PR-C EOS reverts to PR EOS when the pore diameter reaches macro-scale conditions.

[*] Part of this chapter is reprinted with permission from "A Generalized Machine Learning-Assisted Phase-Equilibrium Calculation Model for Shale Reservoirs" by Fangxuan Chen, Sheng Luo, Shihao Wang and Hadi Nasrabadi, 2022. *Fluid Phase Equilibria*, Volume 558, Pages 113423, Copyright [2022] by Elsevier B.V.

[†] Part of this chapter is reprinted with permission from "Multiscale Pressure/Volume/Temperature Simulation of Decreasing Condensate/Gas Ratio at Greater than Dewpoint Pressure in Shale Gas-Condensate Reservoirs" by Sheng Luo, Fangxuan Chen, Dengen Zhou and Hadi Nasrabadi, 2021. *SPE Journal*, Volume 26, Pages 4174-4186, Copyright [2021] by Society of Petroleum Engineers.

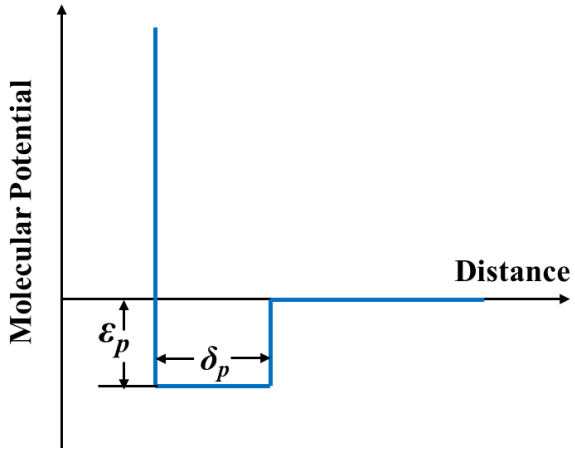


Figure 1: Square-well potential model for the fluid–pore wall interaction.

Based on the generalized van der Waals theory, the canonical partition function (Q) was calculated as follows:

$$Q(T, V, N_1, N_2, \dots, N_{nc}) = \prod_{i=1}^{nc} \left(\frac{q_i^{N_i}}{\lambda_i^{3N_i} N_i!}\right) V_f^N \exp\left(\int_{\infty}^{T} \frac{E_{conf}}{k_B T^2} dT\right) \quad (1)$$

where T denotes the temperature, V represents the total volume of the system, N is the total number of molecules, nc denotes the number of components in the system, q denotes intramolecular contribution, λ is translational contribution, V_f denotes the free volume, and E_{conf} represents the configurational energy. The free volume (V_f) is calculated based on the equations shown below:

$$V_f = V - \sum_{i=1}^{nc} \left(\frac{N_i}{\rho_{max,i}}\right) \quad (2)$$

$$\rho_{max,i}\sigma_i^3 = 1.158 - 0.479 \exp\left(0.621\left(0.5 - \frac{r_p}{\sigma_i}\right)\right) + 0.595 \exp\left(4.014\left(0.5 - \frac{r_p}{\sigma_i}\right)\right) \quad (3)$$

$$\sigma_i = \sqrt[3]{\frac{1.158}{N_{av}} b_i} \quad (4)$$

The parameters $\rho_{max,i}$ and σ_i represent the packing density of pure component i and the effective diameter of component i, respectively. The PR-C EOS divides a pore into two parts: the surface

adsorbed region (Region I and II) and the confined core region (Region III), as shown in Figure 2. The fluid molecules in the confined core region are only affected by fluid-fluid intermolecular interactions, while the fluid molecules in the surface adsorbed region are subject to both fluid-fluid and fluid–pore wall interactions. Based on this assumption, the configurational energy (E_{conf}) is modeled as follows:

$$E_{conf} = -\sum_{i=1}^{nc}\sum_{j=1}^{nc}\left(\frac{N_j}{2}N_{c,ij}\varepsilon_{ij}\right) - \sum_{i=1}^{nc}(N_i F_{p,i}\varepsilon_{p,i}) \quad (5)$$

where $N_{c,ij}$ is the coordination number, ε_{ij} denotes the interaction energy between components *i* and *j*, $\varepsilon_{p,i}$ denotes the interaction energy between component *i* and the pore wall, and $F_{p,i}$ represents the percentage of molecules of component *i* in the surface-adsorbed region. The first part of Equation (5) represents the fluid-fluid interactions, while the second part represents the fluid–pore wall interactions. The term $F_{p,i}$ is modeled by the empirical expression as shown below:

$$F_{p,i} = F_{pr,i} + (1 - F_{pr,i})(1 - \exp(-\frac{\varepsilon_{p,i}}{k_B T}))(1 - \frac{x_i \rho}{\rho_{max,i}})^{\theta^i} \quad (6)$$

$$F_{pr,i} = \frac{(r_p - \sigma_i/2)^2 - (r_p - \sigma_i/2 - \delta_{p,i})^2}{(r_p - \sigma_i/2)^2} \quad (7)$$

$$\theta_i = \frac{r_p}{\delta_{p,i} + \sigma_i/2} \quad (8)$$

The term $F_{pr,i}$ is the volume fraction of the surface-adsorbed region accessible to the mass center of the fluid molecules in a pore (calculated as the volume ratio of Region II over Regions II and III), which is affected by the value of $\delta_{p,i}$. The term $\delta_{p,i}$ is the square-well width of the fluid–pore wall interaction potential of component *i*. When the distance between a fluid molecule and a pore wall is less than $\delta_{p,I}$, the fluid molecule is subject to fluid–pore wall interaction. In this paper, the square-well width (δ_p) of different types of hydrocarbons was set as $0.5\sigma_i$, which is widely used for square-well potential models in the existing literature [69, 70].

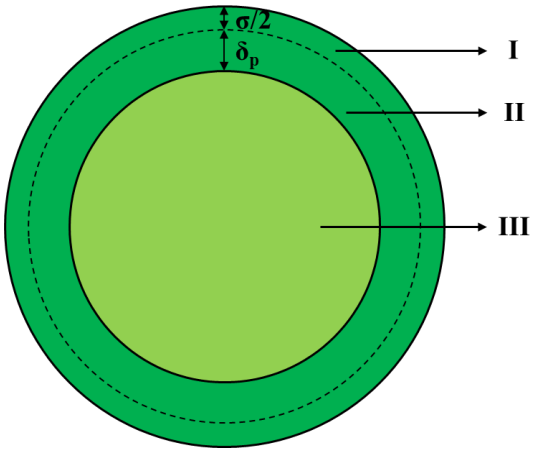


Figure 2: Schematic figure of a cylindrical pore. The dark green regions (Regions I and II) are the surface-adsorbed regions, while the light green region (Region III) is the confined core region. Region II represents the surface-adsorbed region that is accessible to the mass center of the fluid molecules. F_{pr} is defined as the volume ratio of Region II over Regions II and III.

The Helmholtz free energy (A) and chemical potential of component i (μ_i) were obtained using the following thermodynamic relations:

$$A(T, V, N_1, N_2, \dots, N_{nc}) = -k_B T ln Q(T, V, N_1, N_2, \dots, N_{nc}) \tag{9}$$

$$\mu_i = \left(\frac{\partial A}{\partial N_i}\right)_{T,V,N_{j\neq i}} \tag{10}$$

The confinement-modified energy parameter ($a_{p,ij}$) between components i and j and the confinement-modified volume parameter ($b_{p,i}$) of component i were calculated as follows:

$$a_{p,ij} = \sqrt{a_i a_j}\left(1 - \frac{2}{5}\frac{\sigma_{ij}}{r_p}\right) \tag{11}$$

$$b_{p,i} = \frac{N_{av}}{\rho_{max,i}} \tag{12}$$

Applying the mixing rule based on the mole fraction of each component in a mixture, the confinement-modified energy parameter (a_p) and the confinement-modified volume parameter (b_p) of the fluid mixture were calculated as follows:

$$a_p = \sum_{i=1}^{nc} \sum_{j=1}^{nc} (x_i x_j a_{p,ij}) \tag{13}$$

$$b_p = \sum_{i=1}^{nc} x_i b_{p,i} \tag{14}$$

The parameters of the PR EOS used in this paper were determined by the equations shown below:

$$m_i = \begin{cases} 0.3746 + 1.5423w_i - 0.2699w_i^2 & 0 < w_i \leq 0.5 \\ 0.3796 + 1.4850w_i - 0.1644w_i^2 + 0.01667w_i^3 & w_i > 0.5 \end{cases} \tag{15}$$

$$a_i = \frac{0.4572R^2T_{c,i}^2}{P_{c,i}} (1 + m_i(1 - \sqrt{\frac{T}{T_{c,i}}}))^2 \tag{16}$$

$$b_i = \frac{0.07780RT_{c,i}}{P_{c,i}} \tag{17}$$

2.2 Model Description

Shale reservoirs possess a complex and heterogeneous pore size distribution ranging from the nanoscale to the macroscale. Macroscale regions include large natural fractures and hydraulic fractures, and their widths are in the range of millimeters [71, 72]. Small natural fractures and macropores form the mesoscale region, with a width or diameter in the order of micrometers [73]. Nanopores in the organic matter contribute to the nano-porosity of the shale rock [74]. Two types of techniques have been applied to characterize shale pore geometry: fluid intrusion methods and imaging techniques. Fluid intrusion methods consist of mercury injection capillary pressure (MICP) analysis and low-pressure adsorption (LPA) tests. MICP can characterize pores with diameters higher than 2 nm, while low-pressure CO_2 adsorption is able to determine the microporosity of pores less than 2 nm. Together, they can present a whole view of the pore size distribution [3]. Imaging techniques include nuclear magnetic resonance (NMR) [75], focused ion

beam scanning electron microscopy (FIB-SEM) [76], transmission electron microscopy (TEM) [77], and atomic force microscopy (AFM) [78]. Although we can roughly make a classification of pores and fractures, the width or diameter values may vary significantly from one formation to another.

Luo et al. [79] presented a multi-scale PVT model to investigate the phase behavior and depletion process in shale porous media. In the model, the porous space is divided into two parts: the bulk region and nanoscale region. The bulk region refers to natural fractures, hydraulic fractures, and macropores, which are on the macroscale ($d \geq 50$ nm). The nanoscale region includes nanopores ($d < 50$ nm). The nanoscale region is further divided into several regions with specific diameters. In this study, we assume three regions in the multi-scale PVT model: one bulk region and two nanopore regions. The bulk region has a diameter of 10 μm, while the two nanopore regions have diameters of 15 nm and 5 nm. The initial volumes of the three regions apply a ratio of $V_{bulk}:V_{15nm}:V_{5nm} = 3:2:1$ [79]. We adopt this pore size distribution scenario as the base case in multi-scale simulations. A sketch is shown in Figure. 3.

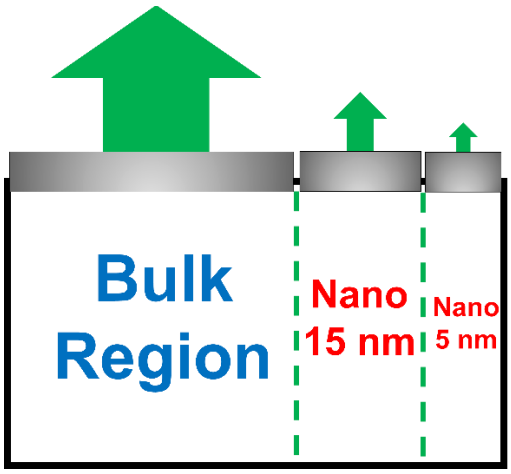


Figure 3: The multi-scale model used in this work with a volume ratio as V_{bulk} : V_{15nm} : V_{5nm} = 3:2:1.

2.3 Reservoir Conditions and Fluid Properties

Two cases are analyzed in this paper, one for a shallow reservoir and the other for a deep reservoir. Reservoir conditions are shown in Table 1 and the reservoir simulation parameters are listed in Table 2.

Table 1: Reservoir conditions of the shallow and deep gas condensate cases

Properties	Case 1 (Shallow reservoir)	Case 2 (Deep reservoir)
Reservoir temperature (°F)	165	210
Initial reservoir pressure (psi)	4,000	8,000
Pore size distribution	V_{bulk} : V_{15nm} : V_{5nm} = 3:2:1	V_{bulk} : V_{15nm} : V_{5nm} = 3:2:1
Reservoir fluid type	Gas condensate	Gas condensate

Table 2: Parameters for reservoir simulation in the shallow and deep reservoir cases

Properties	Values	Unit
n_{og}	2	
n_g	2	
K_{rog} at S_{gc}	1	fraction
K_{rg} at S_{org}	0.8	fraction
S_{org}	0.15	fraction

S_{gc}	**0.03**	**fraction**
Rock Compressibility	**6.0×10^{-6}**	**1/psi**

The bulk-state phase diagram is shown in Figure 4. The reservoir fluids are initially in supercritical state and will go through an isothermal depletion.

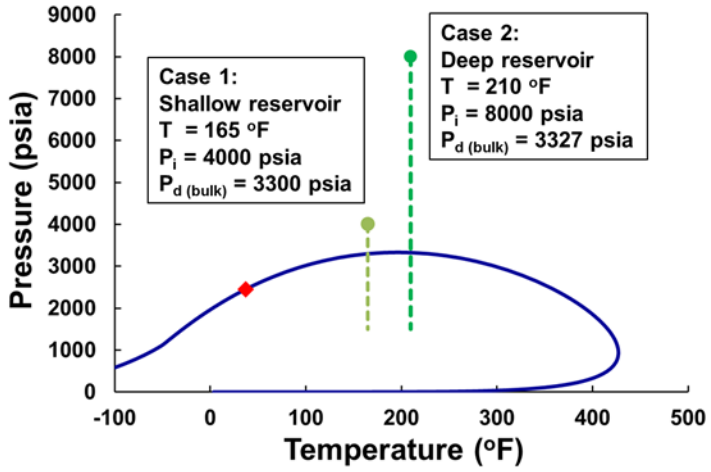


Figure 4: Bulk-state phase diagrams of the shallow and deep reservoirs.

The gas condensate fluid uses a synthetic model of six-pseudo-component system to represent typical condensate sample's compositions from lab test. The parameters of PR-C EOS are listed in Tables 3 and 4. The critical temperature, critical pressure, acentric factors, volume shift parameters, and binary interaction parameters are bulk-state parameters, while the fluid–wall interaction parameters (ε_p) and square well width (δ_p) are confinement parameters. ε_p is obtained by interpolating or extrapolating the fluid–pore surface affinity factor from the work by Luo et al. [79]. Based on the studies conducted by Vega et al. [80], they found that large values of δ_p/σ give deviated phase equilibrium and a range between 0.25 to 0.75 leads to the most accurate results.

Herein, δ_p/σ is assumed to be 0.5 for all pseudo-components, which have generally been used for square well potential models in previous works [69, 70, 81].

Table 3: Extended PR EOS properties for the gas condensate

Species	MW (g/mol)	Tc (R)	Pc (psia)	ω	VSP	ε_p/k_B (K)	δ_p/σ	Mole fraction
C_1-N_2	16.14	343	667	0.008	-0.154	229	0.5	0.6442
C_2-CO_2	30.19	550	711	0.099	-0.124	486	0.5	0.1738
C_3	44.10	666	616	0.152	-0.086	1210	0.5	0.0770
C_4-C_5	62.67	785	522	0.193	-0.051	1836	0.5	0.0537
C_6-C_{10}	103.44	1064	413	0.351	0.029	2044	0.5	0.0372
C_{11}^+	198.56	1285	246	0.603	0.133	2323	0.5	0.0141

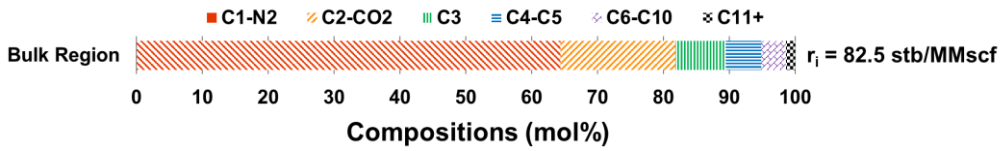
Table 4: Binary interaction parameters for the gas condensate

BIP	C_1-N_2	C_2-CO_2	C_3	C_4-C_5	C_6-C_{10}	C_{11}^+
C_1-N_2	0.000	0.003	0.009	0.015	0.034	0.063
C_2-CO_2	0.003	0.000	0.002	0.005	0.018	0.041
C_3	0.009	0.002	0.000	0.001	0.009	0.026
C_4-C_5	0.015	0.005	0.001	0.000	0.004	0.018
C_6-C_{10}	0.034	0.018	0.009	0.004	0.000	0.005
C_{11}^+	0.063	0.041	0.026	0.018	0.005	0.000

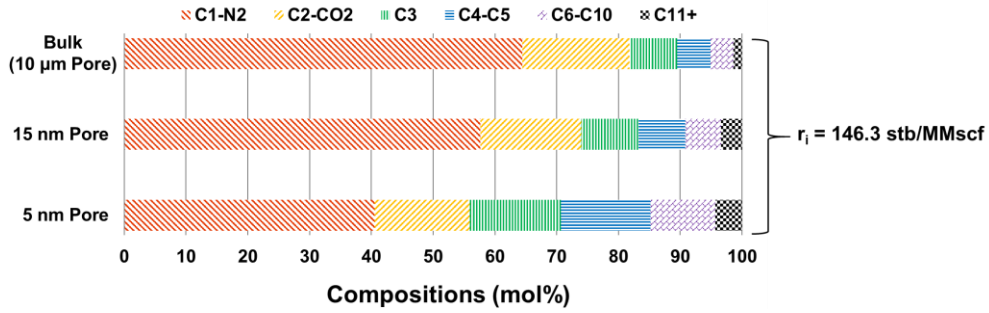
2.4 Results and Analysis

The compositions in both the single bulk-scale and multi-scale shale models of the two cases are compared under initial reservoir conditions. The composition profiles are shown in Figures. 5 and 6. The given mole fractions of the pseudo-components in Table 3 are used as the initial bulk region compositions. It is assumed that after geologic time the fluids in different sizes of pores are in thermodynamic equilibrium. The fluid compositions in 15 and 5 nm pores can be obtained by solving the chemical equilibrium equation shown below:

$$\mu_{i,bulk} = \mu_{i,15nm} = \mu_{i,5nm} \forall i \in 1,2,\dots,NC \tag{18}$$

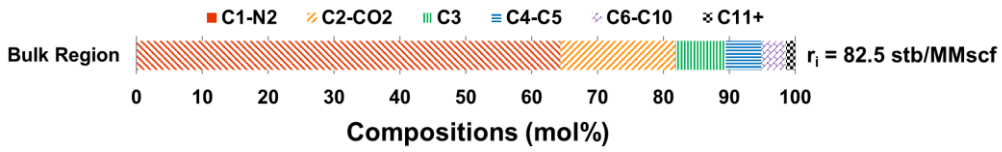


(a) Single bulk model

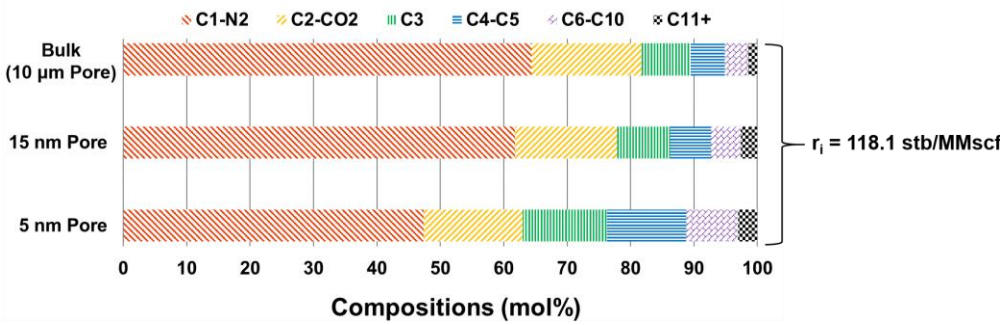


(b) Multi-scale model

Figure 5: Fluid compositions in the shallow reservoir in the (a) single bulk model and (b) multi-scale model



(a) Single bulk model



(b) Multi-scale model

Figure 6: Fluid compositions in the deep reservoir in a (a) single bulk model and (b) multi-scale model

In the multi-scale PVT model, we assume a set of two-stage separators under the conditions of 200 psia, 100 °F and 50 psia, 80 °F. During the depletion process, fluids produced from system are assessed with flash calculations with separators conditions and stock tank condition (14.696 psia and 60 °F). Typical reservoir properties are characterized based on the above process for both single bulk-scale and multi-scale models. For the shallow reservoir, the initial gas formation volume factors ($B_{g,i}$) are 0.716 and 0.776 rb/Mscf, and the initial solution oil-gas ratios (r_i) are 82.5 and 146.3 stb/MMscf for single bulk-scale and multi-scale models, respectively. The reservoir fluid compositions in the nanopores differ from the compositions in the bulk region. Fewer light hydrocarbons (C_1-C_2) but significantly more heavier hydrocarbons (C_{6+}) are observed in the nanopores due to pore adsorption (Figures. 5 and 6). It should be noted that the composition shifts of components occur in different relative ratios: in the shallow reservoir case, C_1-C_2 decreases from 82% at bulk to 74% at 15 nm and 56% at 5nm, while C_{6+} significantly increases from 5% at bulk to 9% at 15 nm and 15% at 5 nm (Figure. 5). Such drastic composition shifts result in change

to original oil and gas in-place: the original gas in place (OGIP) of the multi-scale model decreases 7.7%, while the original oil in place (OOIP) of the multi-scale model increases 63.8% when compared with the corresponding values of the single bulk-scale model. In the deep reservoir, similar trends can be observed. The values of $B_{g,i}$ are 0.603 and 0.645 rb/Mscf, and the values of r_i are 82.5 and 118.1 stb/MMscf for the single bulk-scale and multi-scale models, respectively. The OGIP of the multi-scale model decreases 6.6%, while the OOIP of the multi-scale model increases 33.8% when compared with the corresponding values of the single bulk-scale model. The results show that the existence of nanopores can slightly decrease the OGIP but significantly increase the OOIP, since nanopores contain more intermediate and heavy components. The compositional differences will affect the produced fluids as long as the depletion zone penetrates into the shale matrix. According to the comparison of Figures. 5(b) and 6(b), at a high pressure, such compositional differences are reduced among the three regions, because the pores will be more densely packed under higher pressure, which suppresses the selectivity effect of the nanopores. The OOIP increment is smaller in the deep reservoir case than in the shallow reservoir case, which is associated with the lower fractions of heavy components in the nanopores.

Based on the PR-C EOS, each pore is divided into two regions: the core region and surface-adsorbed region. A sketch of a pore is shown in Figure. 2. Fluid molecules in the core region only interact with fluid molecules, while fluid molecules in the surface-adsorbed region are affected by both fluid molecules and wall molecules. The distribution of fluid molecules is determined by the PR-C EOS considering the effects of temperature, fluid density, confinement degree, and the fluid–wall interaction parameter. Compositions in the multi-scale regions of the two cases are shown in Figures. 7 and 8. Regardless of pore size, the surface-adsorbed regions are rich in intermediate to heavy hydrocarbons (C_3-C_{11+}), while more light hydrocarbons (C_1 and C_2) end up in the core

region. Alfi et al. [82] applied the PR-C EOS and also found the existence of compositional heterogeneity between the small and large pores. The steric hindrance has more significant effect in small pores (d < 6 nm) while the energetic effect becomes dominant in relatively large pores (6 nm < d < 40 nm). The fluid behaves bulk-like when the pore diameter is larger than 40 nm. However, in terms of the total mass percentage in the two regions, the pore size plays a dominant role. For macroscale pores, nearly all molecules gather in the core region. When the diameters of the pores decrease to the nanometer scale, the confinement effect plays a significant role on the fluid phase behavior, as there are more molecules distributed in the surface-adsorbed region.

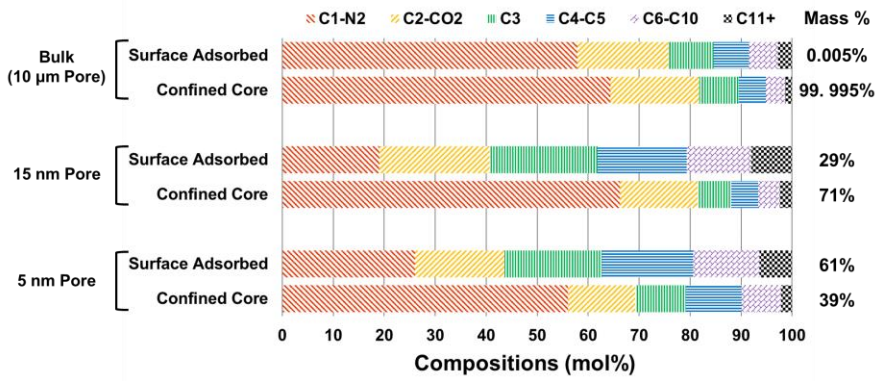


Figure 7: Compositions in the surface-adsorbed region and the core region of the shallow reservoir

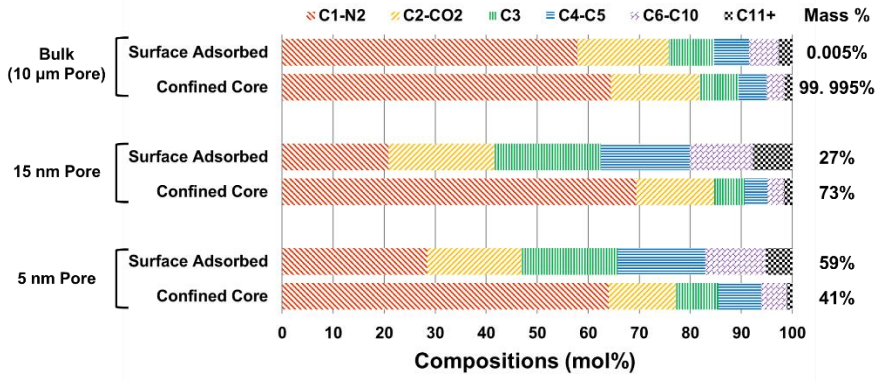


Figure 8: Compositions in the surface-adsorbed region and the core region of the deep reservoir

Constant volume depletion (CVD) processes are simulated using the multi-scale PVT model. First, the fluid phase equilibriums for the different porous geometry sizes are calculated to determine the compositions in each pore. The pore-size-dependent transmissibility (T_a) [83] is defined as:

$$T_a = \frac{A_c K K_r}{\mu} \tag{19}$$

where A_c is the cross-sectional area of cylindrical pores, and K is the pore-size-dependent permeability. Based on the pore-size-dependent transmissibility, the volume increment of pore i (ΔV_i) is determined by

$$\Delta V_i = \Delta V \cdot \frac{T_{a,i}}{\sum_{j=1}^{NR} T_{a,j}} \tag{20}$$

where ΔV denotes the volume increment of the whole system. For a CVD process, the ending volume of pore i ($V_{end,i}$) is first calculated using Equation (21) by incorporating the rock compressibility factor (c_f). The parameters $V_{init,i}$, P_{end}, and P_{init} denote the initial volume of pore i, the ending pressure, and the initial pressure, respectively. Then, the depleted fluid volume of pore i ($V_{depleted,i}$) is calculated by Equation (22), and the depleted fluid compositions of pore i can be determined according to $V_{depleted,i}$. The depleted fluid moles for hydrocarbon j ($n_{j,depleted}$) are the sum of the depleted fluid mole from each pore i, as shown in Equation (23).

$$V_{end,i} = V_{init,i} \cdot exp\ (c_f(P_{end} - P_{init})) \tag{21}$$

$$V_{depleted,i} = V_{init,i} + \Delta V_i - V_{end,i} \tag{22}$$

$$n_{j,depleted} = \sum_{i=1}^{NR} V_{depleted,i} c_{j,i} \tag{23}$$

Since the transmissibility in the bulk region is much higher than the nanopores, the depleted system volume is primarily governed by the depleted volume in the bulk region. Therefore, the moles of depleted fluid of the total system are approximately the same as the depleted moles from the bulk region. The depletion CGR is approximately the same as bulk region fluid oil-gas solution ratio (r_{bulk}).

In the shallow reservoir case ($T_i = 165°F, P_i = 4000\,psia$), the compositions of each component in both the single bulk and multi-scale model under different pressures are shown in Figures. 9 and 10.

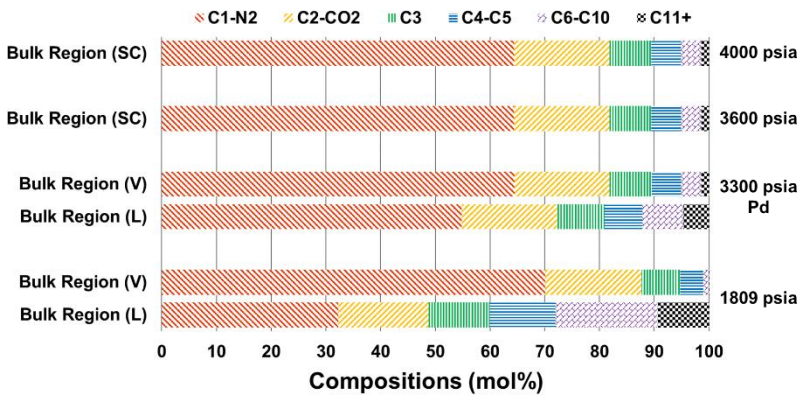


Figure 9: Hydrocarbon compositions of single bulk model under different pressures in the shallow reservoir (L: liquid, V: vapor, SC: supercritical, P_d: dew-point pressure)

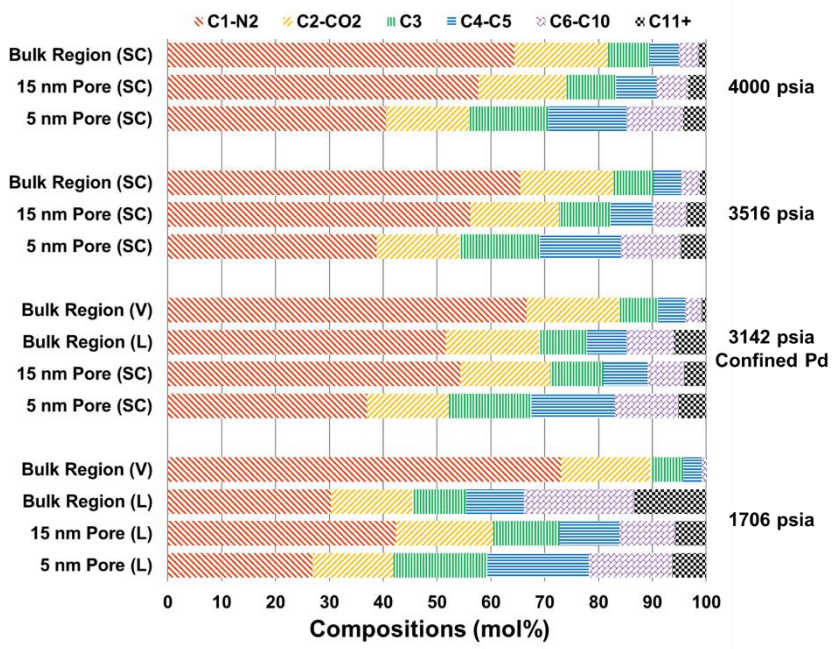


Figure 10: Hydrocarbon compositions of multi-scale model under different pressures in the shallow reservoir (L: liquid, V: vapor, SC: supercritical, confined P_d: dew-point pressure in the confined system)

Due to the nanopore effect, the depletion process of multi-scale system occurs differently than the single bulk system (Figures. 9 and 10). Above dew point, the nanopores selectively release light hydrocarbons as the pressure decreases. As a result, the bulk region hydrocarbon fluid becomes learner and the depletion CGR decreases when the pressure is above dew point, as shown in Figure. 11. This behavior is consistent with field observations from the productions of shale gas condensate reservoirs [84, 85]. Such CGR behavior cannot be modeled with bulk thermodynamics modeling, which predicts a constant production CGR above dew point. It should be noted that in the multi-scale model the initial depletion condensate–gas ratio (CGR_i = 82.5 *stb/MMscf*) is lower than the initial solution oil–gas ratio (r_i = 146.3 *stb/MMscf*), as shown in Figure. 5. The reason is that the depleted fluid compositions are governed by the lean bulk region with a high

transmissibility. The decrease of CGR is sensitive to pore size distribution: as the volume percentage of bulk volume decreases from 70% to 30%, the CGR decline becomes steeper due to the increased nanopore effect. The continuous decrease in the CGR suggests that, in field production, CGR production will decrease above the dew-point pressure as the depletion zone penetrates into the shale matrix.

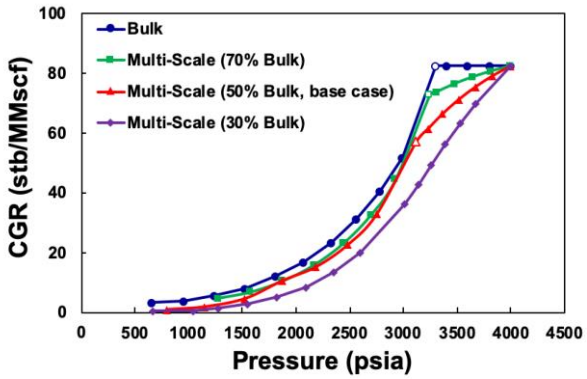


Figure 11: Relations of CGR and pressure of single bulk and multi-scale models in the shallow reservoir (unshaded points represent the dew pressures). Bulk: pure bulk fluid; multi-scale cases: bulk region volume percentage as 30%, 50% and 70%. 15 and 5 nm pore volume are kept in ratio 2:1 for all cases. 50% bulk volume one is considered as the pore-size-distribution base case and also applied in the deep reservoir scenario.

As the bulk region composition changes due to the nanopore effect, the dew-point pressure shifts accordingly. We observe that the dew-point pressure of multi-scale system shifts to 3142 psia compared with the single bulk system dew point of 3300 psia. It is worth noting that dew point only occurs within the bulk region and there is no liquid drop-out happens in nanopore. Throughout the process, the fluid in nanopores stays at single-phase state. Compositional differences become more remarkable at low pressures, as the bulk region is rich in light ends while

the nanopores accumulate intermediate and heavy components. This phenomenon indicates that, during the depletion process, a fraction of the intermediate and heavy components is trapped in the nanopores, which results in condensate recovery loss.

In the deep reservoir case ($T_i = 210°F, P_i = 8000\ psia$), compositions of each component in both the bulk and multi-scale models under different pressures are shown in Figures. 12 and 13.

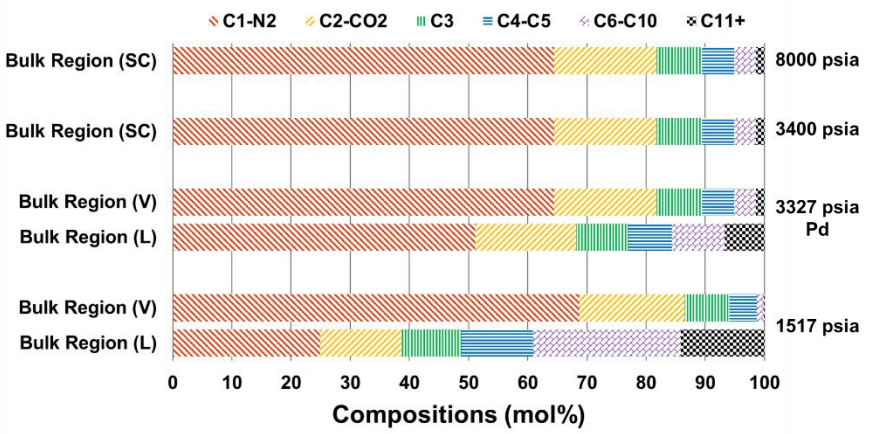


Figure 12: Hydrocarbon compositions of single bulk model under different pressures in the deep reservoir (L: liquid, V: vapor, SC: supercritical)

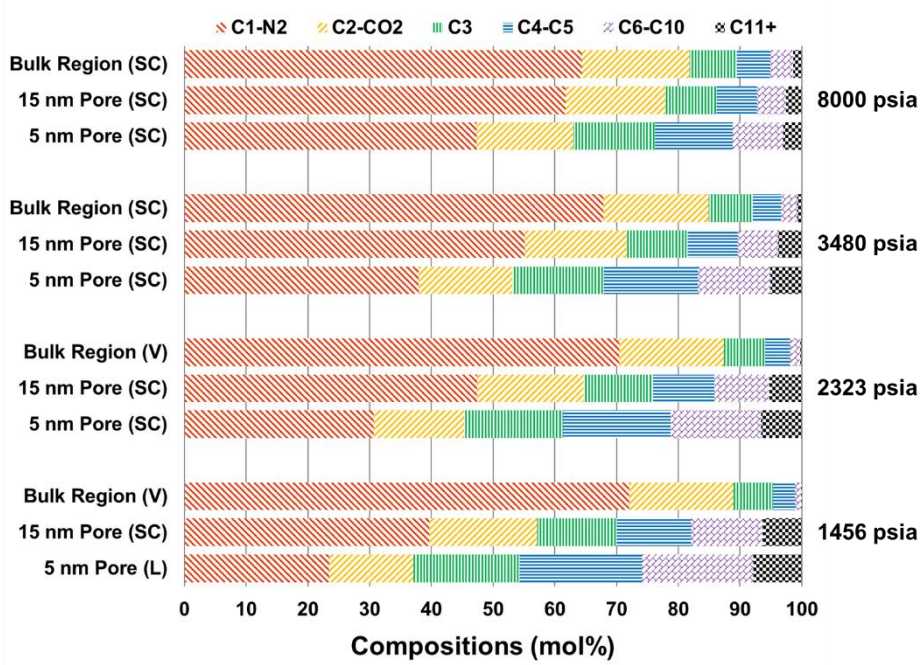


Figure 13: Hydrocarbon compositions of multi-scale model at different pressures in the deep reservoir (L: liquid, V: vapor, SC: supercritical)

The bulk model has a dew-point pressure of 3327 psia, while there is no single- to two-phase transition occurring in any region of the multi-scale model. Supercritical fluids in the bulk region and nanopores turn into subcritical liquid or vapor as the pressure in the corresponding region drops below the critical pressure. In the multi-scale model, the fluid in the bulk region is generally leaner than the fluid in the nanopores, and it becomes increasingly leaner as the pressure decreases. Compositional selectivity between the different regions causes the nanopores to be rich in intermediate and heavy components, which is similar to the shallow reservoir case.

The CGRs of the bulk and multi-scale models under different pressures are summarized in Figure. 14. Similar to the trend of the shallow reservoir case, the bulk CGR remains constant when the pressure is above the dew-point pressure, while multi-scale CGR decreases upon the start of depletion.

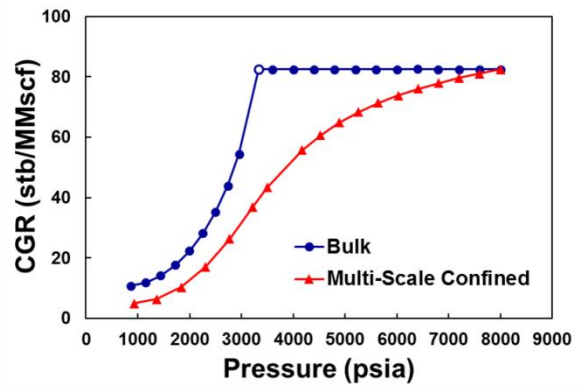


Figure 14: Relations of CGR and pressure of single bulk and multi-scale models in the deep reservoir (Empty points represent the dew pressures)

3. ACCELERATING PHASE-EQUILIBRIUM CALCULATION USING MACHINE LEARNING[*†]

In this section, we propose a generalized ML-assisted phase-equilibrium calculation model to accelerate phase-equilibrium calculation. In our traditional reservoir simulator, the phase-equilibrium calculation is performed using PR-C EOS, which is accurate but time-consuming. The introduction of ML techniques could significantly reduce the running time without decreasing accuracy unacceptably. In this regard, we have found that the combination of PR-C EOS and ML techniques can allow the proxy calculation model to conduct phase-equilibrium calculation in a rapid and accurate way.

3.1 Methodology

The methodologies of traditional phase-equilibrium calculation and ML models are explained in detail in this section.

3.1.1 Traditional Phase-Equilibrium Calculation

The phase-equilibrium calculation consists of two parts: stability test and phase-split computation. We need to first determine if the fluid is in single-phase or two-phase condition, which is the aim of the stability test. If the fluid is in two-phase condition, the phase split computation will be conducted to obtain the compositions in the vapor and liquid phases.

[*] Part of this chapter is reprinted with permission from "A Generalized Machine Learning-Assisted Phase-Equilibrium Calculation Model for Shale Reservoirs" by Fangxuan Chen, Sheng Luo, Shihao Wang and Hadi Nasrabadi, 2022. *Fluid Phase Equilibria*, Volume 558, Pages 113423, Copyright [2022] by Elsevier B.V.

[†] Part of this chapter is reprinted with permission from "A Novel Machine-Learning Assisted Phase-Equilibrium Calculation Model for Liquid-Rich Shale Reservoirs" by Fangxuan Chen, Sheng Luo, Shihao Wang and Hadi Nasrabadi, 2023. SPE Proceedings, SPE-212193-MS, Copyright [2023] by Society of Petroleum Engineers.

The basic principle of stability test using the PR-C EOS is the equality of the chemical potential of component *i* in vapor and liquid phase, as shown in the equation below:

$$\mu_i^V = \mu_i^L \tag{24}$$

where μ_i^V and μ_i^L represent the chemical potential of component *i* in vapor and liquid phase, respectively. We first assume that the fluid is in single-phase condition and calculate the chemical potential of each component (μ_i^I). Then we search for every possible set of composition that has the same chemical potential (μ_i^{II}) as μ_i^I. The Newton-Raphson iteration is used in the searching program. If the new phase composition (n_i^{II}) is the same as the initial single-phase composition (n_i^I), the solution is trivial and the fluid is in single-phase condition. Otherwise, it is in two-phase condition. If the fluid is in two-phase condition, the phase split computation will be conducted to obtain the compositions in the vapor and liquid phase. The flowchart of the stability test is shown in Figure 15.

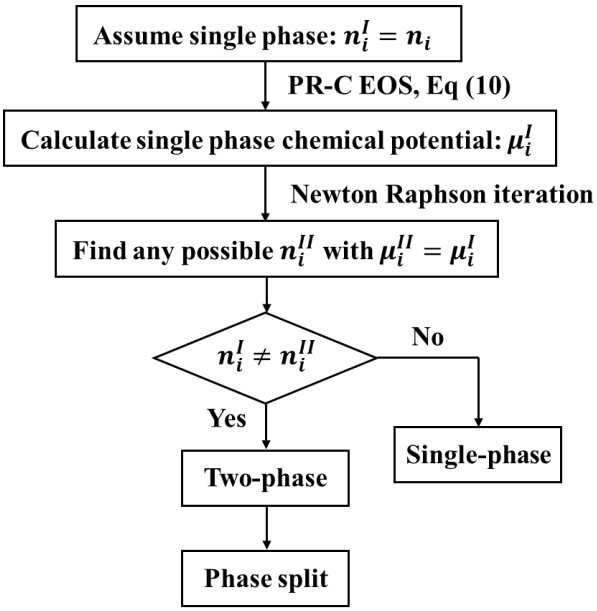


Figure 15: Flowchart of stability test using the PR-C EOS. The fluid is first assumed to be in single-phase condition and the chemical potential of each component (μ_i^I) is calculated. Then every possible set of composition that has the same chemical potential (μ_i^{II}) as μ_i^I is searched using the Newton-Raphson iteration. If the new phase composition (n_i^{II}) is the same as the initial single-phase composition (n_i^I), the solution is trivial and the fluid is in single-phase condition. Otherwise, it is in two-phase condition. If the fluid is in two-phase condition, the phase split computation will be conducted to obtain the compositions in vapor and liquid phase.

The phase split computation is performed by directly minimizing the total Helmholtz free energy (A_{total}), which is calculated by the equation below:

$$A_{total} = A^I + A^{II} \tag{25}$$

where A^I and A^{II} represent the Helmholtz free energy of phase I and II, which is calculated using Equation (9). The variables of the minimization program are equilibrium ratio of each pseudo component (K_i), which can be used to calculate the number of moles of each component (n_i) in the vapor and liquid phases. The initial guess of equilibrium ratio was calculated based on n_i^I and n_i^{II} obtained from the stability test. Since the fluid is determined to be in two-phase condition, the total Helmholtz free energy is minimized by changing the values of K_i. The lowest total Helmholtz free energy corresponds to the desired value of K_i.

3.1.2 Machine Learning Models

In the stability test, the multi-layer perceptron (MLP) was applied to predict fluid phase behavior due to its high flexibility and efficiency. In the phase split computation, the physics-

informed neural network (PINN) was applied to incorporate the PR-C EOS into the ML model and the neural networks were trained by directly minimizing the total Helmholtz free energy.

An artificial neural network (ANN) is a computing system that models information processing by a biological neural network. The basic unit of an ANN is the artificial neuron; "signals" (real numbers) can be transmitted between different neurons [86]. One common type of ANN is the MLP, which consists of at least three layers: an input layer, a hidden layer and an output layer [87]. The layers in MLP are fully connected, which means that a node in the previous layer is connected to all nodes in the next layer. Each node has a weight value and a bias value, reflecting the importance of the node. The training algorithm in MLP is the backpropagation (BP) algorithm [88], which focuses on the derivative of the loss function with respect to the weight (or bias). The connections between nodes in the adjacent layers are augmented by activation functions to include nonlinearity in the system. The commonly used activation functions in MLP include hyperbolic tangent (tanh), sigmoid and rectified linear unit (ReLU) [89]. During the training process, the model is optimized by continuously changing the weight and bias values of each node. The schematic illustration of MLP is shown in Figure 16.

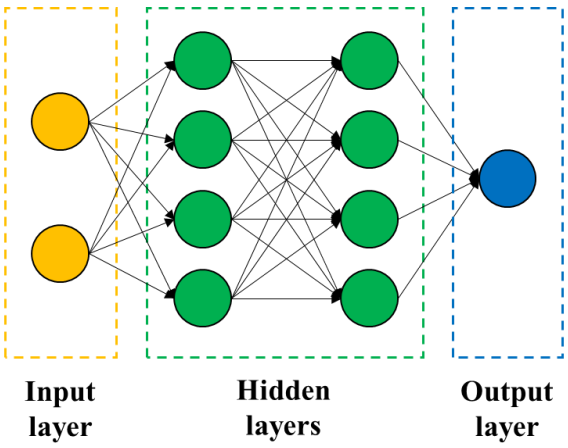


Figure 16: The schematic illustration of MLP. MLP consists of at least three layers: an input layer, a hidden layer and an output layer. The layers in MLP are fully connected layers, which means that a node in the previous layer is connected to all the nodes in the next layer. Each node has a weight value and a bias value. The connections between nodes in the adjacent layers are augmented by activation functions.

PINNs are widely used in different areas, for example fluid mechanics [90] and heat transfer [91]. The structure of a PINN is shown in Figure 17. A PINN mainly consists of two parts: the deep neural network and the partial differential equation (PDE) component [92]. The deep neural network includes one input layer, one or more hidden layers and one output layer. The output parameters in the deep neural network are functions of the input variables, connected by the weights and biases of the neurons. In the PDE portion, the derivatives of output parameters with respect to input variables are used as the target function. The loss function of a PINN may include the output parameters of the neural network, the target function, the initial conditions or the boundary conditions. In the training process of a PINN, the loss is continuously minimized by changing the weights and biases of the neurons. It should be noted that the training of a PINN does not require a data set [92]. By directly minimizing the loss function, the neural network plays a role as the function approximator but not works as a predictor between the input and output variables.

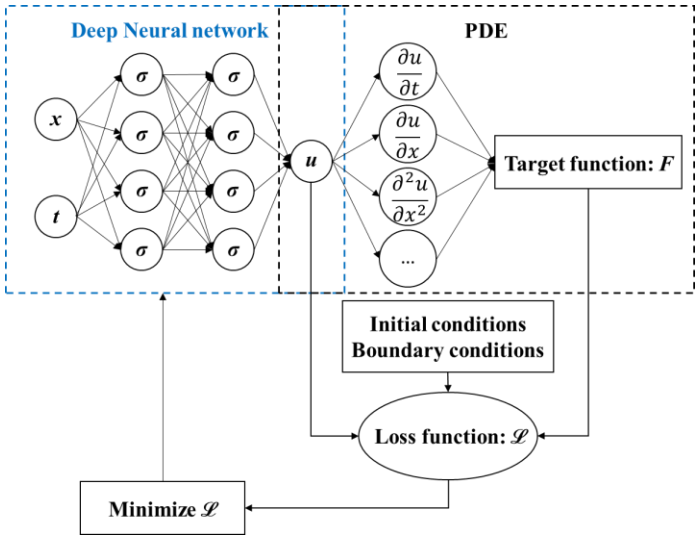


Figure 17: The schematic illustration of PINN. In the deep neural network (shown in blue square), the output is a function of the input variables. The derivatives of output with respect to input variables follow physical laws and can be calculated by PDEs (shown in black square). The loss function of PINN may include the loss of the output of neural network, initial conditions, boundary conditions and PDE results.

3.2 Data Generation

The generation of the dataset is detailly explained in this part. Six pseudo-components are selected as the typical fluid composition. A large range of parameters are considered to include various fluid types.

3.2.1 General Set of Pseudo Components

In the generalized ML-assisted phase-equilibrium calculation model, instead of using a specific set of real components, a general set of pseudo-components with the PR-C EOS parameters is considered. Based on the literature review of compositional analysis [93-95], we select 6 pseudo-components to be the representative composition for different types of reservoir

fluids, including C_1-N_2, C_2-CO_2-C_3, C_4-C_6, C_7-C_{12}, C_{13}-C_{24}, C_{25+}. When the fluid composition is given, it can be lumped into the six pseudo-components and the PR EOS parameters of each pseudo component (n_i, $P_{c,i}$, $T_{c,i}$, MW_i, w_i, VSP_i, and $BIP_{i,j}$) are obtained. If these PR EOS parameters are considered as the input variables of the ML model, the ML model will be suitable for any fluid composition, which extends the application of our approach from fixed-fluid composition to generalized conditions.

3.2.2 Data Range and Acquisition

The input parameters of the generalized ML-assisted phase-equilibrium calculation model include P, T, r_p, n, P_c, T_c, MW, w, VSP, ε_p and BIP. The range of P and T follows the reference [96] and the range of r_p is based on the pore size distribution of shale reservoirs [3], as shown in Table 5. The lower bound of r_p is set as 2nm because the pore diameter should be larger than the effective diameter (σ) of fluid molecules, which is required by the PR-C EOS.

Table 5: Range of parameters of reservoir conditions

Parameter	Unit	Minimum	Maximum
Pressure (P)	**MPa**	**1**	**80**
Temperature (T)	**°C**	**40**	**100**
Pore radius (r_p)	**nm**	**2**	**50**

The values of PR EOS parameters (n, P_c, T_c, MW, w, VSP, and BIP) are determined by applying the mixing rule based on the mole fraction of each real component in a pseudo component, as shown in the equation below:

$$S_{pseudo} = \sum x_{real,i} \cdot S_{real,i} \quad (26)$$

where *S* denotes one of the PR EOS parameters and $x_{real,i}$ is the mole fraction of the real component *i* in a pseudo component. Each pseudo component is categorized under several real components for the simplicity of data generation, as shown in Table 6. The values of PR EOS parameters of real components are obtained from the previous work [97]. To clarify Equation (26), an example of calculating the critical pressure of C_2-CO_2-C_3 is shown below:

$$P_{c,C_2-CO_2-C_3} = P_{c,C_2} \cdot x_{C_2} + P_{c,CO_2} \cdot x_{CO_2} + P_{c,C_3} \cdot x_{C_3} \quad (27)$$

$$x_{C_2} + x_{CO_2} + x_{C_3} = 1 \quad (28)$$

where $P_{c,C_2-CO_2-C_3}$, P_{c,C_2}, P_{c,CO_2}, and P_{c,C_3} denote the critical pressure of C_2-CO_2-C_3, C_2, CO_2, and C_3 respectively and x_{C_2}, x_{CO_2}, and x_{C_3} represent the mole fraction of C_2, CO_2, and C_3 in C_2-CO_2-C_3.

Table 6: The lumped real components of 6 pseudo components

Pseudo components	Real components
C_1-N_2	**C_1**
	N_2
C_2-CO_2-C_3	**C_2**
	CO_2
	C_3
C_4-C_6	**C_4**
	C_5
	C_6

C_7-C_{12}	**C_7**
	C_9
	C_{12}
C_{13}-C_{24}	**C_{13}**
	C_{19}
	C_{24}
C_{25+}	**C_{25}**
	C_{35}
	C_{45}

The PR-C EOS parameter (ε_p) is calculated based on the equation [79] shown below:

$$\kappa = \frac{\varepsilon_p \delta_p}{MW} \tag{29}$$

where κ is the normalized fluid-pore surface affinity factor, which can be obtained from the reference [79]. The value of ε_p of a pseudo-component is calculated using the same method as shown in Equation (26).

In summary, for a six-pseudo-component system, each parameter of *n*, P_c, T_c, *MW*, *w*, *VSP*, and ε_p contains 6 values. For the matrix of *BIP*, it has a symmetric distribution and the values of the diagonal elements in the matrix are zeros. As a result, for a six-pseudo-component system, 36 elements in the *BIP* matrix can be described by 15 different values. Together with the parameters of reservoir conditions (*P*, *T*, r_p), the total number of input parameters for the generalized model is 60. The ranges of different parameters of the PR-C EOS are summarized in Table 7. The critical

temperature starts from 5.2 K for helium and can be too high to measure. Therefore, the critical temperature can be negative for hydrocarbon mixtures.

Table 7: Range of parameters of the PR-C EOS

Parameter	Unit	Minimum	Maximum
Critical pressure (P_c)	**Pa**	**736,768**	**5,455,021**
Critical temperature (T_c)	**K**	**175**	**954**
Molecular weight (MW)	**g/mol**	**16**	**533**
Acentric factor (w)	**Dimensionless**	**0.0084**	**1.3189**
Volume shift parameter (VSP)	**Dimensionless**	**-0.1678**	**0.3535**
Binary interaction parameter (BIP)	**Dimensionless**	**0**	**0.1303**
Square-well depth (ε_p)	**K**	**228**	**2788**

3.2.3 Data Samples

The samples are generated using the Latin Hypercube Sampling technique [98] to ensure the randomness in a multidimensional distribution. A total number of 123,050 samples are generated for the training and test of the ML model.

3.3 Results and Analysis

In this work, all the ML models are trained on Keras with the CPU of Intel Core i7-1065G7 with an installed memory of 16 GB. The comparison between the traditional and ML-assisted phase-equilibrium calculation models is shown below.

3.3.1 Traditional PR EOS Modeling

The average simulation time for the phase-equilibrium calculation using PR EOS and PR-C EOS is compared, as shown in Figure 18. In the previous phase-equilibrium calculation studies [99], the computation time per phase-equilibrium calculation using cubic EOS is about 10 μs, which is much faster compared with our simulator. However, the aim of this section is not to optimize and accelerate traditional phase-equilibrium calculation model, but to show that the phase-equilibrium calculation using PR-C EOS is much slower compared with phase-equilibrium calculation using PR EOS. The reasons are two-folded: 1) the interaction between the fluid and pore wall is incorporated in PR-C EOS. As shown in Section 3.1.1, more parameters are introduced in PR-C EOS when the effect of fluid-wall interaction is taken into consideration. Compared with PR EOS, these extra computations require more CPU time; 2) the algorithm of solving the chemical equilibrium is more difficult in PR-C EOS modeling. In the phase-equilibrium calculation using PR EOS, the method of tangent plane distance (TPD) is often applied to find the solutions. However, in the phase-equilibrium calculation using PR-C EOS, the TPD method is not effective and a more time-consuming search method is used. Therefore, the application of ML technique provide a more significant speedup in the phase-equilibrium calculation using PR-C EOS.

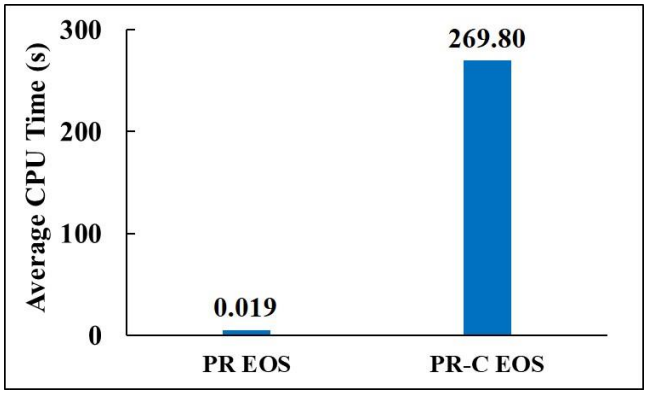


Figure 18: The average CPU times for phase-equilibrium calculation using PR EOS and PR-C EOS, respectively.

3.3.2 Stability Test

The function of the ML model of stability test is to determine the fluid phase state: single-phase or two-phase. MLP is applied in the stability test. The input variables of the MLP are the reservoir conditions and EOS parameters. The output variable is the label of the phase condition, either single-phase or two-phase condition. The structure of the MLP is shown in Figure 19, including an input layer, 5 hidden layers and an output layer. The activation functions of the input layer and hidden layers are ReLU whereas the activation function of the output layer is sigmoid. The loss function is set as binary cross-entropy because the stability test is a binary classification problem. The Adam optimization algorithm [100] is used in the MLP, which is computationally efficient and requires less memory. All input variables are rescaled based on the normalization equation shown below:

$$y_{rescaled} = (y - y_{min})/(y_{max} - y_{min}) \tag{30}$$

where *y* denotes a variable among the 60 input variables.

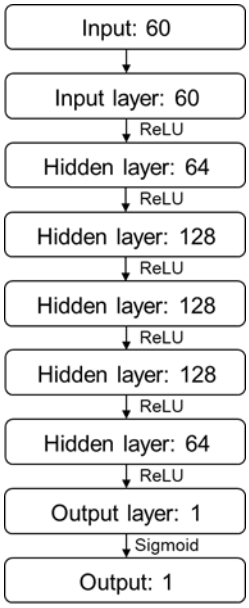


Figure 19: Structure of the MLP of stability test. The number represents the number of nodes in each layer.

A total of 123,050 samples are used for the stability test. The *k*-folds cross validation technique [101, 102] is applied in the training process. In this work, *k* is set as 10, which means the samples are divided into ten parts. Each time one part is selected as the test set and the remaining nine parts are used for the training set. The training process is performed 10 times until each part is selected as the test set. The accuracy obtained using *k*-folds cross validation is more robust and trustworthy compared with the method of splitting data directly because each sample is considered for training and test.

To increase the accuracy of our ML model, several parameters of the MLP are optimized separately. Firstly, the number of hidden layers is optimized to provide a proper structure of the MLP. The results show that the structure with 5 hidden layers has the highest accuracy. Less layers may not fully interpret the complex relationship between the input and output variables while too many layers may lead to the problem of overfitting. Then, different combinations of nodes in each

layer are tested and the results show that this specific set of nodes (64, 128, 128, 128, 64) has the best performance. By balancing the accuracy and CPU time, the optimized batch size and the optimized number of epochs are 64 and 80, respectively. The optimal learning rate of Adam is 0.001. The optimized parameters of MLP are summarized in Table 8.

Table 8: The optimized parameters of MLP in the stability test

Parameter of MLP	Optimal results
Number of hidden layers	**5**
Number of nodes	**64, 128, 128, 128, 64**
Batch size	**64**
Number of epochs	**80**
Learning rate of Adam	**0.001**

Based on the optimized MLP of stability test, the average accuracy of the test set using *k*-folds cross validation technique ($k = 10$) is 96.96%. If the activation function of the output layer is changed to ReLU, the average accuracy is 92.55%, which is lower than the condition using sigmoid. Sigmoid function is generally not used in hidden layers due to the gradient vanish problem. However, in the output layer, the output result is between 0 and 1. If the output result is equal to or higher than 0.5, the label is considered as 1. Otherwise, the label is 0. Therefore, sigmoid is a good choice for a binary classification problem. We randomly select 500 samples from the data set and perform the stability test by applying the algorithm in the traditional phase-equilibrium calculation method and the MLP of stability test. The average CPU times for performing the stability test using the two methods are shown in Figure 20. The results show that

the utilization of ML technique can significantly improve the computational efficiency by 963 times compared to the traditional method.

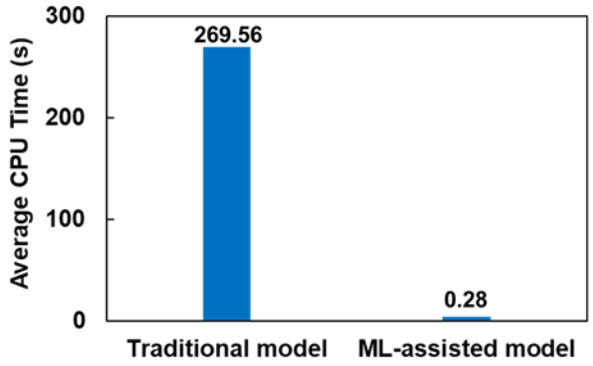


Figure 20: The average CPU time for performing the stability test using the traditional phase-equilibrium calculation method and the MLP.

3.3.3 Phase-Split Computation

The function of the ML model of phase split computation is to obtain the initial guess of equilibrium ratio for the optimization program of total Helmholtz free energy (A_{total}). A good initial guess is shown to largely reduce the number of iterations and, therefore, cut down the CPU time. In the traditional phase-equilibrium calculation, the initial guess of equilibrium ratio is obtained from the stability test. In the generalized ML-assisted phase-equilibrium calculation model, the MLP and PINN models of phase-split computation are trained to obtain the equilibrium ratio as the initial estimate for the Helmholtz free energy minimization program.

The structure of the MLP of the phase-split computation is shown in Figure 21. The input variables are the parameters of the reservoir conditions and EOS, whereas the output variables are the equilibrium ratios of the six pseudo-components. Both the input and output variables are normalized using Equation (30). The optimization algorithm used for this model is the Adam

algorithm. The loss function is the mean square error (MSE), which is calculated based on the equation shown below:

$$MSE = \frac{1}{ns*nc}\sum_{j=1}^{ns}\sum_{i=1}^{nc}(\widehat{K}_{i,j} - K_{i,j})^2 \quad (31)$$

where *ns* represents the number of samples and $\widehat{K}_{i,j}$ is the predicted equilibrium ratio of component *i* in sample *j*.

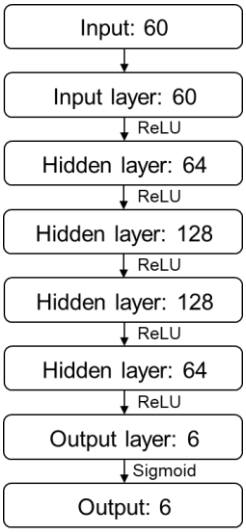


Figure 21: Structure of the MLP of phase-split computation. The numbers represent the number of nodes in each layer.

The total number of samples for phase-split computation is 52,102. The MLP of phase-split computation is optimized following the same method for the optimization of the MLP of the stability test. The optimized results are summarized in Table 9.

Table 9: The optimized parameters of MLP in the phase-split computation

Parameter of MLP	Optimal results

Number of hidden layers	**4**
Number of nodes	**64, 128, 128, 64**
Batch size	**64**
Number of epochs	**40**
Learning rate of Adam	**0.001**

The flow chart of the PINN of phase-split computation is shown in Figure 22. For the neural network, the input variables are the parameters of the reservoir conditions and EOS, whereas the output parameters are the equilibrium ratios. The structure of the neural network in the PINN is the same as the neural network of the MLP for the phase-split computation, as shown in the blue dashed box in Figure 22. For conventional NNs (e.g., MLP), the loss function represents the differences between the real values and predicted values of output parameters. The training of the neural networks is achieved by minimizing the deviations. However, for the PINN, the loss function does not require the real values of the output parameters. In this case, the initial estimate of K_i is used to calculate A_{total} using PR-C EOS. A_{total} is calculated using Equations (9) and (25). It is a function of the 60 input variables, including reservoir conditions and EOS parameters. Instead of directly minimizing the A_{total}, we minimize the loss function shown below:

$$\mathcal{L} = \frac{1}{ns}\sum_{j=1}^{ns}(A_{total,j}/10^7) \tag{32}$$

where $\mathcal{L}$ is the loss function, *ns* represents the number of samples and A_{total} denotes the total Helmholtz free energy. The reason is that A_{total} is at the scale of 10^7 while K_i is in the range of 0 and 1. A slight change of K_i can significantly affect the value of A_{total}. For a node in the neural network, it consists of a weight and a bias. The training process focuses on the derivative of the

loss function with respect to the weight. If A_{total} changes significantly, the gradient will change significantly, which makes the neural network not stable. Therefore, A_{total} is normalized to ensure the neural network is trained smoothly. During the iteration process, the loss function and A_{total} are continuously minimized. The neural network (blue dashed box) is trained to predict more accurate K_i. In the training of neural network, the real values of K_i are not used. The neural network is trained to find the minimal A_{total} based on PR-C EOS and provide accurate prediction of K_i.

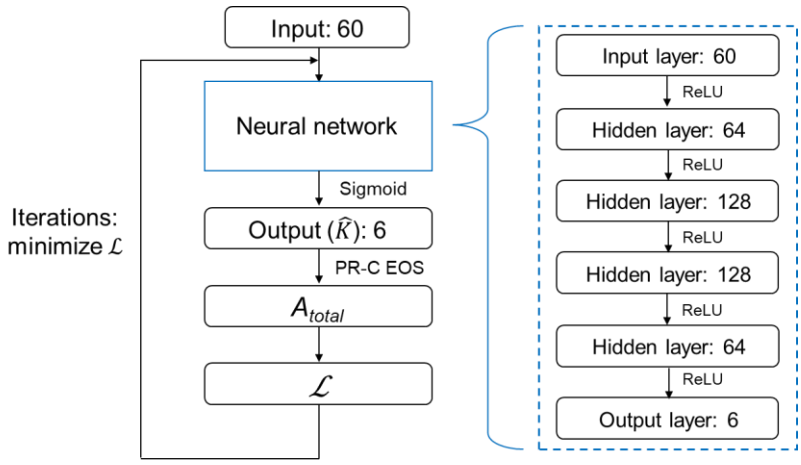


Figure 22: Flow chart of the PINN of phase-split computation. The structure of the neural network is shown in the blue dashed box. The input variables are the reservoir conditions and EOS parameters, whereas the output variables of the neural network are the equilibrium ratios. The total Helmholtz free energy is calculated based on Equations (9) and (25) using the equilibrium ratios predicted by the neural network. During the iterations, the loss function is continuously minimized and therefore, the neural network is trained to give more accurate predictions of the equilibrium ratios.

To evaluate the performance of traditional method, MLP and PINN on the initial estimate of K_i, we select the mean absolute error (MAE) as the metric. The equation of MAE is shown below:

$$MAE = \frac{1}{ns \times nc} \sum_{j=1}^{ns} \sum_{i=1}^{nc} \left| \frac{\hat{K}_{i,j} - K_{i,j}}{K_{i,j}} \right| \tag{33}$$

where *nc* represents the number of pseudo-components and $\hat{K}_{i,j}$ represents the predicted value of the equilibrium ratio of pseudo-component *j* in the *i*th sample. A low value of MAE means a good prediction. The performance of different methods on the initial estimate of K_i is summarized in Figure 23. The MAE of K_i obtained from the stability analysis in the traditional phase-equilibrium calculation model (Method a) is relatively high. The MAEs of K_i obtained from MLP trained with a large dataset (Method c) and PINN (Method d) are reduced by 77.26% and 82.12% compared with Method a, respectively. However, the performance of MLP trained with a small dataset (Method b) becomes worse compared with Method a. The comparison between Methods b and c indicate that the performance of MLP is largely affected by the number of samples. A large dataset is required to ensure the prediction accuracy of MLP. For the PINN, it shows good performance even when it is fed with a small dataset. The reason is that the PR-C EOS is incorporated into the training of PINN. Instead of working as a black-box model, the neural network learns to minimize the A_{total} and provides more reasonable predictions of K_i. PINN becomes more effective than conventional neural networks when the samples are hard to generate or collect.

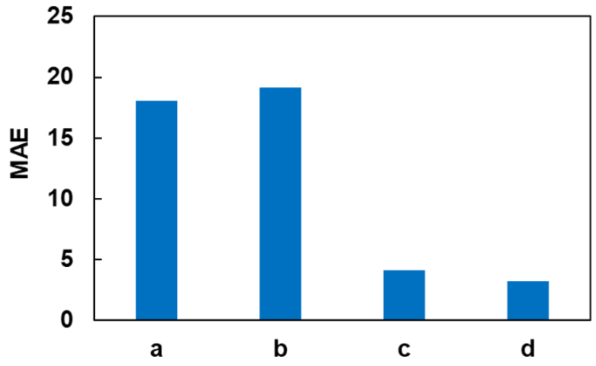


Figure 23: The mean absolute error (MAE) of the initial estimates of equilibrium ratios obtained from different methods: a) the results of stability analysis in the traditional VLE calculation models, b) MLP trained with 1,000 samples, c) MLP trained with 52,102 samples, d) PINN fed with 1,000 samples.

3.3.4 Phase-Equilibrium Calculation

The stability test and phase-split calculation are combined as the phase-equilibrium calculation. The algorithms of traditional and ML-assisted phase-equilibrium calculation models are shown in Figure 24. In the traditional phase-equilibrium calculation model, the Newton-Raphson iteration is used to search any possible phase II that has chemical equilibrium with phase I. The results obtained from the stability test are directly used as the initial estimate of K_i. In the ML-assisted phase-equilibrium calculation model, the MLP is used to determine the phase-state condition, which significantly reduces the computational time. The initial estimates of K_i are obtained from PINN. More accurate estimates of K_i can accelerate the minimization of A_{total}. The average CPU time of VLE calculation using the traditional and ML-assisted phase-equilibrium calculation models are shown in Figure 25. The ML-assisted phase-equilibrium calculation model reduces the computational time by two orders of magnitude compared with the traditional phase-equilibrium calculation model.

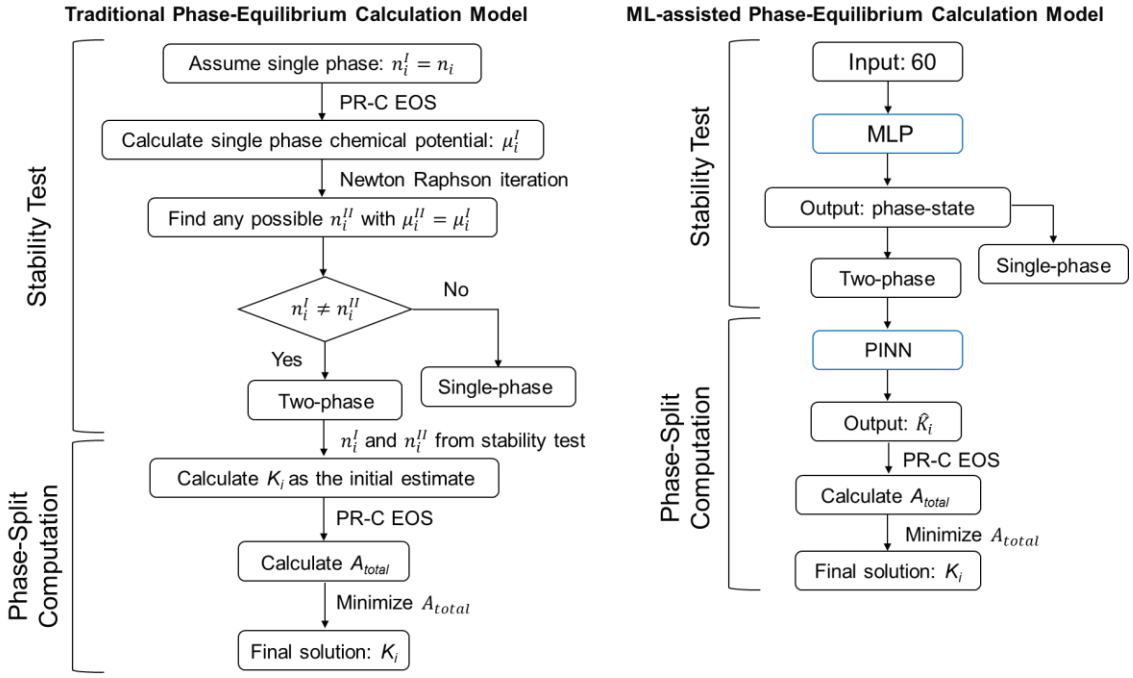


Figure 24: The algorithms of traditional and ML-assisted phase-equilibrium calculation models.

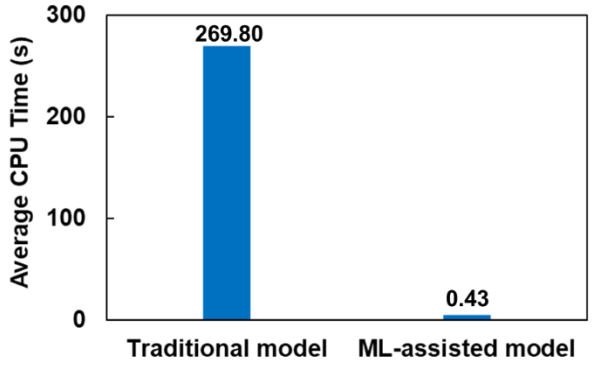


Figure 25: The average CPU time of phase-equilibrium calculation using traditional and ML-assisted phase-equilibrium calculation models.

4. ANALYSIS OF HYDROGEN-HYDROCARBON MIXTURE PHASE BEHAVIOR USING MOLECULAR SIMULATION[*†]

Unlike traditional fossil fuels, hydrogen (H_2) is a low-energy-density fuel; therefore, an extremely large storage volume is required for H_2 to meet energy demand [13, 14]. To solve this problem, underground H_2 storage has been considered due to its large storage capacity, safety and low cost. Conventional geologic H_2 storage sites include depleted reservoirs, salt caverns and saline aquifers. With the development of shale oil and gas, a large number of shale gas reservoirs have been depleted; such reservoirs may be attractive candidates for H_2 storage. In this section, we present an investigation of the potential of H_2 storage in depleted shale gas reservoirs.

4.1 Simulation Model

The simulation methods and molecular settings of different Monte Carlo simulation models are described in this section.

4.1.1 Single Bulk Conditions

As discussed in Section 1.2, the GEMC method is designed to analyze fluid density and phase behavior. There are two types of GEMC simulations, the global volume-imposed GEMC (NVT-GEMC) and the pressure-imposed GEMC (NPT-GEMC). Because pressure, temperature and fluid composition are provided in our cases, the NPT-GEMC simulation is applied in this

[*] Part of this chapter is reprinted with permission from "Molecular Simulation of Hydrogen-Shale Gas System Phase Behavior under Multiscale Conditions: A Molecular-Level Analysis of Hydrogen Storage in Shale Gas Reservoirs" by Fangxuan Chen, Mohamed Mehana and Hadi Nasrabadi, 2023. *Energy & Fuels*, Volume, Pages, Copyright [2023] by American Chemical Society.

[†] Part of this chapter is reprinted with permission from "Molecular Simulation of Competitive Adsorption of Hydrogen and Methane: Analysis of Hydrogen Storage Feasibility in Depleted Shale Gas Reservoirs" by Fangxuan Chen, Shihao Wang and Hadi Nasrabadi, 2023. SPE Proceedings, SPE-212218-MS, Copyright [2023] by Society of Petroleum Engineers.

work. The schematic figure of NPT-GEMC simulation at bulk condition is shown in Figure 26. The two boxes (Box I and II) are designed for the potential two phases: one for the liquid phase and the other one for the gas phase. The advantage is that the separation of two phases into two simulation boxes avoids the consideration of explicit interfaces between the two phases. Periodic boundary conditions (PBCs) are applied at all three directions to simulate the bulk condition. During the simulation process, the total number of fluid molecules (N), system pressure (P) and system temperature (T) are kept constant. Three types of Monte Carlo moves are performed in the simulation process: 1) particle displacement (in the current box), 2) particle swap (between two boxes) and 3) volume change.

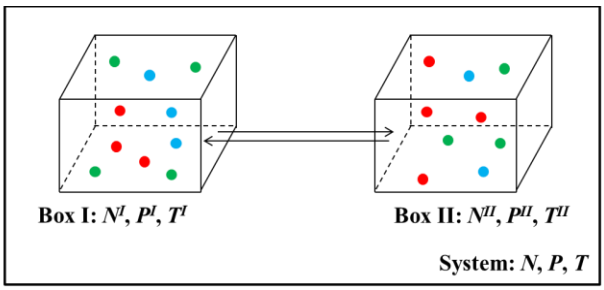


Figure 26: Schematic figure of NPT-GEMC simulation at bulk condition

The fluid phase behavior is determined by the fluid densities in the two boxes. The fluid is in single gas phase if the fluid densities in the two boxes are both gas-phase densities. When the fluid density in one box is gas-phase density while the other one is liquid-phase density, the fluid is in two-phase condition. In this work, the reference gas-phase density or liquid-phase density at specific pressure and temperature is estimated by PR EOS [10].

4.1.2 Single Nanoscale Conditions

As discussed in Section 1.2, the GCMC simulation was designed to investigate fluid adsorption characteristics. To analyze the fluid adsorption behavior in nanopores, we applied GCMC simulations under single nanoscale conditions. The simulation models are shown in Figure 27. Slit pore was selected because it is the commonly observed pore shape in shale reservoirs [3, 103, 104]. We consider two types of boundary materials, namely graphite as organic material and quartz as inorganic material. We later found that the adsorption behavior simulated in the graphite layer was similar to the adsorption behavior simulated in the kerogen boundary. For the inorganic boundary, we selected quartz as a representative material because quartz accounts for the highest portion (40%) of inorganic compositions in North American shale rocks [105, 106]. PBCs were applied at the extended directions of the boundary (x and y directions in Figure 27). During the simulation process, the chemical potential of each component (μ), system volume (V) and system temperature (T) are kept constant. Two types of Monte Carlo moves were performed in the simulation process: 1) particle displacement (in the current box) and 2) particle insertion or deletion.

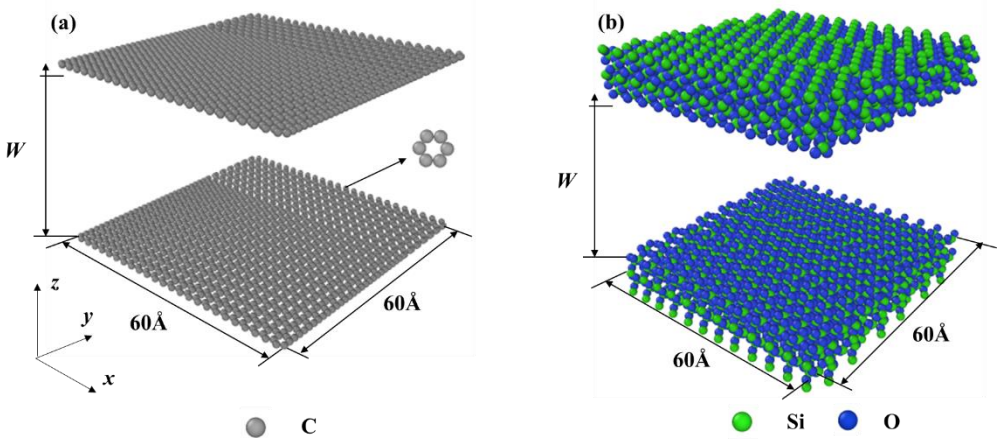


Figure 27: Slit pores in the simulation: a) graphite, b) quartz. For the graphite layer, the unit is a regular hexagon shape of carbon atoms, which is magnified for clarification. The crystal

structure of quartz is α-quartz. The grey, green and blue balls represent the atoms of carbon, silicon, and oxygen, respectively.

The workflow of the simulation process is shown in Figure 28. Generally, the given conditions are temperature, pressure and fluid composition. The PR EOS is applied to first obtain the fluid density. For a specific volume, we can calculate the number of molecules of each component. Then, the chemical potential of each component is obtained using Widom's particle insertion method [107] using the GEMC simulation, which has been used in previous studies [108, 109], under bulk conditions. Based on the chemical equilibrium between bulk and confined fluid, we can obtain the chemical potential of each component under nanoscale conditions. After obtaining this potential, a GCMC simulation is performed under nanoscale condition to obtain the distribution of fluid molecules. Based on the fluid distribution, we can quantitatively analyze adsorption behaviors.

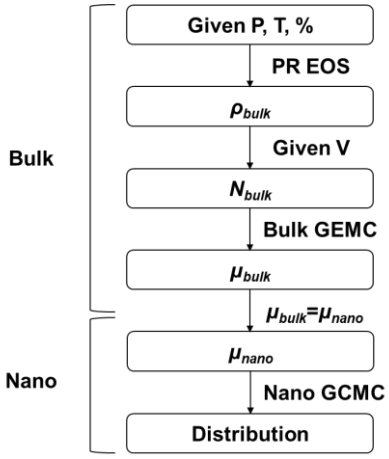


Figure 28: Workflow of the simulation process.

4.1.3 Multiscale Conditions

To simulate the multi-scale condition, we designed the modified NPT-GEMC method [110], as shown in Figure 29. Different from the traditional NPT-GEMC method discussed above, an extra gauge box is added to the modified NPT-GEMC method to simulate the confined region [111]. In Figure 29, the two cubic boxes (Box I and Box II) are used to simulate the gas and liquid phases of fluid in bulk condition, which is similar to the traditional NPT-GEMC method. The three types of MC moves are executed (particle displacement, volume change, and particle swap) in Box I and Box II to exchange particles and tune volume according to current pressure. A gauge cell (Box III) with fixed volume is utilized to simulate the confined region. Since the volume of Box III is fixed, only two types of MC moves (particle displacement and particle swap) are executed. It should be noted that the pressure (P^{III}) in Box III is not the same as the system pressure (P) because molecular simulation uses the Virial pressure [112], which is affected by the distribution of particles. Due to the confinement effect, particles in Box III tend to occupy the near-boundary region. The heterogeneous distribution largely affects the value of Virial pressure in Box III and therefore P^{III} is higher than P [113]. For a multi-scale system with M components, the constraints are shown in the following equations:

$$N_i^I + N_i^{II} + N_i^{III} = N_i \quad i = 1,2,\dots,M \tag{34}$$

$$P^I = P^{II} = P \tag{35}$$

$$T^I = T^{II} = T^{III} = T \tag{36}$$

$$N_i, P, T, V^{III} = constant \quad i = 1,2,\dots,M \tag{37}$$

Since Box I and II are designed as the bulk condition, PBCs are applied to all three dimensions. For Box III as the confined region, the PBC is applied only in the extended direction.

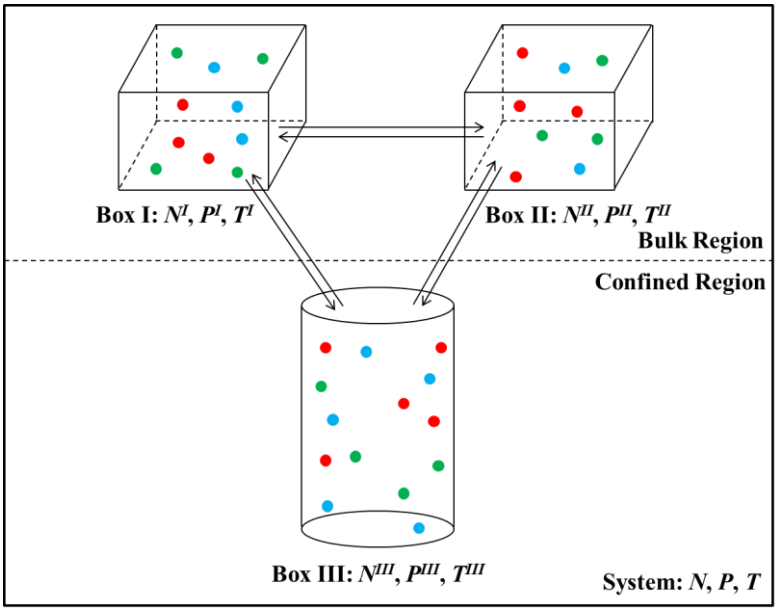


Figure 29: The schematic of modified NPT-GEMC method. Box I and II represent the bulk region while Box III simulates the confined region. The pressures in Box I and II are the same as system pressure and the temperatures of all three boxes are equal to system temperature. The volume of Box III is fixed.

4.1.4 Molecular Settings

In the simulation process, particle displacement and particle swap are conducted by randomly selecting a molecule in the system, inserting the molecule at a random position, and deleting the initial molecule. This method is effective for transferring a small molecule in a sparse environment. However, when it comes to inserting a long-chain molecule in a densely-packed simulation box, the acceptance rate would be much lower due to the overlap with other molecules, which prolongs the simulation time to reach equilibrium. To solve this problem, the configurational-bias Monte Carlo (CBMC) insertion move is applied. Instead of inserting a whole molecule at one time, CBMC inserts one atom in the molecule at a position randomly and tries to insert another atom in this molecule by evaluating the possible positions around the previous atom

[114]. CBMC insertion move significantly improves the acceptance rate of long-chain molecule insertion and therefore reduces the running time. In our simulations, 10 trial positions are set for each growing atom in the MC moves of particle displacement and particle swap.

The nonbonded interaction energy ($U(r_{ij})$) between particle i and j is calculated by the Lennard Jones (LJ) 12-6 potential, as shown below:

$$U(r_{ij}) = 4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] \quad (38)$$

where ε_{ij} represents the potential well depth, r_{ij} refers to the distance between particle i and j and σ_{ij} is the separation distance between particle i and j where the LJ potential is zero. The Lorentz-Berthelot mixing rule [115] is applied to calculate the potential parameters of diverse particles:

$$\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2} \quad (39)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \cdot \varepsilon_j} \quad (40)$$

The cutoff distance is set as 10 Å and the long-range tail correction is incorporated in the simulation.

The transferable-potentials-for-phase-equilibria united-atom (TraPPE-UA) force field [116] was used for hydrocarbons, graphite boundary and helium (He). The transferable-potentials-for-phase-equilibria explicit-hydrogen (TraPPE-EH) force field [117] was applied for N_2. In the TraPPE-EH force field, the N_2 molecule was modeled as a three-site molecule with two nitrogen atoms bonded to the center-of-mass (COM). The parameters of quartz (SiO_2) were obtained from the ClayFF force field [118]. The TIP4P model [119] was selected to simulate water (H_2O) molecules due to its effective performance in molecular simulations [120, 121]. The symbol *M* shown in Table 10 represents the negative charge site of H_2O molecule. The H_2 molecule was considered as a single site molecule, and the potential parameters were obtained from the reference

[122], which exhibited good agreement with experimental isotherms. The parameter settings are summarized in Table 10.

Table 10: The potential parameters used in the simulations

	ε/k_B **(K)**	σ **(Å)**
C (graphite)	**30**	**3.70**
He	**4**	**3.11**
CH_4	**148**	**3.73**
CH_3 (hydrocarbon)	**98**	**3.75**
CH_2 (hydrocarbon)	**46**	**3.95**
N (N_2)	**36**	**3.31**
COM (N_2)	**0**	**0**
Si (SiO_2)	**0**	**3.30**
O (SiO_2)	**78**	**3.17**
H (H_2O)	**0**	**0**
O (H_2O)	**78**	**3.15**
M (H_2O)	**0**	**0**
H_2	**37**	**2.96**

The structures of different boundaries are shown in Figures 27 and 30. As shown in Figure 27, the graphite layer consists of carbon atoms, and the unit of graphite layer is in the shape of a regular hexagon. The unit of quartz boundary is α-quartz; the crystal structure of α-quartz is

obtained from American Mineralogist Crystal Structure Database (AMCSD) [123]. As shown in Figure 30, a buckytube is used as the boundary of Box III in the modified NPT-GEMC simulation. The buckytube [124] is a single-carbon-atom layer in a cylindrical shape, and the unit of the buckytube is in the shape of a regular hexagon.

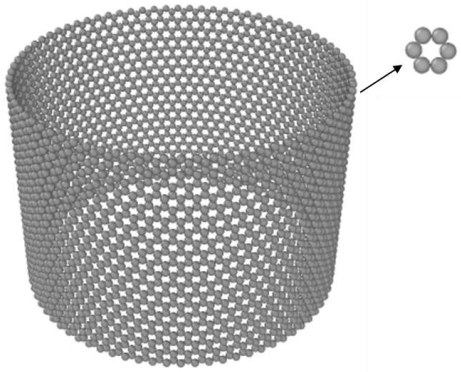


Figure 30: Structure of buckytube. The buckytube consists of carbon atoms, with an inner diameter of 6 nm. The unit of buckytube is in the shape of regular hexagon.

The Monte Carlo for Complex Chemical Systems (MCCCS) Towhee [125] is modified to conduct all the simulations in this paper.

4.2 Model Validation

To demonstrate the accuracy of our simulation model and molecular settings, we performed a series of simulations on fluid phase behavior and adsorption characteristics. The simulation results were compared with the experimental results and other simulation outputs.

4.2.1 Fluid Phase Behavior

To ensure the accuracy of potential parameters of H_2, we simulate the density characteristics and phase behaviors of pure H_2 and hydrogen-hydrocarbon (H_2-HC) systems at bulk condition.

The density profiles of pure H_2 at bulk condition is shown in Figure 31. The blue line represents the experimental results [126] while the red dots are the simulation results. Since H_2 is always in single gas phase, only one simulation box is used to measure the fluid density. A series of simulations with isothermal-isobaric ensemble (constant NPT ensemble) are performed. We can observe that the simulation results are consistent with the experimental results.

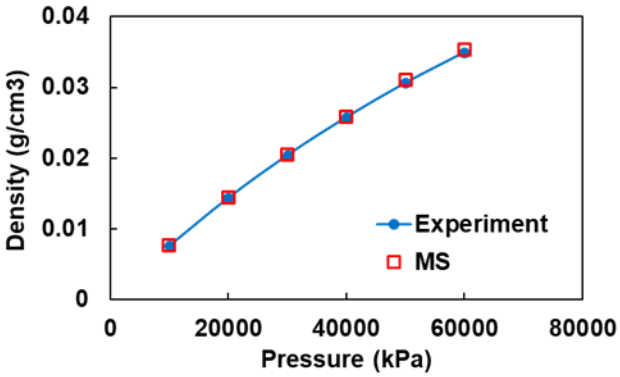
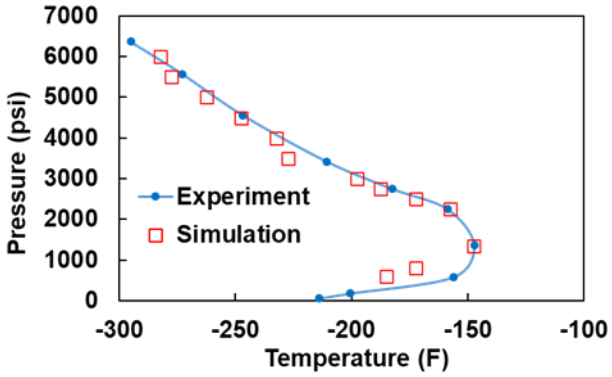


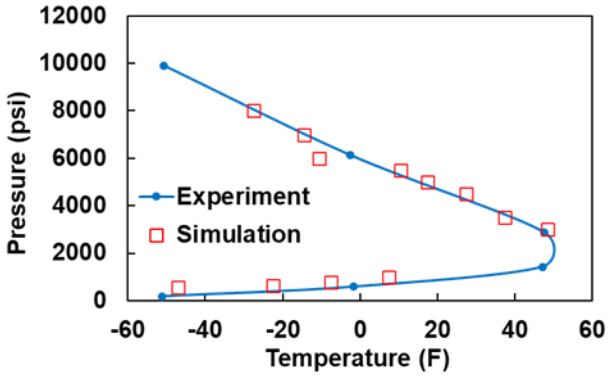
Figure 31: Density profile of pure H_2 at 300 K under bulk condition. The blue line represents the experimental results while the red dots are the simulation results.

To the best of the authors' knowledge, no experimental results of H_2-SG system phase behavior are found. However, there are several experimental works focusing on the phase behaviors of H_2-HC binary mixtures, including H_2-C_1 [127], H_2-C_2 [128], H_2-C_3 [128, 129], H_2-C_4 [130] and H_2-C_6 [131] binary mixtures. Since shale gas consists of more than 90% of light hydrocarbons (C_1~C_3) [132], we validate the accuracy of potential parameters and simulation settings using H_2-light hydrocarbon binary mixtures (H_2-C_1, H_2-C_2 and H_2-C_3). The phase

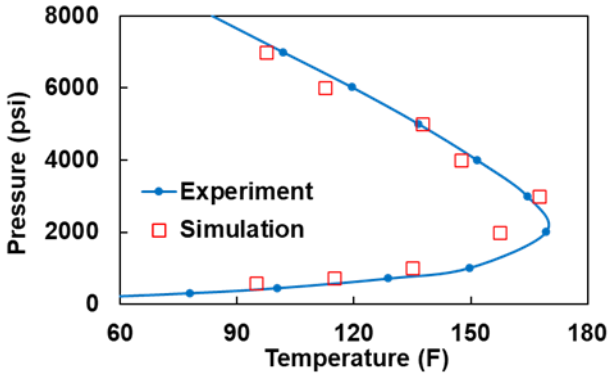
behaviors of H_2-C_1, H_2-C_2 and H_2-C_3 binary mixtures are shown in Figure 32. The experimental data of H_2-C_1 binary mixture is obtained from the previous study [127] while the experimental results of H_2-C_2 and H_2-C_3 binary mixtures are obtained from the previous work [128]. The blue lines represent the experimental results while the red dots are the simulation results. The simulation follows the method discussed in Section 4.1.1. For a specific simulation point (red dot), many parallel simulations with NPT ensembles are performed at the same pressure but different temperatures. With the variation of temperatures, there exists a temperature that the fluid changes from single-phase to two-phase condition, which is the temperature of phase transition we seek. Combined with the fixed pressure, we can obtain the pressure and temperature of the phase transition point, which is the coordinate of red dots shown in the figure. The reason we change temperature is that the phase transition is sensitive to temperature, which means that a small change of temperature can have significant effect on phase behavior. In Figure 32, we can observe that the phase transition points of three binary mixtures by simulation are in agreement with corresponding experimental results, indicating that our potential parameters and simulation settings are reasonable to perform further simulations.



(a) C_1 60% and H_2 40%



(b) C_2 50% and H_2 50%



(c) C_3 50% and H_2 50%

Figure 32: Phase behaviors of H_2-HC binary mixtures under bulk conditions: a) C_1 60% and H_2 40%, b) C_2 50% and H_2 50%, and c) C_3 50% and H_2 50%. The blue line represents the experimental results while the red dots are the simulation results.

4.2.2 Adsorption Characteristics

As shown in Figure 27, the pore size is defined as the distance (W) between the mass centers of the boundary atoms. The surface area (S_A) is the area of the upper or lower boundary surface. However, due to the shape and size of molecules, the fluid molecules cannot overlap with boundary atoms. Therefore, the effective pore volume cannot be simply calculated based on the simulation box shape ($W \times S_A$). Zhang et al. [133] measured the effective pore volume using He because He

does not exhibit strong adsorption behavior under confined conditions [108], and their simulation results showed that the effective pore volume was the same at different pressures. Similarly, we followed their methods and measured the effective pore volume of our slit pore model ($60\text{Å} \times 60\text{Å} \times 40\text{Å}$) at 338.15 K. The equation used to calculate the effective pore volume (V_p) is as follows:

$$V_p = \frac{N_{He}}{N_A \cdot \rho_{He,b}} \tag{41}$$

where N_{He} is the number of He molecules in the slit pore, N_A is the Avogadro constant, and $\rho_{He,b}$ is the molar density of He under bulk condition, which is obtained from the National Institute of Standards and Technology (NIST) Chemistry Webbook [134].

The simulation process is similar to the workflow shown in Figure 28. Based on the He bulk density obtained from the NIST, we can determine the number of He molecules and perform the GEMC simulation to obtain the chemical potential of He under bulk conditions. According to the chemical potential equilibrium, the chemical potential of He in confined conditions is obtained and GCMC simulations are performed to determine the number of He molecules in the slit pore. We performed a series of GCMC simulations of He ranging from 2 MPa to 16 MPa, and the number of He molecules at each pressure was obtained. Based on Equation (41), we could obtain the effective pore volume at each pressure. The average effective pore volume was 133,643 Å^3, which was less than the slit pore volume ($W \times S_A = 144{,}000\text{Å}^3$). These results are consistent with Tian et al.'s work [108]. The deviations between the effective pore volumes at different pressures and average effective pore volume were less than 3%, indicating that the effective pore volume had not been affected by pressure. The excess adsorption capacity of C_1 is calculated based on the following equation:

$$m_{ex} = \frac{N_{C1}/N_A - V_p \cdot \rho_{C1,b}}{2S_A} \quad (42)$$

where N_{C1} is the number of C_1 molecules in the slit pore, $\rho_{C1,b}$ is the molar density of C_1 under bulk conditions, and m_{ex} is the excess adsorption capacity (mmol/m^2).

Generally, the adsorption data from experiments is expressed in unit per unit mass of the adsorbent (e.g., mmol/g). Chen et al. [135] have suggested that it would be easier to compare experimental data and simulation results by converting the adsorption data to unit per unit surface area of the adsorbent (e.g., mmol/m^2). Based on previous measurements [136], the specific surface area of kerogen is approximately 250 m^2/g. The experimental data [2] and simulation results [1] for C_1 adsorption are converted to the unit of mmol/m^2. The excess adsorption capacities of C_1 between 2 and 16 MPa at 338 K are shown in Figure 33. Our simulation results are consistent with the experimental data, indicating that our simulation model and molecular settings are reasonable. In addition, our simulation results are similar to the simulation results of Wang et al. [1], who used kerogen as a boundary material, which implies that graphite layers can be used to simulate adsorption behaviors on kerogen without a large loss of accuracy.

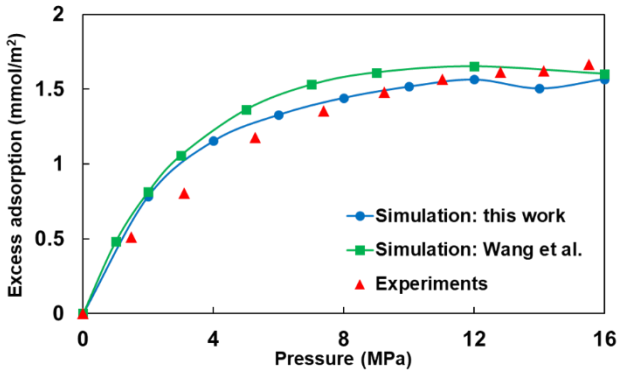


Figure 33: Excess adsorption capacity of C_1 at 338 K. The blue line represents the simulation results of this work while the green line represents the simulation results of Wang et al. [1]. The red dots are the experimental results [2].

4.3 Results and Analysis

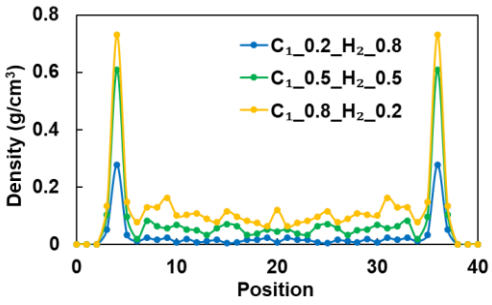
Investigations of the phase behavior, compositional heterogeneity and adsorption behaviors of H_2-HC systems at different conditions are presented in this section. The simulation results can be used to analyze the potential for H_2 storage in shale reservoirs.

4.3.1 H_2-Hydrocarbon Under Nanoscale Conditions

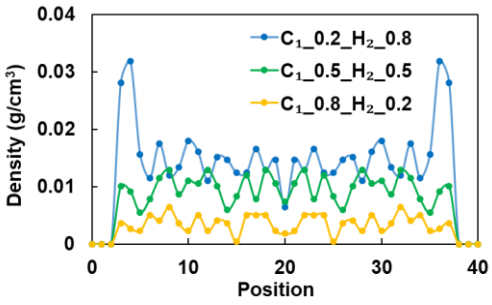
A series of MC simulations were performed to investigate the effect of pore size, temperature, pressure, boundary surface, and fluid compositions on C_1-H_2 competitive adsorption. In this work, we consider the pore sizes of 2, 4, and 10 nm. Jin and Firoozabadi [109] have shown that when the slit pore size is large than 4 nm, the fluid molecules in the pore center behave in a bulk-like way. Therefore, the pore size of 10 nm is representative for large mesopores. The temperature and pressure ranges are selected based on shale reservoirs conditions. The average depth of shale gas reservoirs in the US is in the range of 1,500 to 3,500 m [137]. Assuming the temperature gradient of 0.03°C/m [138] and a pressure gradient of 0.012 MPa/m [139], the temperature range of shale gas reservoirs is 340 to 400 K and the pressure range is 18 to 42 MPa. Here, we consider simulation temperatures of 340, 370, and 400 K and the simulation pressures of 10, 30, and 50 MPa. Graphite and quartz are used as the organic and inorganic boundary surfaces, respectively. For the fluid composition, we consider the mole fractions of C_1 in a C_1-H_2 system as 20%, 50%, and 80%. In addition, we analyze the effect of H_2O on C_1-H_2 competitive adsorption.

Experimental results [140] show that the water saturation is in the range of 17 to 43%. In this work, we set the mole fraction of H_2O in the fluid system as 0%, 15% and 30%. The base case is the equimolar C_1 and H_2 system at 370 K and 30 MPa in the 4 nm graphite slit pore.

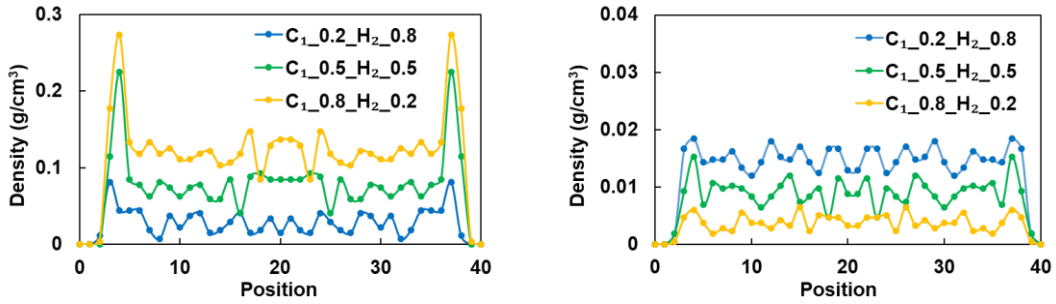
The density profiles of C_1-H_2 mixtures with different compositions are shown in Figure 34. Position 20 represents the center of the slit pore, and positions 0 and 40 represent the lower and upper boundaries. When the mole fraction of C_1 is relatively high (≥ 50%), C_1 is adsorbed near the boundary while H_2 is freely distributed in the slit pore regardless of the boundary material. The reason for this effect is that C_1 is preferentially adsorbed near the boundary compared with H_2. Due to the limited space of the adsorbed region, C_1 molecules first occupy the region and force H_2 molecules away from the boundary. The preferential adsorption will be later quantitatively analyzed. When the mole fraction of C_1 is 20%, in the graphite layer, H_2 forms an adsorption layer because fewer C_1 molecules occupy the adsorbed region, leaving space for H_2 molecules. A similar phenomenon is not observed in the quartz boundary because the organic boundary has higher adsorption effects than the inorganic boundary [141].



(a) C_1 in graphite layer



(b) H_2 in graphite layer

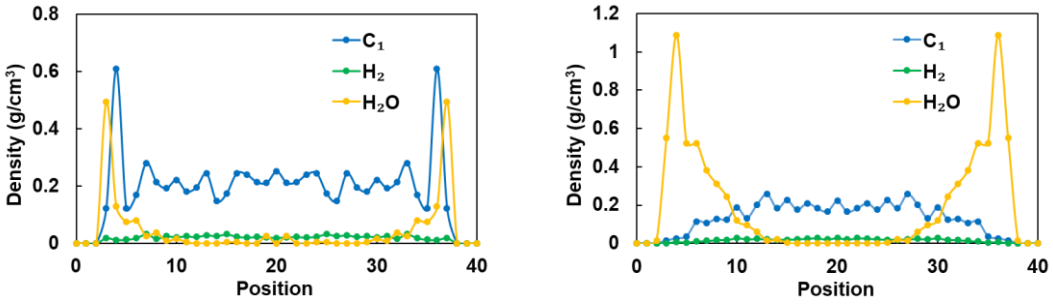


(c) C_1 in quartz **(d) H_2 in quartz**

Figure 34: Density profiles of C_1-H_2 mixtures with different compositions at 30 MPa and 370 K under 4 nm slit pore: a) C_1 in graphite layer, b) H_2 in graphite layer, c) C_1 in quartz, d) H_2 in quartz. The blue, green and orange lines represent the fluid densities of the fluid compositions of 20%, 50% and 80% C_1, respectively. The position 20 represents the center of the slit pore while the positions 0 and 40 are the lower and upper boundaries.

The density profiles of equimolar C_1-H_2 mixtures at different pressures, temperatures and pore sizes have similar trends to those of the equimolar C_1-H_2 mixture at 30 MPa and 370 K under 4 nm slit pore. Therefore, the density profiles have not been shown to avoid redundancy. The density profiles of C_1-H_2-H_2O mixtures in the graphite layer and the quartz boundary are shown in Figure 35. For the graphite layer, both C_1 and H_2O form adsorption layers near the boundary while H_2 molecules are freely distributed in the slit pore. In addition, the H_2O adsorption layer is closer to the boundary compared with C_1 because the H_2O molecules have stronger interactions with boundary atoms than C_1 [140]. As a polar molecule, H_2O is partially charged. The interaction between H_2O molecules and boundary atoms consists of the van der Waals force and electrostatic force. However, due to the electric neutrality of C_1 molecules, the interactions between C_1 molecules and boundary atoms are only based on van der Waals forces. The stronger interactions between H_2O molecules and boundary atoms allow H_2O to accumulate near the boundary surface

and force C_1 molecules away from the boundary [142]. For the quartz boundary, H_2O forms an adsorption layer near the boundary while C_1 molecules tend to accumulate in the center of the slit pore. The reason for this effect is that quartz is hydrophilic and has higher affinity for H_2O molecules [109]. H_2O is more competitive than C_1 and therefore, occupies the adsorbed region [143, 144]. Experimental studies [2, 105] have shown that the existence of H_2O weakens the adsorption capacity of C_1 by 20–60% and 60–95% in organic-rich and inorganic-rich shales, respectively. Our simulation results show the same trend. The hydrophilicity of quartz leads to more significant affinity of H_2O and has stronger effects on C_1 adsorption capacity than the graphite layer.



(a) Graphite layer **(b) quartz boundary**

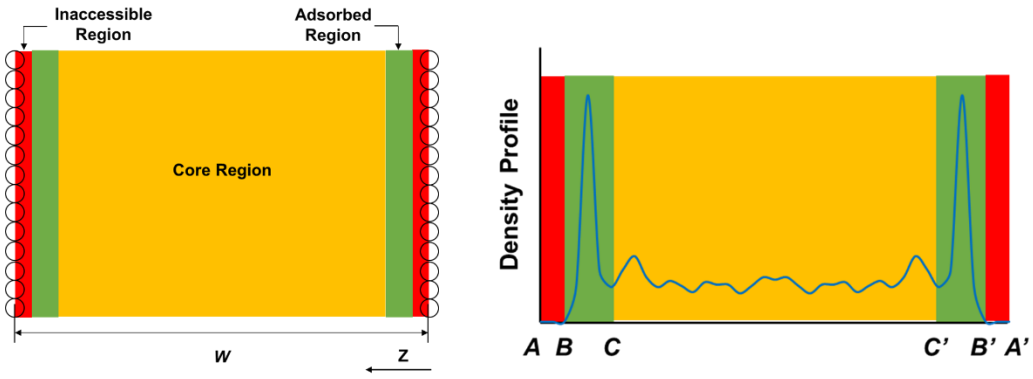
Figure 35: Density profiles of C_1-H_2-H_2O mixtures (bulk composition: C_1 35%, H_2 35% H_2O 30%) at 30 MPa and 370 K under 4 nm slit pore: a) graphite layer, b) quartz boundary. The blue, green and orange lines represent the fluid densities of C_1, H_2 and H_2O, respectively. The position 20 represents the center of the slit pore while the positions 0 and 40 are the lower and upper boundaries.

The adsorption characteristics of C_1, H_2 and H_2O are also analyzed. A schematic figure of slit pores is shown in Figure 36, following the definitions from previous work [108, 141]. The slit

pore is divided into three regions: the inaccessible region, the adsorbed region and the core region. The inaccessible region refers to the region that does not have molecules due to the shape and size of boundary and fluid molecules. It is defined as the region between points A (A') and B (B') based on the density profile. The adsorbed region refers to the region that has an adsorption layer of fluid molecules. In this work, we consider the adsorption behavior of C_1 as monolayer adsorption, which has been shown in previous studies [145-147]. The width (l_{BC} or $l_{B'C'}$) of the adsorbed region is close to the effective diameter (σ) of the adsorbate molecule [148, 149]. The adsorbed phase density is calculated as the average fluid density in the adsorbed regions. Based on the symmetric characteristic of adsorption, the relevant equation is as follows

$$\rho_a = \int_B^C \rho(z)dz/l_{BC} \tag{43}$$

where ρ_a is the adsorbed phase density and l_{BC} is the width of the adsorbed region. The remaining region in the slit pore is the core region. The fluid density in this region is similar to the bulk fluid density under the same conditions [108].



(a) Slit pore **(b) Density profile**

Figure 36: Schematic figures of (a) slit pores, (b) density profile. The red, green and yellow regions are the inaccessible region, adsorbed region and core region, respectively. Points A

and A' are the positions of the mass centers of boundary atoms in the *z* direction. Points B and B' are the positions to separate the inaccessible region and adsorbed region. Points C and C' are the positions to separate the adsorbed region and core region.

The adsorbed phase densities of C_1 at different pore sizes, temperatures, pressures, boundary surfaces, and fluid compositions are shown in Figure 37. Because H_2 molecules are freely distributed in the slit pore, adsorbed phase densities of H_2 are not analyzed. The adsorbed phase densities of C_1 at different pore sizes (Figure 37(a)) are similar because the change of pore size only affects the volume of the core region. The adsorbed region holds similar number of molecules when other conditions are the same. The adsorbed phase density increases with increasing pressure and mole fraction but with decreasing temperature. The same trend has been observed in previous studies [109, 141]. For constant temperature and volume, an increase in pressure leads to an increase in fluid molecules in the simulation box. Due to the confinement, fluid molecules first occupy the adsorbed region [36], which causes an increase in adsorbed phase density (Figure 37(b)). Similarly, an increase of the mole fraction of C_1 adds C_1 molecules to the system, which leads to an increase in the adsorbed phase density of C_1 (Figure 37(c)). An increase in temperature has a negative effect on adsorption (Figure 37(d)). The reason for this effect is that the adsorption process is exothermic and the increase in temperature suppresses the adsorption process [150]. In addition, the increased temperature provides molecules with more energy to surmount the layer barrier and escape from the adsorbed region [151]. Therefore, an increase in temperature leads to a decrease in the adsorbed phase density of C_1. The boundary surface does not change the trend in adsorbed phase density. The higher adsorbed phase densities of C_1 in the graphite boundary are due to the strong interactions between C_1 and graphite layer [141].

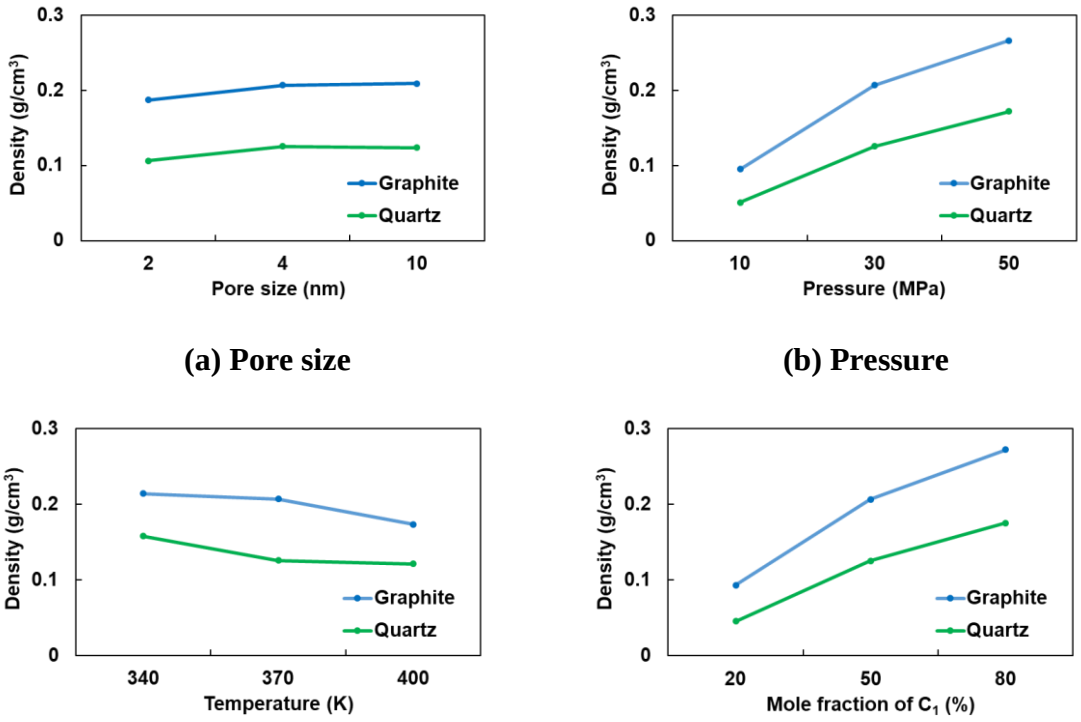


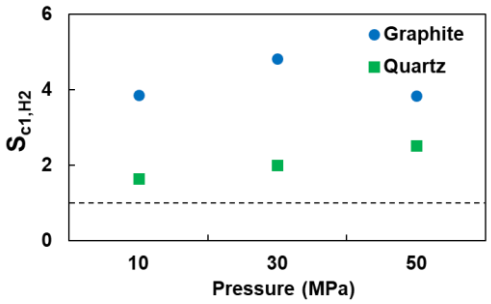
Figure 37: Adsorbed phase densities of C_1 at different conditions in graphite layer and quartz boundary: a) pore size, b) pressure, c) temperature, and d) fluid composition. The blue and green lines represent the adsorbed phase densities of C_1 in graphite layer and quartz boundary, respectively.

As shown in Figures 35 and 36, C_1 and H_2O molecules accumulate in the adsorbed region while H_2 molecules are freely distributed in the slit pore. The preferential adsorption is caused by the selectivity effect, which means that the surface adsorption capacities of different components are not the same. To quantitatively evaluate the boundary selectivity effect, the concept of relative selectivity has been introduced [152]. Relative selectivity refers to the ratio of the adsorbed fractions of the two components, which can be calculated by the following equation:

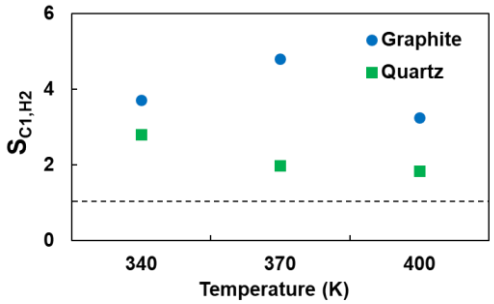
$$S_{i,j} = \frac{x_i/x_j}{y_i/y_j} \tag{44}$$

where x_i and x_j denote the mole fractions of components i and j in the adsorbed region while y_i and y_j represent the mole fractions of components i and j in the core region. Specifically, x_i/y_j is termed as the absolute selectivity of component i [152]. If $S_{i,j} > 1$, component i is preferentially adsorbed by the boundary compared with component j. A high relative selectivity refers to strong adsorption preference.

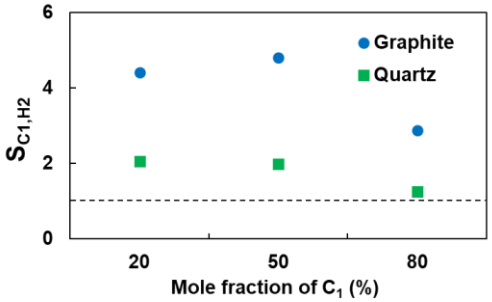
Many studies [1, 153] have applied relative selectivity to investigate the competitive adsorption of C_1 and CO_2. Their results have shown that $S_{CO_2,C_1} > 1$ and CO_2 is preferentially adsorbed near the boundary compared with C_1. The preferential adsorption of CO_2 contributes to the CO_2 sequestration in depleted gas reservoirs. The results of selectivity effect in C_1-H_2 and C_1-H_2-H_2O mixtures are shown in Figures 38 and 39.



(a) Pressure



(b) Temperature



(c) Mole fraction of C_1

Figure 38: Relative selectivity of C_1-H_2 mixture in graphite layer and quartz boundary: a) pressure, b) temperature, c) mole fraction of C_1. The blue and green dots represent the relative selectivity of C_1 and H_2 in graphite layer and quartz boundary. The dashed line represents the relative selectivity of 1.

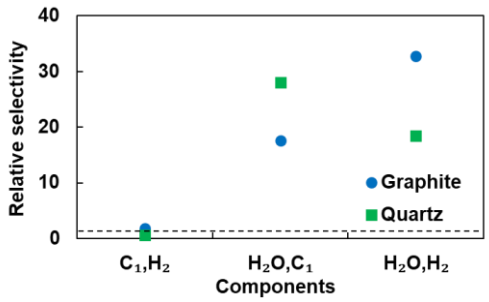


Figure 39: Relative selectivity of C_1-H_2-H_2O mixture (bulk composition: C_1 35%, H_2 35% H_2O 30%) at 30 MPa and 370 K under 4 nm slit pore with graphite layer or quartz boundary.

As shown in Figure 38, S_{C_1,H_2} is greater than 1 regardless of the pressure, temperature, fluid composition and boundary condition, indicating that C_1 has been preferentially adsorbed compared with H_2. In CO_2 sequestration, the hope is that CO_2 can be permanently stored underground and that the preferential adsorption of CO_2 can be beneficial for long-term CO_2 sequestration. However, in H_2 storage, H_2 is temporarily stored underground and is extracted when needed. The preferential adsorption of C_1 allows H_2 molecules to remain freely in the pore, which ensures that a large amount of H_2 can be produced during the H_2 recovery process. There is no obvious trend of S_{C_1,H_2} with changes in pressure, temperature or fluid composition because the change in conditions may have affected the number of molecules in the adsorbed region and core region simultaneously. Relative selectivity is affected by the mole fraction of fluid molecules in both the adsorbed region and the core region; therefore, an increase in the number of molecules in the two

regions cannot be used to predict whether the relative selectivity will increase or decrease. For the boundary condition, S_{C_1,H_2} is higher under the graphite layer compared to the quartz boundary. The reason for this effect is that C_1 has strong interactions with the graphite layer [141] and more C_1 molecules have accumulated in the adsorbed region. The high percentage of C_1 in the adsorbed region leads to a large value of S_{C_1,H_2}.

As shown in Figure 39, S_{H_2O,C_1} and S_{H_2O,H_2} are greater than 1 regardless of the boundary conditions, indicating that H_2O has been preferentially adsorbed compared to C_1 and H_2. When the boundary is the graphite layer, S_{C_1,H_2} is greater than 1 because C_1 still forms the adsorption layer, as shown in Figure 35(a). However, in the slit pore with a quartz boundary, S_{C_1,H_2} is less than 1 because C_1 has been forced away to the core region due to the presence of H_2O molecules, as shown in Figure 35(b).

The above analysis shows the adsorption behaviors of C_1-H_2 and C_1-H_2-H_2O mixtures. For the C_1-H_2 mixture, C_1 is preferentially adsorbed compared with H_2. For the C_1-H_2-H_2O mixture, H_2O molecules are preferentially adsorbed compared to C_1 and H_2. These conclusions are made based on the results of relative selectivity. Preferential adsorption ensures that H_2 is freely distributed in the shale reservoirs and can be extracted easily during the H_2 recovery stage. The pressure, temperature and boundary material do not have significant effect on the adsorption characteristics of C_1 and H_2. However, when the mole fraction of C_1 is low, H_2 forms an adsorption layer, as shown in Figure 34(b). The adsorbed H_2 cannot be easily extracted during the H_2 recovery process, resulting in a waste of H_2. The adsorption of H_2 highlights the importance of cushion gas. A certain amount of cushion gas must be injected to avoid the consumption of working gas. Our results show that 50% cushion gas is sufficient for H_2 storage in depleted shale gas reservoirs if C_1 is used as the cushion gas.

4.3.2 H_2-HC Under Multi-scale Condition

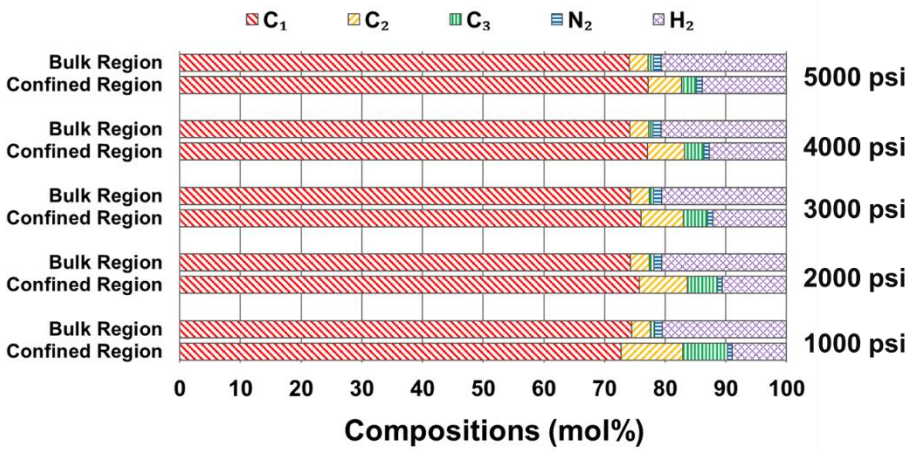
In this section, we will analyze the phase behavior and compositional distribution of hydrogen-shale gas (H_2-SG) system under multiscale conditions. The composition of shale gas is simplified from the work [132] by ignoring the components with mole fraction less than 1%. The composition of the simulated shale gas is summarized in Table 11. The initial pressure of shale gas reservoir is set as 5,000 psi while the reservoir temperature is considered as 130°F based on the references [154, 155]. To reduce the calculation errors, a total of 10,000 fluid molecules are simulated in each case. Different mole fractions of H_2 in the fluid systems are considered to observe the effect of shale gas percentage on H_2 storage.

Table 11. Shale gas composition in the simulation

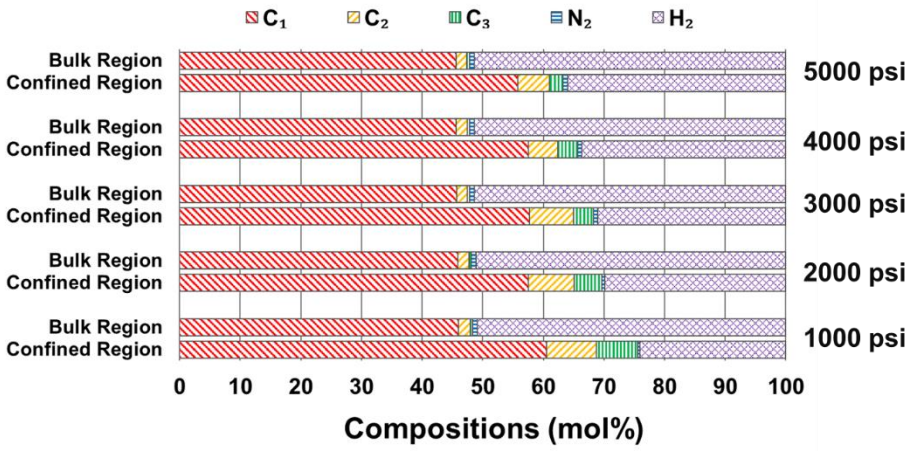
Component	Mole fraction (%)
C_1	93
C_2	4.2
C_3	1.2
N_2	1.6

The compositional distribution of H_2-SG system is shown in Figure 40. The red, yellow, green, blue and purple color bars represent the mole fractions of C_1, C_2, C_3, N_2 and H_2 in the bulk or confined regions, respectively. We can observe that the compositional heterogeneity exists between the fluids in the bulk and confined regions regardless of the fluid composition, which can be demonstrated from the different lengths of the color bars of a specific component in the two regions. Specially, hydrocarbons (C_1~C_3) tend to stay in the confined region while more H_2

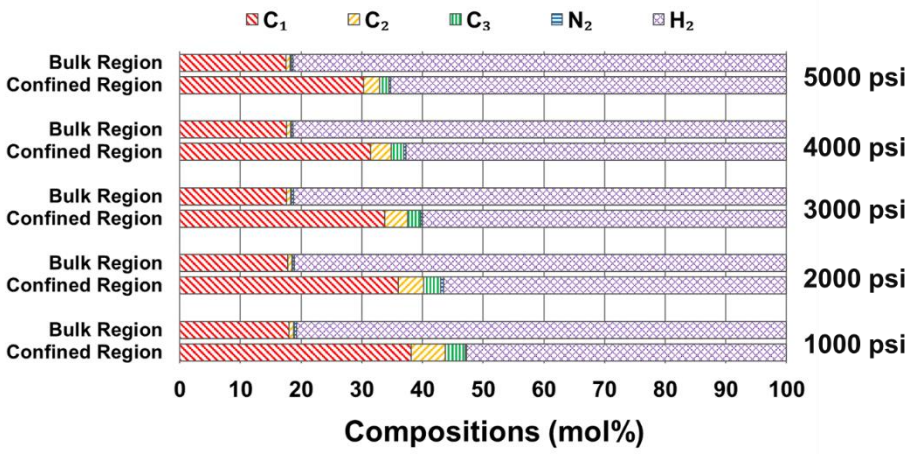
molecules are present in the bulk region. The reason for this is that hydrocarbons have higher interaction energies with the boundary in the confined region (Box III) compared with N_2 and H_2. The nonbonded interaction energy is calculated based on Equation (38) and the potential parameters can be found in Table 10. A higher value of ε_p leads to a stronger interaction energy. Because hydrocarbons have higher values of ε_p, the nonbonded interaction energy between graphite boundary and hydrocarbons is higher. With the decrease of pressure, the compositional heterogeneity becomes more significant, which can be observed from the change of the mole fraction differences of H_2 in the two regions.



(a) SG 80% and H_2 20%



(b) SG 50% and H_2 50%



(c) SG 20% and H_2 80%

Figure 40: Compositional distribution of H_2-SG system at different pressures under multiscale condition: a) SG 80% and H_2 20%, b) SG 50% and H_2 50%, and c) SG 20% and H_2 80%. The red, yellow, green, blue and purple color bars represent the mole fractions of C_1, C_2, C_3, N_2 and H_2 in the bulk or confined regions, respectively.

One important insight we can get from the above analysis is that H_2 has higher mole fraction in the bulk region than initial state (50%). During the extraction of H_2, the fluid in the bulk region is first produced due to the high mobility. Therefore, the extracted fluid has higher purity of H_2, which makes depleted shale gas reservoir a better candidate for H_2 storage than conventional gas reservoirs. With the decrease of pressure, the mole fraction of H_2 becomes higher in the bulk region, which leads to a higher purity of H_2 in the extracted fluid. The result indicates that shallow shale gas reservoirs are favorable candidates for H_2 storage. In addition, the volume percentage of nanopores in the system also affects the mole fraction of H_2 in the bulk region and therefore, affects the H_2 purity in the extracted fluid. In Figure 40(b), the volume percentage of nanopores at 3000 psi is about 5.30% in the whole system, which corresponds to 51.32% of H_2 in the bulk region. As

we increase the volume percentage of nanopores at 3000 psi to about 15.48%, the mole fraction of H_2 is 53.72% as shown in Figure 41. The increase of the volume percentage of nanopores causes more fluids flowing into nanopores. Therefore, more hydrocarbon molecules accumulate in the nanopores, while more H_2 molecules show up in the bulk region, which leads to a higher mole fraction of H_2 in the bulk region.

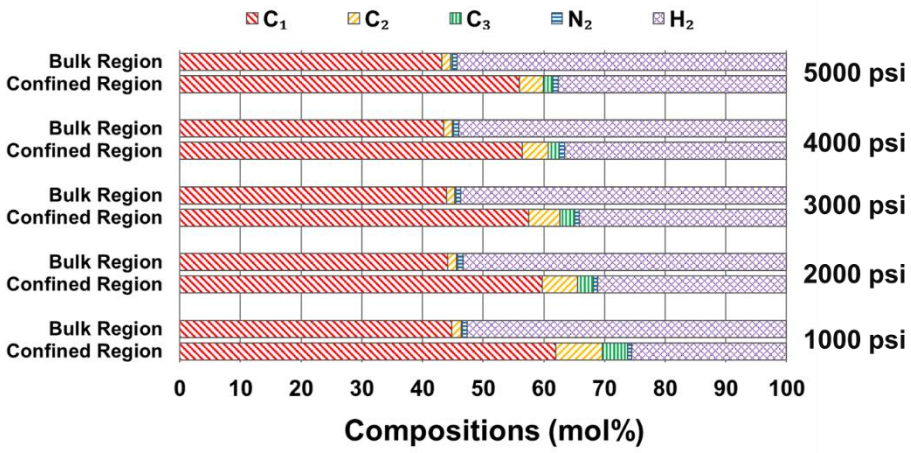
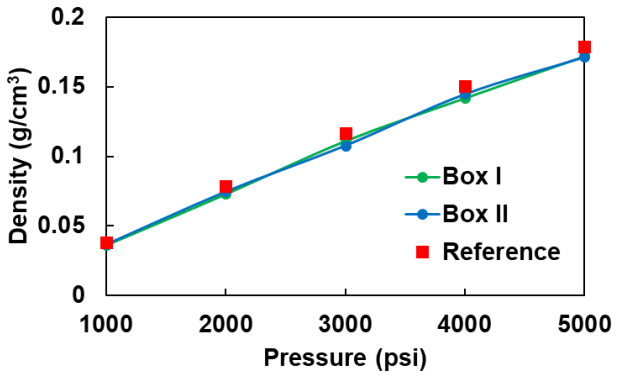


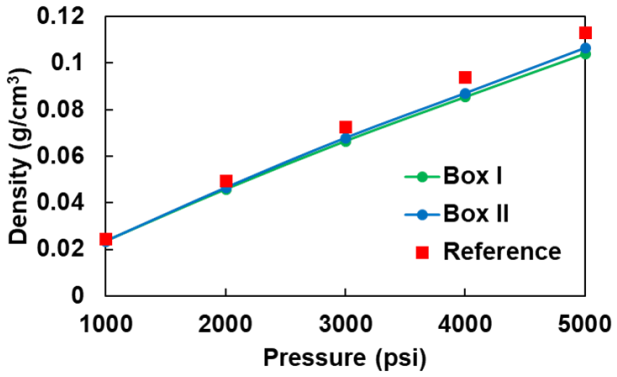
Figure 41. Compositional distribution of equimolar H_2-SG system at different pressures under multiscale condition. The volume percentage of nanopores at 3,000 psi is about 15.48%. The red, yellow, green, blue and purple color bars represent the mole fractions of C_1, C_2, C_3, N_2 and H_2 in the bulk or confined regions, respectively.

Next, we analyze the bulk fluid phase behavior of H_2-SG system. The fluid phase behavior in Monte Carlo simulation is determined by the fluid densities in the simulation boxes. Following the method mentioned in Section 4.1.1, we can compare the fluid densities of equimolar H_2-SG system in Box I and II with the reference gas-phase density to obtain the bulk fluid phase behavior, as shown in Figure 42. The fluid densities in Box I and II are both gas-phase densities, indicating that the bulk fluid is always in single gas phase at different pressures. The same method cannot be

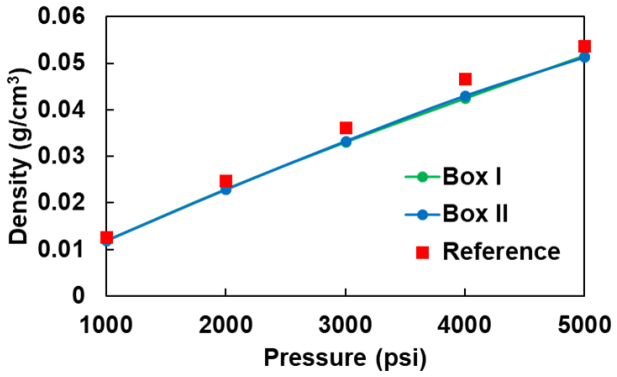
used to determine the fluid phase behavior in nanopores (Box III) because the fluid molecules in the nanopores are heterogeneously distributed due to the confinement effect. The high fluid densities in nanopores are caused by the adsorption of hydrocarbons, which is discussed later.



(a) SG 80% and H_2 20%



(b) SG 50% and H_2 50%



(c) SG 20% and H_2 80%

Figure 42: Density profile of H_2-SG system: a) SG 80% and H_2 20%, b) SG 50% and H_2 50%, c) SG 20% and H_2 80%. The green and blue lines represent the simulated fluid densities in Box I and II, respectively. The red dots are the reference gas-phase density of H_2-SG system, which is calculated by PR EOS.

As mentioned before, the extracted fluid with high H_2 purity is from the bulk region. When the bulk fluid is gradually extracted, fluid in nanopores will be released into the bulk region. However, the fluid released into the bulk region does not have the same composition as the fluid in the nanopores, which is affected by the adsorption of fluid molecules. The density profiles of H_2-HC system in nanopores are shown in Figures 34 and 35. Previous analysis shows that C_1 is preferentially adsorbed compared with H_2 in the C_1-H_2 mixture while H_2O molecules are preferentially adsorbed compared with C_1 and H_2 in the C_1-H_2-H_2O mixture regardless of pressure, temperature and boundary material. The preferential adsorption ensures that H_2 molecules are not adsorbed and have high mobility compared with hydrocarbons and H_2O. When the fluid in nanopores is released to bulk region, more H_2 molecules will move to the bulk region due to the high mobility. The adsorption of hydrocarbons and H_2O limits their migration and therefore, tends to trap them in the nanopores. During the extraction process, the increase of H_2 molecules released from the nanopores contributes to the high percentage of H_2 in bulk fluid, which further increases the purity of H_2 in the extracted fluid.

4.3.3 Implication for H_2 storage in depleted shale gas reservoirs

Here, we summarize the mechanism of H_2 storage in shale gas reservoirs. When H_2 is injected into shale gas reservoirs, the remaining shale gas and injected H_2 form the H_2-SG system. Due to the confinement effect, the compositional heterogeneity is observed between the bulk region and nanopores. The bulk region has higher percentage of H_2 while more hydrocarbons show up in nanopores. In the nanopores, the fluid molecules are heterogeneously distributed. The hydrocarbons are adsorbed near the boundary whereas H_2 molecules are freely distributed. During the extraction process, the bulk fluid is first recovered. Since the bulk fluid has higher percentage of H_2, the extracted fluid has higher purity of H_2, which improves the efficiency of H_2 extraction. With the continuous extraction of bulk fluid, the fluid in nanopores is released to the bulk region. Due to the mobility difference between hydrocarbon and H_2 molecules, more H_2 molecules are released to the bulk region, which further increases the percentage of H_2 in bulk fluid and the H_2 purity in extracted fluid. Therefore, we can conclude that the existence of nanopores in shale gas reservoirs allows higher proportion of H_2 in the bulk region and the increase of H_2 purity during H_2 extraction, which benefits the H_2 storage process.

5. SUMMARY AND FUTURE WORK

5.1 Summary

In this work, EOS modeling, ML and molecular simulation were applied to investigate fluid phase behavior in shale reservoirs. Due to the existence of widely distributed nanopores in shale reservoirs, fluids in shale reservoirs behave differently compared to the fluids in conventional reservoirs. EOS modeling estimated the fluid phase behavior in multiscale condition at the reservoir level, and ML was used to accelerate the phase-equilibrium calculation of the EOS modeling. Molecular simulation focused on the interactions between fluid molecules, which provided the foundation for analysis at the molecular level.

In the EOS modeling, the PR-C EOS was incorporated in the phase-equilibrium calculation to accurately simulate fluid phase behavior under both bulk and nanoscale conditions. The confinement effect on shale gas-condensate reservoir fluid is analyzed, and the phenomenon of abnormal depletion CGR was explained. In LRS production, oil recovery from LRS reservoirs is much lower compared to oil recovery from a conventional reservoir with the same drawdown. The reason for this difference is the compositional heterogeneity in LRS reservoirs. Hydrocarbons distribute heterogeneously with respect to pore size in nanoscale. There are fewer intermediate to heavy hydrocarbons (C_3-C_{11+}) but more light ends (C_1-C_2) in the bulk region. Because fluids are mainly produced from the bulk region, the leaner composition in the bulk region leads to relatively lower oil recovery. In addition, the variation of bulk fluid composition leads to a change in depletion CGR. The bulk fluid composition is not constant when the pressure is higher than the dew point pressure.

Although EOS modeling with the PR-C EOS was more accurate under multiscale condition, it is much more complex and time-consuming. In our simulations, the average CPU

time for phase-equilibrium calculation using the PR-C EOS was four orders of magnitude higher than the CPU time for phase-equilibrium calculation using the PR EOS. To accelerate the phase-equilibrium calculation using the PR-C EOS, the MLP and PINN were applied in the stability test and phase-split computation, respectively. In the stability test, the MLP was trained based on the input variables of the reservoir conditions and the EOS parameters to determine whether the fluid was in a single-phase or a two-phase condition. If the fluid was in two-phase condition, the PINN model was used to obtain an initial estimate of the equilibrium ratio for the Helmholtz free energy minimization algorithm instead of obtaining the estimate in stability test. In total, the average CPU time for phase-equilibrium calculation using the generalized ML-assisted phase-equilibrium model was reduced by more than two orders of magnitude while maintaining an accuracy of 97%.

Various molecular simulation models were applied to investigate the potential for H_2 storage in depleted shale gas reservoirs. The single bulk model was used to simulate H_2-HC phase behavior under bulk condition. The simulation results were highly consistent with experimental results, demonstrating the accuracy of the simulation process and molecular settings. The single nanoscale model was applied to investigate the adsorption behavior of an H_2-HC mixture in nanopores. For the C_1-H_2 mixture, C_1 was preferentially adsorbed compared with H_2. For the C_1-H_2-H_2O mixture, H_2O molecules were preferentially adsorbed compared with C_1 and H_2. The preferential adsorption ensured that H_2 was freely distributed in nanopores and was more likely to be released to the bulk region during the H_2 recovery stage. In the multiscale model, compositional heterogeneity was observed for the H_2-HC system. A higher proportion of H_2 appeared in the bulk region while more hydrocarbons existed in the nanopores. Because the bulk fluid contained a higher percentage of H_2, the extracted fluid had a higher purity of H_2 as well, which would improve the efficiency of H_2 extraction.

5.2 Future work

Extensions of the current work can focus on the improvement of EOS modeling, different ML techniques and wider application of molecular simulation.

The EOS modeling in this work was based on the PR-C EOS. Although the PR-C EOS is accurate in simulating fluid phase behavior under both bulk and nanoscale conditions, some of the equations and parameters are empirical or must be obtained from experiments. For example, in Equations (7) and (8), the square-well width (δ_p) has been set as $0.5\sigma_i$ regardless of fluid types. However, δ_p should be determined for each component to improve the accuracy. Equation (6) is an empirical expression of the term $F_{p,i}$, which refers to the percentage of molecules of component i in the surface-adsorbed region.. Luo et al. [36] have proposed a new method to calculate $F_{p,i}$ by considering two layers of adsorption. More efforts can be focused on improving the PR-C EOS.

The ML techniques used to accelerate the phase-equilibrium calculation were MLP and PINN. The MLP had the advantage of flexibility and simplicity, while the PINN could incorporate the PR-C EOS in the training process to achieve high accuracy with limited samples. The optimization of the generalized ML-assisted model focused on the structures of the MLP and PINN rather than the testing of different ML techniques. In future work, many other ML techniques can be used in the model to determine whether the model produces higher accuracy with other ML strategies.

A series of molecular simulations was performed with different hydrocarbon mixtures under various conditions. The same simulation model could be used for more complex cases. For example, for the single nanoscale model, one slit pore was used to investigate the fluid adsorption characteristics. The boundary materials used in this work were graphite and quartz. Further analyses can test other boundary materials, including kerogen and minerals (e.g., montmorillonite,

kaolinite, and illite). For the multiscale model, the compositional heterogeneity of the H_2-SG system was investigated under different conditions. Further analysis can include H_2O as a component to observe the effect of H_2O in the multiscale model.

REFERENCES

1. Wang, T., et al., *Molecular simulation of CO2/CH4 competitive adsorption on shale kerogen for CO2 sequestration and enhanced gas recovery.* The Journal of Physical Chemistry C, 2018. **122**(30): p. 17009-17018.
2. Gasparik, M., et al., *Geological controls on the methane storage capacity in organic-rich shales.* International Journal of Coal Geology, 2014. **123**: p. 34-51.
3. Clarkson, C.R., et al., *Pore structure characterization of North American shale gas reservoirs using USANS/SANS, gas adsorption, and mercury intrusion.* Fuel, 2013. **103**: p. 606-616.
4. Sing, K.S., *Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984).* Pure and applied chemistry, 1985. **57**(4): p. 603-619.
5. Rouquerol, J., et al., *Recommendations for the characterization of porous solids (Technical Report).* Pure and applied chemistry, 1994. **66**(8): p. 1739-1758.
6. Liu, X. and D. Zhang, *A review of phase behavior simulation of hydrocarbons in confined space: Implications for shale oil and shale gas.* Journal of Natural Gas Science and Engineering, 2019. **68**: p. 102901.
7. Kanfar, M.S. and C.R. Clarkson, *Reconciling flowback and production data: A novel history matching approach for liquid rich shale wells.* Journal of Natural Gas Science and Engineering, 2016. **33**: p. 1134-1148.
8. Alharthy, N., et al., *Enhanced oil recovery in liquid–rich shale reservoirs: laboratory to field.* SPE Reservoir Evaluation & Engineering, 2018. **21**(01): p. 137-159.

9. Whitson, C.H. and S. Sunjerga. *PVT in liquid-rich shale reservoirs*. in *SPE Annual Technical Conference and Exhibition*. 2012. OnePetro.

10. Peng, D.-Y. and D.B. Robinson, *A new two-constant equation of state.* Industrial & Engineering Chemistry Fundamentals, 1976. **15**(1): p. 59-64.

11. Travalloni, L., et al., *Thermodynamic modeling of confined fluids using an extension of the generalized van der Waals theory.* Chemical Engineering Science, 2010. **65**(10): p. 3088-3099.

12. Travalloni, L., M. Castier, and F.W. Tavares, *Phase equilibrium of fluids confined in porous media from an extended Peng–Robinson equation of state.* Fluid Phase Equilibria, 2014. **362**: p. 335-341.

13. White, C., R. Steeper, and A. Lutz, *The hydrogen-fueled internal combustion engine: a technical review.* International journal of hydrogen energy, 2006. **31**(10): p. 1292-1305.

14. Modisha, P.M., et al., *The prospect of hydrogen storage using liquid organic hydrogen carriers.* Energy & fuels, 2019. **33**(4): p. 2778-2796.

15. Muhammed, N.S., et al., *A review on underground hydrogen storage: Insight into geological sites, influencing factors and future outlook.* Energy Reports, 2022. **8**: p. 461-499.

16. Mouli-Castillo, J., N. Heinemann, and K. Edlmann, *Mapping geological hydrogen storage capacity and regional heating demands: An applied UK case study.* Applied Energy, 2021. **283**: p. 116348.

17. Chen, F., et al., *Capacity assessment and cost analysis of geologic storage of hydrogen: A case study in Intermountain-West Region USA.* International Journal of Hydrogen Energy, 2022.

18. Chen, F., et al. *Cost Analysis of Geologic Hydrogen Storage: A Case Study in the US Intermountain-West Regions*. in *Fall Meeting 2022*. 2022. AGU.

19. Wang, S., et al., *The potential of hydrogen storage in depleted unconventional gas reservoirs: A multiscale modeling study.* International Journal of Hydrogen Energy, 2023.

20. Chen, F., M. Mehana, and H. Nasrabadi, *Molecular Simulation of Hydrogen-Shale Gas System Phase Behavior under Multiscale Conditions: A Molecular-Level Analysis of Hydrogen Storage in Shale Gas Reservoirs.* Energy & Fuels, 2023.

21. Carrott, M.R., et al., *Adsorption of nitrogen, neopentane, n-hexane, benzene and methanol for the evaluation of pore sizes in silica grades of MCM-41*. Microporous and mesoporous materials, 2001. **47**(2-3): p. 323-337.

22. Yun, J.-H., et al., *Adsorption of methane, ethane, and their binary mixtures on MCM-41: experimental evaluation of methods for the prediction of adsorption equilibrium.* Langmuir, 2002. **18**(7): p. 2693-2701.

23. Qiao, S., S. Bhatia, and D. Nicholson, *Study of hexane adsorption in nanoporous MCM-41 silica.* Langmuir, 2003. **20**(2): p. 389-395.

24. Russo, P.A., M.M.L.R. Carrott, and P.J. Carrott, *Hydrocarbons adsorption on templated mesoporous materials: effect of the pore size, geometry and surface chemistry.* New Journal of Chemistry, 2011. **35**(2): p. 407-416.

25. Luo, S., H. Nasrabadi, and J.L. Lutkenhaus, *Effect of confinement on the bubble points of hydrocarbons in nanoporous media.* AIChE Journal, 2016. **62**(5): p. 1772-1780.

26. Luo, S., J.L. Lutkenhaus, and H. Nasrabadi, *Confinement-induced supercriticality and phase equilibria of hydrocarbons in nanopores.* Langmuir, 2016. **32**(44): p. 11506-11513.

27. Luo, S., J.L. Lutkenhaus, and H. Nasrabadi, *Use of differential scanning calorimetry to study phase behavior of hydrocarbon mixtures in nano-scale porous media.* Journal of Petroleum Science and Engineering, 2018. **163**: p. 731-738.

28. Alfi, M., H. Nasrabadi, and D. Banerjee, *Experimental investigation of confinement effect on phase behavior of hexane, heptane and octane using lab-on-a-chip technology.* Fluid Phase Equilibria, 2016. **423**: p. 25-33.

29. Alfi, M., H. Nasrabadi, and D. Banerjee. *Effect of confinement on bubble point temperature shift of hydrocarbon mixtures: Experimental investigation using nanofluidic devices*. in *SPE Annual Technical Conference and Exhibition*. 2017. OnePetro.

30. Parsa, E., X. Yin, and E. Ozkan. *Direct observation of the impact of nanopore confinement on petroleum gas condensation*. in *SPE Annual Technical Conference and Exhibition*. 2015. OnePetro.

31. Yang, Q., et al., *Direct visualization and molecular simulation of dewpoint pressure of a confined fluid in sub-10 nm slit pores.* Fuel, 2019. **235**: p. 1216-1223.

32. Yang, Q., et al., *Direct Observation of the Vapor–Liquid Phase Transition and Hysteresis in 2 nm Nanochannels.* Langmuir, 2022. **38**(32): p. 9790-9798.

33. Dong, X., et al., *Phase equilibria of confined fluids in nanopores of tight and shale rocks considering the effect of capillary pressure and adsorption film.* Industrial & Engineering Chemistry Research, 2016. **55**(3): p. 798-811.

34. Siripatrachai, N., T. Ertekin, and R. Johns. *Compositional simulation of discrete fractures incorporating the effect of capillary pressure on phase behavior*. in *SPE Improved Oil Recovery Conference*. 2016. OnePetro.

35. Devegowda, D., et al. *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*. 2012. OnePetro.

36. Luo, S., et al., *A novel pore-size-dependent equation of state for modeling fluid phase behavior in nanopores.* Fluid Phase Equilibria, 2019. **498**: p. 72-85.

37. Chang, Q., L. Huang, and X. Wu, *A Molecular Dynamics Study on Low-Pressure Carbon Dioxide in the Water/Oil Interface for Enhanced Oil Recovery.* SPE Journal, 2022: p. 1-10.

38. Jin, L., et al., *Natural gas density under extremely high pressure and high temperature: Comparison of molecular dynamics simulation with corresponding state model.* Chinese Journal of Chemical Engineering, 2021. **31**: p. 2-9.

39. Chang, Q., L. Huang, and X. Wu. *Combination of simplified local density theory and molecular dynamics simulation to study the local density distribution of hydrocarbon gas in shale gas reservoir*. in *SPE Eastern Regional Meeting*. 2019. OnePetro.

40. Sun, Z., et al., *Molecular dynamics of methane flow behavior through realistic organic nanopores under geologic shale condition: pore size and kerogen types.* Chemical Engineering Journal, 2020. **398**: p. 124341.

41. Pitakbunkate, T., et al., *Effect of confinement on pressure/volume/temperature properties of hydrocarbons in shale reservoirs.* SPE Journal, 2016. **21**(02): p. 621-634.

42. Lowry, E. and M. Piri, *Effects of chemical and physical heterogeneity on confined phase behavior in nanopores.* Microporous and Mesoporous Materials, 2018. **263**: p. 53-61.

43. Panagiotopoulos, A.Z., *Adsorption and capillary condensation of fluids in cylindrical pores by Monte Carlo simulation in the Gibbs ensemble.* Molecular Physics, 1987. **62**(3): p. 701-719.

44. Neimark, A.V. and A. Vishnyakov, *Gauge cell method for simulation studies of phase transitions in confined systems.* Physical Review E, 2000. **62**(4): p. 4611.

45. Vasileiadis, M., et al., *Transport properties of shale gas in relation to kerogen porosity.* The Journal of Physical Chemistry C, 2018. **122**(11): p. 6166-6177.

46. Jin, B. and H. Nasrabadi, *Phase behavior of multi-component hydrocarbon systems in nano-pores using gauge-GCMC molecular simulation.* Fluid Phase Equilibria, 2016. **425**: p. 324-334.

47. Jin, B., R. Bi, and H. Nasrabadi, *Molecular simulation of the pore size distribution effect on phase behavior of methane confined in nanopores.* Fluid Phase Equilibria, 2017. **452**: p. 94-102.

48. Jin, B. and H. Nasrabadi, *Phase behavior in shale organic/inorganic nanopores from molecular simulation.* SPE Reservoir Evaluation & Engineering, 2018. **21**(03): p. 626-637.

49. Mehana, M., et al., *Machine-learning predictions of the shale wells' performance.* Journal of Natural Gas Science and Engineering, 2021. **88**: p. 103819.

50. Wang, Y., H. Liu, and Y. Zhou, *Development of a deep learning-based model for the entire production process of steam-assisted gravity drainage (SAGD).* Fuel, 2021. **287**: p. 119565.

51. Syed, F.I., et al., *Smart shale gas production performance analysis using machine learning applications.* Petroleum Research, 2022. **7**(1): p. 21-31.

52. Tang, J., et al., *A new ensemble machine-learning framework for searching sweet spots in shale reservoirs.* SPE Journal, 2021. **26**(01): p. 482-497.

53. Groven, S.D., C. Desgranges, and J. Delhommelle, *Prediction of the boiling and critical points of polycyclic aromatic hydrocarbons via Wang-Landau simulations and machine learning.* Fluid Phase Equilibria, 2019. **484**: p. 225-231.

54. Wang, K., et al., *Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations.* Journal of Computational Physics, 2020. **401**: p. 109013.

55. Wang, Y., et al., *Image recognition model based on deep learning for remaining oil recognition from visualization experiment.* Fuel, 2021. **291**: p. 120216.

56. Wu, Y., et al., *Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales.* Fuel, 2019. **253**: p. 662-676.

57. Mohanty, S., *Estimation of vapour liquid equilibria for the system carbon dioxide–difluoromethane using artificial neural networks.* International Journal of Refrigeration, 2006. **29**(2): p. 243-249.

58. Sharma, R., et al., *Potential applications of artificial neural networks to thermodynamics: vapor–liquid equilibrium predictions.* Computers & chemical engineering, 1999. **23**(3): p. 385-390.

59. Petersen, R., A. Fredenslund, and P. Rasmussen, *Artificial neural networks as a predictive tool for vapor-liquid equilibrium.* Computers & chemical engineering, 1994. **18**: p. S63-S67.

60. Nguyen, V.D., et al., *Prediction of vapor–liquid equilibrium data for ternary systems using artificial neural networks.* Fluid phase equilibria, 2007. **254**(1-2): p. 188-197.

61. Gaganis, V. and N. Varotsis, *An integrated approach for rapid phase behavior calculations in compositional modeling.* Journal of Petroleum Science and Engineering, 2014. **118**: p. 74-87.

62. Burges, C.J., *A tutorial on support vector machines for pattern recognition.* Data mining and knowledge discovery, 1998. **2**(2): p. 121-167.

63. Rachford, H.H. and J. Rice, *Procedure for use of electronic digital computers in calculating flash vaporization hydrocarbon equilibrium.* Journal of Petroleum Technology, 1952. **4**(10): p. 19-3.

64. Kashinath, A., M.L. Szulczewski, and A.H. Dogru, *A fast algorithm for calculating isothermal phase behavior using machine learning.* Fluid Phase Equilibria, 2018. **465**: p. 73-82.

65. Tipping, M.E., *Sparse Bayesian learning and the relevance vector machine.* Journal of machine learning research, 2001. **1**(Jun): p. 211-244.

66. Mesbah, M., et al., *Vapor liquid equilibrium prediction of carbon dioxide and hydrocarbon systems using LSSVM algorithm.* The Journal of Supercritical Fluids, 2015. **97**: p. 256-267.

67. Huang, H.-Z., et al., *Support vector machine based estimation of remaining useful life: current research status and future trends.* Journal of Mechanical Science and Technology, 2015. **29**(1): p. 151-163.

68. Wang, S., et al., *Accelerating and stabilizing the vapor-liquid equilibrium (VLE) calculation in compositional simulation of unconventional reservoirs using deep learning based flash calculation.* Fuel, 2019. **253**: p. 209-219.

69. Carley, D.D., *Thermodynamic properties of a square‐well fluid in the liquid and vapor regions.* The Journal of Chemical Physics, 1983. **78**(9): p. 5776-5781.

70. Chapela, G.A., S.E. Martínez‐Casas, and C. Varea, *Square well orthobaric densities via spinodal decomposition.* The Journal of chemical physics, 1987. **86**(10): p. 5683-5688.

71. Gale, J.F., et al., *Natural fractures in shale: A review and new observationsNatural Fractures in Shale: A Review and New Observations.* AAPG bulletin, 2014. **98**(11): p. 2165-2216.

72. Warpinski, N., *Measurement of width and pressure in a propagating hydraulic fracture.* Society of Petroleum Engineers Journal, 1985. **25**(01): p. 46-54.

73. Ko, L.T., et al., *Origin and characterization of Eagle Ford pore networks in the south Texas Upper Cretaceous shelf.* AAPG Bulletin, 2017. **101**(3): p. 387-418.

74. Curtis, M.E., et al. *Transmission and scanning electron microscopy investigation of pore connectivity of gas shales on the nanoscale*. in *North American unconventional gas conference and exhibition*. 2011. OnePetro.

75. Sigal, R.F., *Pore-size distributions for organic-shale-reservoir rocks from nuclear-magnetic-resonance spectra combined with adsorption measurements.* SPE Journal, 2015. **20**(04): p. 824-830.

76. Driskill, B., et al., *11 Applications of SEM Imaging to Reservoir Characterization in the Eagle Ford Shale, South Texas, USA.* 2013.

77. Bernard, S., et al., *Geochemical evolution of organic-rich shales with increasing maturity: A STXM and TEM study of the Posidonia Shale (Lower Toarcian, northern Germany).* Marine and Petroleum Geology, 2012. **31**(1): p. 70-89.

78. Javadpour, F., M. Moravvej Farshi, and M. Amrein, *Atomic-force microscopy: a new tool for gas-shale characterization.* Journal of Canadian Petroleum Technology, 2012. **51**(04): p. 236-243.

79. Luo, S., J.L. Lutkenhaus, and H. Nasrabadi, *Multiscale fluid-phase-behavior simulation in shale reservoirs using a pore-size-dependent equation of state.* SPE Reservoir Evaluation & Engineering, 2018. **21**(04): p. 806-820.

80. Vega, L., et al., *Phase equilibria and critical behavior of square‐well fluids of variable width by Gibbs ensemble Monte Carlo simulation.* The Journal of chemical physics, 1992. **96**(3): p. 2296-2305.

81. Singh, J.K. and S.K. Kwak, *Surface tension and vapor-liquid phase coexistence of confined square-well fluid.* The Journal of chemical physics, 2007. **126**(2): p. 024702.

82. Alfi, M., et al. *Pore Size Variability and Sieving Effect in Liquid Shale—A Multiple Permeability Approach and Eagle Ford Case Study*. in *SPE Reservoir Simulation Conference*. 2017. OnePetro.

83. Luo, S., et al., *Multiscale Pressure/Volume/Temperature Simulation of Decreasing Condensate/Gas Ratio at Greater than Dewpoint Pressure in Shale Gas-Condensate Reservoirs.* SPE Journal, 2021. **26**(06): p. 4174-4186.

84. Uzun, I., B. Kurtoglu, and H. Kazemi, *Multiphase rate-transient analysis in unconventional reservoirs: theory and application.* SPE Reservoir Evaluation & Engineering, 2016. **19**(04): p. 553-566.

85. Dachanuwattana, S., et al., *Application of proxy-based MCMC and EDFM to history match a shale gas condensate well.* Journal of Petroleum Science and Engineering, 2018. **167**: p. 486-497.

86. Abiodun, O.I., et al., *State-of-the-art in artificial neural network applications: A survey*. Heliyon, 2018. **4**(11): p. e00938.

87. Ramchoun, H., et al., *Multilayer perceptron: Architecture optimization and training.* 2016.

88. Schiffmann, W., M. Joost, and R. Werner, *Optimization of the backpropagation algorithm for training multilayer perceptrons.* University of Koblenz: Institute of Physics, 1994.

89. Nair, V. and G.E. Hinton. *Rectified linear units improve restricted boltzmann machines*. in *Icml*. 2010.

90. Cai, S., et al., *Physics-informed neural networks (PINNs) for fluid mechanics: A review*. Acta Mechanica Sinica, 2022: p. 1-12.

91. Cai, S., et al., *Physics-informed neural networks for heat transfer problems.* Journal of Heat Transfer, 2021. **143**(6).

92. Raissi, M., P. Perdikaris, and G.E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations.* Journal of Computational physics, 2019. **378**: p. 686-707.

93. Fevang, Ø., K. Singh, and C.H. Whitson. *Guidelines for choosing compositional and black-oil models for volatile oil and gas-condensate reservoirs*. in *SPE annual technical conference and exhibition*. 2000. OnePetro.

94. Li, Y.-K., L.X. Nghiem, and A. Siu, *Phase behaviour computations for reservoir fluids: effect of pseudo-components on phase diagrams and simulation results.* Journal of Canadian petroleum technology, 1985. **24**(06).

95. Newley, T. and R. Merrill, *Pseudocomponent selection for compositional simulation.* SPE reservoir engineering, 1991. **6**(04): p. 490-496.

96. Akkurt, R., et al. *Accelerating and enhancing petrophysical analysis with machine learning: a case study of an automated system for well log outlier detection and reconstruction*. in *SPWLA 59th Annual Logging Symposium*. 2018. OnePetro.

97. Chen, F., et al., *A generalized machine learning-assisted phase-equilibrium calculation model for shale reservoirs.* Fluid Phase Equilibria, 2022. **558**: p. 113423.

98. Stein, M., *Large sample properties of simulations using Latin hypercube sampling.* Technometrics, 1987. **29**(2): p. 143-151.

99. Michelsen, M.L., W. Yan, and E.H. Stenby, *A comparative study of reduced-variables-based flash and conventional flash.* SPE Journal, 2013. **18**(05): p. 952-959.

100. Kingma, D.P. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.

101. Abou-Moustafa, K. and C. Szepesvári, *An a priori exponential tail bound for k-folds cross-validation.* arXiv preprint arXiv:1706.05801, 2017.

102. Marcot, B.G. and A.M. Hanea, *What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?* Computational Statistics, 2021. **36**(3): p. 2009-2031.

103. Chalmers, G.R., R.M. Bustin, and I.M. Power, *Characterization of gas shale pore systems by porosimetry, pycnometry, surface area, and field emission scanning electron microscopy/transmission electron microscopy image analyses: Examples from the Barnett, Woodford, Haynesville, Marcellus, and Doig units.* AAPG bulletin, 2012. **96**(6): p. 1099-1119.

104. Kuila, U. and M. Prasad, *Specific surface area and pore‐size distribution in clays and shales.* Geophysical Prospecting, 2013. **61**(2-Rock Physics for Reservoir Exploration, Characterisation and Monitoring): p. 341-362.

105. Ross, D.J. and R.M. Bustin, *The importance of shale composition and pore structure upon gas storage potential of shale gas reservoirs.* Marine and petroleum Geology, 2009. **26**(6): p. 916-927.

106. Wu, S., et al., *Reservoir quality characterization of upper triassic Chang 7 shale in Ordos basin.* Earth Science, 2015. **40**(11): p. 1810-1823.

107. Widom, B., *Some topics in the theory of fluids.* The Journal of Chemical Physics, 1963. **39**(11): p. 2808-2812.

108. Tian, Y., C. Yan, and Z. Jin, *Characterization of methane excess and absolute adsorption in various clay nanopores from molecular simulation.* Scientific reports, 2017. **7**(1): p. 1-21.

109. Jin, Z. and A. Firoozabadi, *Effect of water on methane and carbon dioxide sorption in clay minerals by Monte Carlo simulations.* Fluid Phase Equilibria, 2014. **382**: p. 10-20.

110. Bi, R. and H. Nasrabadi, *Molecular simulation of the constant composition expansion experiment in shale multi-scale systems.* Fluid Phase Equilibria, 2019. **495**: p. 59-68.

111. Chen, F., R. Bi, and H. Nasrabadi. *Molecular simulation of multi-scale multi-component hydrocarbon phase behavior in liquid-rich shale reservoirs*. in *Unconventional Resources Technology Conference, 26–28 July 2021*. 2021. Unconventional Resources Technology Conference (URTeC).

112. Ungerer, P., B. Tavitian, and A. Boutin, *Applications of molecular simulation in the oil and gas industry: Monte Carlo methods*. 2005: Editions Technip.

113. Gubbins, K.E., Y. Long, and M. Śliwinska-Bartkowiak, *Thermodynamics of confined nano-phases.* The Journal of Chemical Thermodynamics, 2014. **74**: p. 169-183.

114. Martin, M.G. and A.L. Frischknecht, *Using arbitrary trial distributions to improve intramolecular sampling in configurational-bias Monte Carlo.* Molecular Physics, 2006. **104**(15): p. 2439-2456.

115. Lorentz, H.A., *Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase.* Annalen der physik, 1881. **248**(1): p. 127-136.

116. Martin, M.G. and J.I. Siepmann, *Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. 2. United-atom description of branched alkanes.* The Journal of Physical Chemistry B, 1999. **103**(21): p. 4508-4517.

117. Potoff, J.J. and J.I. Siepmann, *Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen.* AIChE journal, 2001. **47**(7): p. 1676-1682.

118. Cygan, R.T., J.-J. Liang, and A.G. Kalinichev, *Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field.* The Journal of Physical Chemistry B, 2004. **108**(4): p. 1255-1266.

119. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, 1983. **79**(2): p. 926-935.

120. Vega, C., et al., *What ice can teach us about water interactions: a critical comparison of the performance of different water models.* Faraday discussions, 2009. **141**: p. 251-276.

121. Guevara-Carrion, G., J. Vrabec, and H. Hasse, *Prediction of self-diffusion coefficient and shear viscosity of water and its binary mixtures with methanol and ethanol by molecular simulation.* The Journal of chemical physics, 2011. **134**(7): p. 074508.

122. Getman, R.B., et al., *Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal–organic frameworks.* Chemical reviews, 2012. **112**(2): p. 703-723.

123. Downs, R.T. and M. Hall-Wallace, *The American Mineralogist crystal structure database.* American Mineralogist, 2003. **88**(1): p. 247-250.

124. Maddox, M. and K. Gubbins, *Molecular simulation of fluid adsorption in buckytubes.* Langmuir, 1995. **11**(10): p. 3988-3996.

125. Martin, M.G., *MCCCS Towhee: a tool for Monte Carlo molecular simulation.* Molecular Simulation, 2013. **39**(14-15): p. 1212-1222.

126. Yang, Q. and C. Zhong, *Molecular simulation of adsorption and diffusion of hydrogen in metal− organic frameworks.* The Journal of Physical Chemistry B, 2005. **109**(24): p. 11862-11864.

127. Benham, A.L., D. Katz, and R. Williams, *Phase behavior of hydrogen‐light‐hydrocarbon systems.* AIChE Journal, 1957. **3**(2): p. 236-241.

128. Williams, R.B. and D.L. Katz, *Vapor-liquid equilibria in binary systems. hydrogen with ethylene, ethane, propylene, and propane.* Industrial & Engineering Chemistry, 1954. **46**(12): p. 2512-2520.

129. Burriss, W., et al., *Phase behavior of the hydrogen-propane system.* Industrial & Engineering Chemistry, 1953. **45**(1): p. 210-213.

130. Klink, A., H. Cheh, and E. Amick Jr, *The vapor‐liquid equilibrium of the hydrogen—n‐butane system at elevated pressures.* AIChE Journal, 1975. **21**(6): p. 1142-1148.

131. Nichols, W., H. Reamer, and B. Sage, *Volumetric and phase behavior in the hydrogen‐n‐hexane system.* AIChE Journal, 1957. **3**(2): p. 262-267.

132. Demirbas, A., N. Cek, and S. Acar, *Chemical analyses of shale gas and conventional natural gas.* Petroleum Science and Technology, 2018. **36**(20): p. 1690-1695.

133. Zhang, T., et al., *Effect of organic-matter type and thermal maturity on methane adsorption in shale-gas systems.* Organic geochemistry, 2012. **47**: p. 120-131.

134. Linstrom, P.J. and W.G. Mallard, *The NIST Chemistry WebBook: A chemical data resource on the internet.* Journal of Chemical & Engineering Data, 2001. **46**(5): p. 1059-1063.

135. Chen, G., et al., *Keys to linking GCMC simulations and shale gas adsorption experiments.* Fuel, 2017. **199**: p. 14-21.

136. Cao, T., et al., *A comparative study of the specific surface area and pore structure of different shales and their kerogens.* Science China Earth Sciences, 2015. **58**(4): p. 510-522.

137. Liu, H., et al., *A study on geological characteristics of marine shale gas reservoir in Southern China.* Energy Exploration & Exploitation, 2015. **33**(3): p. 299-308.

138. Pedersen, K. and N. Lindeloff. *Simulations of compositional gradients in hydrocarbon reservoirs under the influence of a temperature gradient*. in *SPE Annual Technical Conference and Exhibition*. 2003. OnePetro.

139. Zhang, Y., et al., *Geological characteristics and abnormal pore pressure prediction in shale oil formations of the Dongying depression, China.* Energy Science & Engineering, 2020. **8**(6): p. 1962-1979.

140. Li, W., et al., *Molecular simulation study on methane adsorption capacity and mechanism in clay minerals: Effect of clay type, pressure, and water saturation in shales.* Energy & Fuels, 2019. **33**(2): p. 765-778.

141. Sun, J., et al., *Asymmetric competitive adsorption of CO2/CH4 binary mixture in shale matrix with heterogeneous surfaces.* Chemical Engineering Journal, 2021. **422**: p. 130025.

142. Chalmers, G.R. and M.R. Bustin, *PS the effects and distribution of moisture in gas shale reservoir systems.* GR Chalmers, MR Bustin, 2010.

143. Fan, E., et al., *Methane sorption capacity of organics and clays in high-over matured shale-gas systems.* Energy Exploration & Exploitation, 2014. **32**(6): p. 927-942.

144. Ji, L., et al., *Experimental investigation of main controls to methane adsorption in clay-rich rocks.* Applied Geochemistry, 2012. **27**(12): p. 2533-2545.

145. Zeng, K., et al., *Molecular simulation of carbon dioxide and methane adsorption in shale organic nanopores.* Energy & Fuels, 2018. **33**(3): p. 1785-1796.

146. Li, J., et al., *An adsorption model for evaluating methane adsorption capacity in shale under various pressures and moisture.* Journal of Natural Gas Science and Engineering, 2020. **81**: p. 103426.

147. Zhou, S., et al., *Investigation of methane adsorption mechanism on Longmaxi shale by combining the micropore filling and monolayer coverage theories.* Advances in Geo-Energy Research, 2018. **2**(3): p. 269-281.

148. Didar, B.R. and I.Y. Akkutlu. *Pore-size dependence of fluid phase behavior and properties in organic-rich shale reservoirs*. in *SPE international symposium on oilfield chemistry*. 2013. OnePetro.

149. Li, Z., D. Cao, and J. Wu, *Layering, condensation, and evaporation of short chains in narrow slit pores.* The Journal of chemical physics, 2005. **122**(22): p. 224701.

150. Chong, L. and E.M. Myshakin, *Molecular simulations of competitive adsorption of carbon dioxide–methane mixture on illitic clay surfaces.* Fluid Phase Equilibria, 2018. **472**: p. 185-195.

151. Hu, X., et al., *Characterization of CO2/CH4 competitive adsorption in various clay minerals in relation to shale gas recovery from molecular simulation.* Energy & Fuels, 2019. **33**(9): p. 8202-8214.

152. Talbot, J., *Analysis of adsorption selectivity in a one‐dimensional model system.* AIChE journal, 1997. **43**(10): p. 2471-2478.

153. Wang, T., et al., *Selective adsorption of supercritical carbon dioxide and methane binary mixture in shale kerogen nanopores.* Journal of Natural Gas Science and Engineering, 2018. **50**: p. 181-188.

154. Osholake, T., J. Yilin Wang, and T. Ertekin, *Factors affecting hydraulically fractured well performance in the Marcellus shale gas reservoirs.* Journal of Energy Resources Technology, 2013. **135**(1).

155. Agrawal, A., Y. Wei, and S. Holditch, *A technical and economic study of completion techniques in five emerging US gas shales: A Woodford Shale example.* SPE Drilling & Completion, 2012. **27**(01): p. 39-49.