

Modeling Nanoconfinement Effects Using Active Learning

Published as part of *The Journal of Physical Chemistry* virtual special issue "Machine Learning in Physical Chemistry".

Javier E. Santos,* Mohammed Mehana, Hao Wu, Maša Prodanović, Qinjun Kang, Nicholas Lubbers, Hari Viswanathan, and Michael J. Pyrcz



Cite This: *J. Phys. Chem. C* 2020, 124, 22200–22211



Read Online

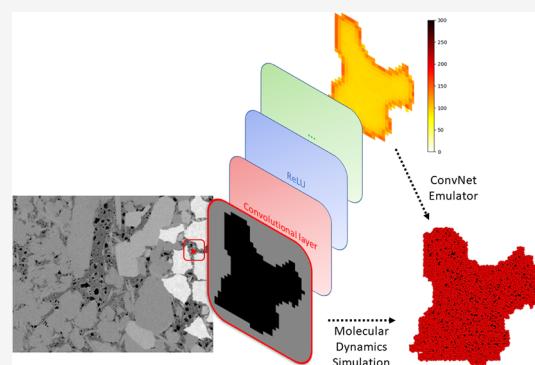
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Predicting the spatial configuration of gas in nanopores is relevant in applications such as fluid flow forecasting and hydrocarbon reserves estimation. For example, shale reservoirs have suffered from computationally intractable multiscale problems, since fluid properties such as viscosity, density, and adsorption must be calculated by using expensive molecular dynamics (MD) simulations within each nanopore, whereas flow through these connected nanopores must be simulated at the micrometer scale. We utilize machine learning techniques to quickly and accurately model nanoscale confinement effects as an important step toward bridging the nano and micro scales. Our workflow is based on building and training physics-based deep-neural-networks models by learning from a database of MD calculations. The model accounts for the adsorption phenomenon by predicting the statistical distribution of gas inside nanopores. Because large databases of MD calculations are expensive to create, we investigate active learning (AL) as a data set construction strategy. In this workflow, new data are selected based on the model uncertainty via the query-by-committee approach. We show that our workflow obtains accurate models that generalize to real scanning electron microscopy geometries with 1/10th of the number of MD calculations required vs random data set generation. Our method enables the possibility of modeling nanoconfinement effects at the mesoscale, where complex connected sets of nanopores affect flow.



1. INTRODUCTION

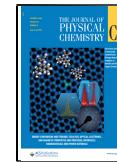
Understanding nanoscale phenomena in the conditions of naturally occurring porous media provides a scientific challenge with numerous applications in energy and environment, such as hydrocarbon recovery, CO₂ sequestration, and energy storage and conversion. The demand of natural gas as an energy source for industrial processes and domestic consumption has encouraged the production of unconventional resources, predominantly from shale formations. Methane, the principal hydrocarbon constituent in shale gas plays, is an attractive asset due to its energy density, ease of transport, and environmental advantages over resources like coal. Recent technological advances (hydraulic fracturing, slick water, and horizontal drilling) have made possible the economic utilization of these resources.^{1,2} Because of these, the world has witnessed a significant increase of natural gas production from shale reservoirs in the past decade. Nevertheless, accurately estimating hydrocarbon reserves (gas in place) and forecasting well performance are tasks that remain unsolved.^{3–6} One of the main reasons is that most of the pore bodies in these formations are under 50 nm in size; hence, nanoconfinement effects must be integrated to properly model fluid behavior.^{7–9} Molecular dynamics (MD) simulations can

describe nanoconfinement effects accurately.^{10–16} These simulations have shown that nanoconfinement effects affect fluid properties such as viscosity, density, critical point, and adsorbed concentration. However, the computational cost of MD does not allow modeling more than a few connected nanopores; also, realistic pore shapes are not commonly modeled in practice. Therefore, there is a pressing need to accelerate these simulations to allow large-scale modeling of the porous media while preserving the physical effects accompanying nanoconfinement. Similar concerns apply to porous media in carbon sequestration, energy storage, and energy conversion, such that all of these applications require accurate simulations of nanoconfinement effects to predict and optimize the system behavior.

Received: August 14, 2020

Revised: September 15, 2020

Published: September 18, 2020



With the recent increase of hydrocarbon production coming from shale reservoirs, there is also a pressing need of better physical models that can explain the field data better. Contrary to conventional hydrocarbon reservoirs (i.e., sandstones and carbonates), 80% of the pores in shale are less than 50 nm in size.¹⁷ As a result, the actual *in situ* gas volumes significantly deviate from predictions made using conventional methods (up to 2.5 times¹⁸). Likewise, for dynamic problems, research has shown that conventional methods greatly underestimate flow through nanopores, finding a discrepancies up to 2 orders of magnitude compared with the Hagen–Poiseuille flow solution.¹⁹

In general, characterizing fluid properties under confinement is an active research topic.^{19–22} Physical properties deviate from analytical models when the size of the bounding walls (in this case the pore) approaches the size of the free fluid molecule.^{23,24} Several experimental and field observations have confirmed that the fluid behavior deviates from the bulk at this scale.^{21,25} Among the physical phenomena that take place in shale formations, the adsorption of gas into the solid pore walls is the most relevant, to the extent that in some cases most of the gas exists as a dense adsorbed layer (instead of in free gas form) in the vicinity of the pore wall. This causes a non-negligible deviation in volume estimates carried out with equations of state and also hampers flow at the microscale because the adsorbed layers reduce the volume of void space available for flow and affect the slippage of the fluids.²⁶ Therefore, it is imperative to investigate the methane adsorption behavior in shale plays at the appropriate scale.

MD simulations are well suited for representing domains with these characteristics thanks to their ability to track the individual molecules of a given system. MD simulations are needed in these types of systems because they present fluid–fluid and fluid–solid interactions that do not have known analytical or coarse-grained representations. To date, these simulations have provided new insights into the confinement effect on the phase properties of hydrocarbons, showing how nanoconfinement affected condensation conditions in nanopores.²⁷ Nevertheless, previous computational studies have not addressed the effects of a large body of complex, realistic pore shapes.^{14,19,28,29}

One of the main bottlenecks toward a better understanding of adsorption in shale formations is their heterogeneity in composition. Shale formations are composed by two main components: the inorganic (which includes kaolinite, illite, smectite, quartz, carbonates, and sulfide minerals) and the organic (mainly composed by kerogen). Each of these materials exhibits unique adsorption characteristics that have been studied in previous work. For instance, Jin and Firoozabadi³⁰ studied the adsorption characteristics of both methane and CO₂ in montmorillonite (smectite group) pores, finding that the surface charge and area mainly affected the adsorbed volumes. Xiong et al.³¹ investigated the methane adsorption in quartz pores, concluding the adsorption was only important in pores larger than 6 nm. On the other hand, the organic fraction of shale changes with maturity, affecting how it interacts with the fluids. Huang et al.³² observed in their simulations that the adsorption capacity for methane is linearly correlated to the thermal maturity. Moreover, Ho et al.²⁹ saw that these adsorbed layers change the strain of the organic material. All these studies provide important insights into the molecular interactions of gas with a simplified surface of a given material. In this study, we build upon these works and

construct more complex structures for our simulation domains. It is worth noting that the lattice Boltzmann method (LBM) allows the simulation of larger physical domains where complex connected sets of nanopores control fluid behavior.³³ Nonetheless, LBM needs to be informed by MD (in the form of pairwise interactions³⁴ to account for the nanoconfinement effects). Currently, there is no functional relationship to calculate these *a priori*.

Recently, deep learning (DL)³⁵ has shown outstanding results achieved by trained models that capture physical phenomena of interest for a variety of scientific fields.^{36–40} These are relevant because they provide accurate predictions that are orders of magnitude faster in computational speed than traditional full-physics numerical simulators. The success of DL models partially stems from the availability of these to learn from large amounts of data during training.⁴¹ DL models have shown an unprecedented ability to improve as more data becomes available and can process inputs in their raw form.³⁵ The main bottleneck for deploying physics-based DL models is that simulating physical processes (MD simulation in this work) is computationally intensive. An efficient approach to tackle this issue would be to selectively sample points in space with the goal of creating a data set to improve the model's generalization capabilities and accuracy when applied to data not used to train the model. The ML framework of active learning (AL) attempts to overcome the bottleneck of insufficient training data by building a model that requests new data where there is no sufficient information; AL focuses on minimizing the size of data set required by the networks via designing a self-improving system.⁴² AL is of particular interest for systems where the training data comprises computer simulations, so that an automated workflow can continuously collect data and retrain the model.^{43,44} AL reduces the computational expense of creating comprehensive high-fidelity data sets. Therefore, the domain of molecular dynamics simulations provides an ideal problem context for AL techniques. In this paper, we show the effectiveness of using AL as a means for possibly be able to bring molecular information into mesoscale models of nanoporous media.

In our proposed method, we build a neural network model that describes absorption under nanoconfinement in complex and realistic pore geometries. This model allows at least 3 orders of magnitude speed-up compared to the conventional MD workflow and enables the application of describing complex domains accurately. We also show an efficient method for creating training sets via AL that is also able to estimate the uncertainty of the prediction. Although we focus our efforts in adsorption, the workflow that we propose is applicable for other nanoconfinement effects, such as effective viscosity, density, and the shifting of critical points.

2. METHODS

In this section, we present the techniques that we utilize in our workflow. We first describe our different nanopore geometries, which consist of three subsets: simple synthetic geometries, complex synthetic geometries, and empirically obtained geometries. From these subsets, we use MD to obtain fluid density distributions under nanoconfinement. We model these densities with convolutional neural network models, which naturally capture the spatial correlations of the data. Finally, we describe our active learning workflow, which is powered by a query-by-committee ensemble⁴⁵ that efficiently selects new

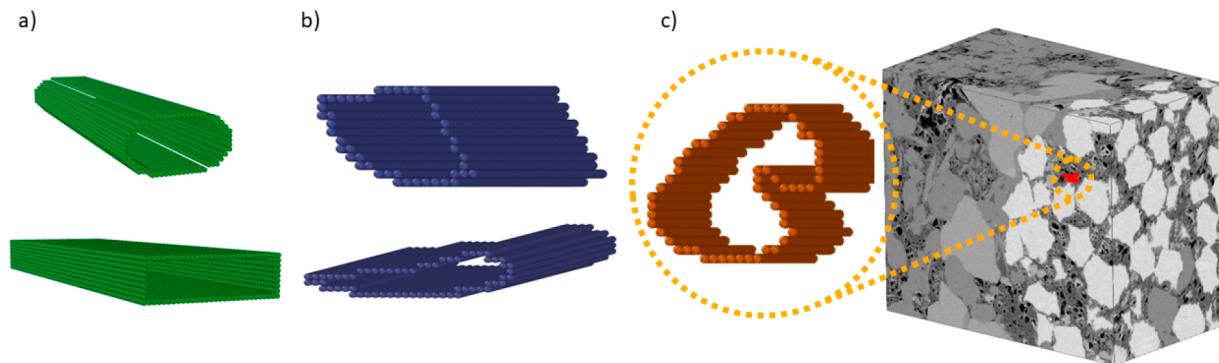


Figure 1. Three simulation domain subsets: (a) simple geometries, (b) irregular geometries, and (c) real pores from a 3D image of the Vaca Muerta formation. The solid spheres show the walls of our domains (which are impermeable). The images were created by using the Ovito software.⁴⁷

data to augment the training set and retrain the models as needed.

2.1. Nanopore Simulation Domains. To create a comprehensive data set for training and testing the models, we produced three different domain subsets with increasing complexity. All of them are bounded in a $120 \times 120 \times 373 \text{ \AA}^3$ box with periodic boundaries in all directions. These examples are 3D geometries with a constant cross section in the Z-axis, forming a nanotube. The motivation behind having three different data sets is to be able to assess how the different models perform with increasingly intricate domains.

The three types of nanopores considered here are the following:

- *Simple geometries*: ellipses and rectangles of varying cross-sectional area, centered in different locations across the plane (Figure 1a). These types of geometries are amenable to analytical solutions.
- *Irregular geometries*: synthetic geometries with different number of corners, concave and convex shapes, and varying cross-sectional areas (Figure 1b).
- *Vaca Muerta pores*: discretized pores from an 3D FIB-SEM image of the Vaca Muerta formation in Argentina⁴⁶ (Figure 1c). This image is obtained using a scanning electron microscope (SEM) combined with a focused ion beam (FIB). The setup enables to image volumes, which fully describe minerals, pore shapes, and connectivity. Using this imaging technique, it is possible to capture inter- and intraparticle pores and organic matter pores that typically are in the nanometer range.

The shape factor,⁴⁸ the ratio of the cross-sectional area to the square of the perimeter, can be used to display the diversity of shapes in these training sets. Section S1 of the Supporting Information discusses this, demonstrating that simple, irregular, and Vaca Muerta pores represent increasingly diverse sets of pore geometries.

Our goal is to train a model to perform accurately in real pores using only synthetic geometries for training (without leaking any information about this subset's characteristics *a priori*), while minimizing the amount of training data needed for this task.

2.2. Molecular Dynamics (MD) Simulation. Because our target pores are nanometer-sized, continuum and mesoscopic approaches fail to capture the relevant physics at this scale. Therefore, we utilize molecular dynamics to simulate the system of interest. We chose the LAMMPS software⁴⁹ to carry

out our simulations. Specifically, we aim to estimate the density profile of a fluid inside a pore under confinement conditions. The density profile will be significantly affected by liquid–solid interactions at the pore wall, especially for pores less than 50 nm. It should be noted that methane adsorption is not a critical process that affects hydrocarbon extraction for larger pores (i.e., sandstones);³¹ in fact, methane adsorption can often be neglected in these cases. However, for smaller pores, because of the large surface-to-volume ratio and proximity to the wall, gas adsorbed to the solid pore wall could account for up to 85% of the total hydrocarbon volume in a pore, acting as a major factor affecting production.

We used molecular dynamics simulations in this paper since ability of the tool to capture the absorption phenomenon has been validated by experimental measurements.^{50,51} Another promising molecular simulation technique is the grand canonical Monte Carlo (GCMC). Song et al.⁵² showed the big impact that different pore geometries have in the adsorption isotherm using this method. Additionally, Cao et al.⁵³ showed that for adsorption both methods yield the same result; while MD was faster, the GCMC has more flexibility to build different pore shapes.

For our proof of concept, we used frozen methane molecules to simulate the solid walls of our structures, since this gives us the ability to easily construct any given three-dimensional structure. Because the material has affinity for itself, this will attract the free gas molecules to the wall, forming adsorbed layers. The gas inside the pore is assumed to follow the united-atom model at a given initial density (controlled by the number of molecules placed inside the domain). We used the TraPPE force field⁵⁴ to describe the molecular interactions with a cutoff range of 10 Å. We would like to emphasize that in our workflow various molecular dynamics schemes could be employed to simulate a variety of conditions, using e.g. different interaction parameters for different wall materials or all-atom force fields. Different materials will exhibit unique affinities to the free fluid, which will affect the adsorbed layer thickness and amplitude across the pore geometry.

To simulate accurately the adsorption phenomenon at the nanoscale, the equilibrium of the system must be ensured. To accomplish this, we start the simulation with an energy-minimization stage that avoids the overlapping of atoms, followed by an equilibration stage where a canonical ensemble is used for 500 ps. Finally, we carry out a sampling phase for 1 ns in a constant temperature ensemble using a Nosé–Hoover

thermostat. All the simulations achieved the steady state (measured by the lack of change in local density for 250 ps).

Figure 2 shows the evolution of the molecular profile of a nanopore extracted from the SEM image of the Vaca Muerta

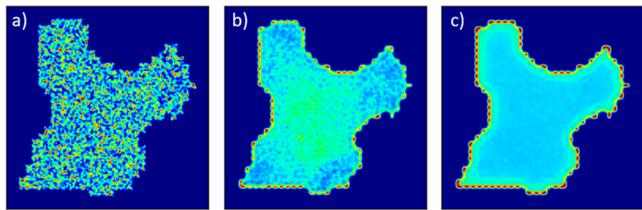


Figure 2. Steps of our simulation workflow showing the evolution in position of the gas molecules: (a) the initial domain (homogeneous gas density), (b) the density profile after 500 ps, and (c) the density profile after equilibrium. These are in arbitrary units; we aim to show how the molecules are preferentially located near the solid boundary.

formation through these stages. During the equilibration, the system structure is arranged where the fluid molecules adhere to the solid interface, which macroscopically translates to the adsorption phenomenon.

To create a diverse data set of pore geometries at different initial densities, a random number of methane molecules were distributed inside the nanopores (over a range of bulk average density between 50 and 250 kg/m³). After the simulation run (with computational time of about 1 h in a desktop computer), we average the gas density in the Z-coordinate to calculate a representative density profile. Finally, we bin these results into a 32 × 32 mesh (with a resolution of 3.75 Å per bin, slightly larger than the effective diameter of the methane molecule).

In the simulation results, we observe a heterogeneous density profile across the pore, where an adsorbed layer circumscribes the solid walls (**Figure 3**). We emphasize that for

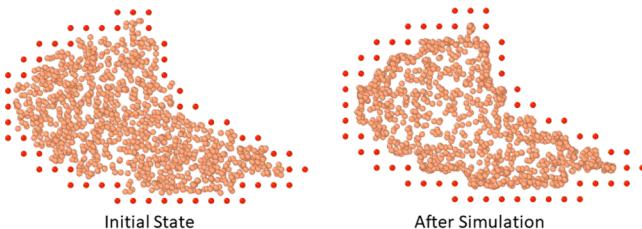


Figure 3. Cross-sectional view of one sample of the irregular geometry subset. The left figure shows the initial state of the system (before the canonical ensemble simulation), where the free gas molecules (light brown) are placed homogeneously in the space bounded by the impermeable pore walls (red). The average density of this system is 50 kg/m³. The figure to the right shows the converged state after the MD simulation. The amount of gas molecules (density) near the walls increases due to the adsorption phenomenon (220% higher than the bulk density for this case).

irregularly shaped nanopores the adsorption profile is complex and greatly influences the hydrocarbon transport. The molecules in this layer are held by van der Waals interactions between the solid walls and the fluid molecules. We observed a gradual decrease in the methane density as we move from the adsorbed layer to the bulk fluid (pore center). This behavior is strongly influenced by the pore geometry and the initial density of the gas. To highlight this, we performed four simulations at different initial density in the same domain.

Figure 4 shows the big impact that geometry and initial density have on the adsorption profile. As the initial density increases, the adsorption peaks diminish. With regard to the pore geometry, where fluid molecules interact with multiple solid surfaces, they form higher density clusters.

2.3. Models Investigated. Because it would be impossible to have a training set that includes every possible geometry/gas density that could occur in nature, we use the simple shape data set (**Figure 1a**) to train different models, followed by a performance assessment on the irregular geometries set (**Figure 1b**). This process shows the generalization capability of each model. As a baseline, we compare deep learning models against linear regression on geometric features. The fitting of the model was done using 1200 samples (70% fitting, 20% validation, and 10% test) of the simple shape data set. We also test these models on the irregular synthetic geometries (1500 samples) to investigate how well the model can generalize to more complex data.

2.3.1. Linear Regression. Linear regression is a simple model that provides a highly interpretable baseline to assess machine learning (ML) models with more parameters. Linear regression models a scalar relationship between distinct input features and an output response (in this case, gas density) by assigning a weight to each of the input variables, which correspond to how the function changes with the value of the input data. These models are easy to fit and can provide a reference point to compare the performance of our deep learning models. The features that we use to train the linear models are the initial density (ρ_0), Euclidean distance (E_{dist}), and the sample porosity (Φ). The Euclidean distance measures the separation between a point inside a pore with the closest wall. The porosity is obtained by dividing the area of the pore over the total cross-sectional area.

The results of linear regression are displayed in **Table 1**. The initial density gets a weight very close to one in every case because of the fact that mass is conserved in the MD simulation, so a weight close to one minimizes the least-squares regression. This leads to good predictions in the center of big pores, where the accuracy decreases near the walls (see **Figure S2**). These results confirm the need of a ML model that takes in account the spatial relationships of the pore structure with its MD solution.

2.3.2. Convolutional Neural Networks. In this subsection, we attempt to obtain the density profile of a random pore geometry under different initial bulk densities using convolutional neural networks (ConvNets). **Figure 5** illustrates our workflow, wherein we learn the density profile from a data set of molecular dynamics calculations on varying pore geometries. Because the pore geometry, initial density, and final distribution of fluid molecules can be represented as images, we propose employing ConvNets, well-known for their capabilities in image processing, to capture the relationship between the inputs and outputs of MD.

The input to the neural network model is an image where the solid boundaries and the outside of the pore are labeled with zeros (gray in the initial domain picture of **Figure 5**), and the rest of the image is labeled with the initial density of the gas. Therefore, the image will contain only two values: zero and a uniform value with initial density in every voxel of the pore (labeled in black in **Figure 5**). This image represents the initial uniform density at bulk conditions, which will be disturbed by the presence of the solids. Additionally, another input representing different solid interactions could be passes

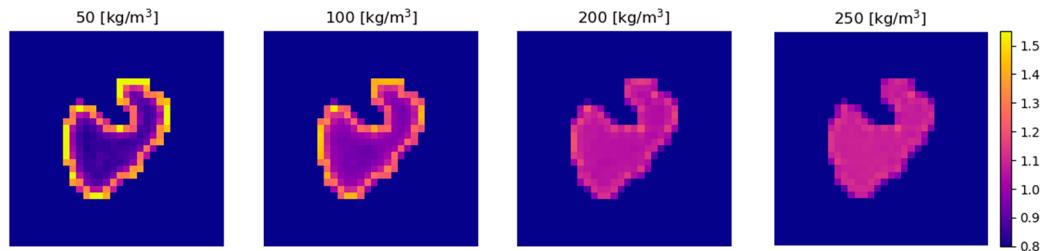


Figure 4. Effect of geometry and initial density in the final solution. The plots show the results of four simulations performed in the same domain using a different initial density (specified on top of each image). We normalized the results using the initial density of each domain. These results are within the ranges observed in the literature.^{13,55}

Table 1. Results of the Linear Regression Models Using Different Input Features

	inputs	test set error, % (20% of the simple shapes)	irregular shapes set error, % (1500 samples)	resultant equation
1	initial density	9.01	10.65	$\rho = 1.0035\rho_0$
2	initial density and Euclidean distance	7.41	10.50	$\rho = 0.998\rho_0 - 4.17E_{\text{dist}}$
3	initial density, Euclidean distance, and porosity	7.04	8.83	$\rho = 1.0034\rho_0 - 4.93E_{\text{dist}} + 0.2772\Phi$

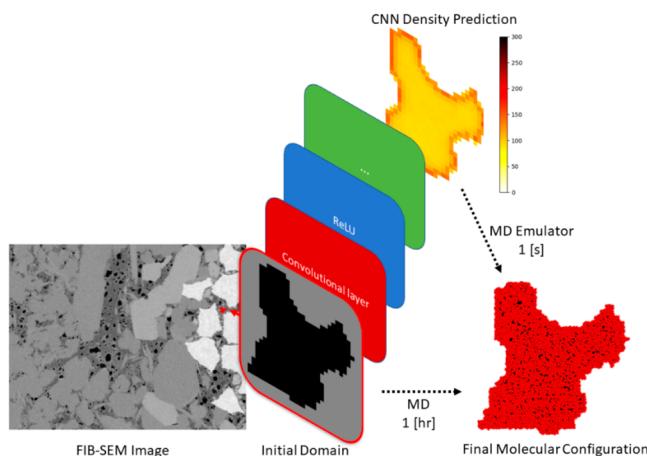


Figure 5. Workflow. Conventionally, starting from an initial domain (in this case, a pore from a FIB-SEM image of the Vaca Muerta formation), we generate the binarized pore geometry. In a traditional workflow, an MD simulation is performed to obtain the final configuration of gas in the system. Each MD calculation takes on the order of 1 h to complete. A neural network model learns from a database of such calculations to make predictions in less than 1 ms.

to account for solids with heterogeneous composition. The output of the model is the density profile as obtained with MD, where the gas takes on a heterogeneous structure due to the formation of adsorbed layers. This workflow is agnostic to the application and could be adopted to predict other properties in porous media (either in the solid or through the void). In this paper, we work with images that can fit an entire closed system (pore with impermeable walls) to present our methodology, since adsorption happens locally. 3D open systems can be handled with the addition of features that inform the model about both the local and the global aspects of the system.³⁶

ConvNets have excelled in the field of computer vision outperforming classical machine learning methods.^{41,56} These

models have shown a remarkable capacity to identify complex spatial relations in many data sets. By utilizing the convolution operation, the network layers process local spatial relationships across the domain. ConvNets utilize filters that are much smaller than the input image (in our case, filters are 3 by 3 pixels), which extract general and meaningful information about the domain in an efficient manner. By stacking convolutional layers, the network extracts features at different levels of abstraction (the first layers learn about the pore corner and edges, while the later ones learn about larger pore structures) with an increasingly wider receptive field (the first layer only sees with a field of 3 pixels while the last one can see the whole domain). However, stacking convolutional layers alone only provides a linear model; a nonlinear *activation function* is interleaved between convolutions to provide the model with greater expressivity.³⁵ A schematic of this is shown in Figure 6. The key aspect of convolution layers is that they are equivariant to translation, which means that if the input images are shifted by a number of pixels across the image plane, their output will be shifted by the same amount. This means that the exact location of the pore within the plane does not matter. This is particularly useful in pattern recognition because ConvNets allow inputs of variable size. By use of this structure, a network can be trained to learn complex, nonlinear relationships between inputs and outputs via an optimization algorithm (in this case, backpropagation⁵⁷), training the network by minimizing a *loss function* which quantifies the error of the network. For our problem, we use the mean absolute error as the loss function.

To build a ConvNet that maps pore geometry initially filled with gas at a certain, homogeneous density to with the density profile at equilibrium, we tested different convolutional neural network architectures to find a model that can take into account the pore geometry and density to make accurate predictions. Because the MD density fields are defined over the same geometry as the input, we selected an architecture that is structured by stacked layers that do not coarse-grain the image (so-called *pooling* in ConvNet terminology), preserving its dimensions at every stage of the process. The network architecture consists of n -stacks of 2D convolutions, a batch normalization layer,⁵⁸ and a ReLU (rectified linear unit)⁵⁹ nonlinear activation layer (as depicted in Figure 6). The ReLU function is given by

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

This nonlinear function allows the network to set threshold to activate certain outputs of the feature maps from the previous layer. In our work, the first convolutional layer is assigned a set number of filters. The subsequent convolutional layers have twice as many filters as their predecessors, similar to the image

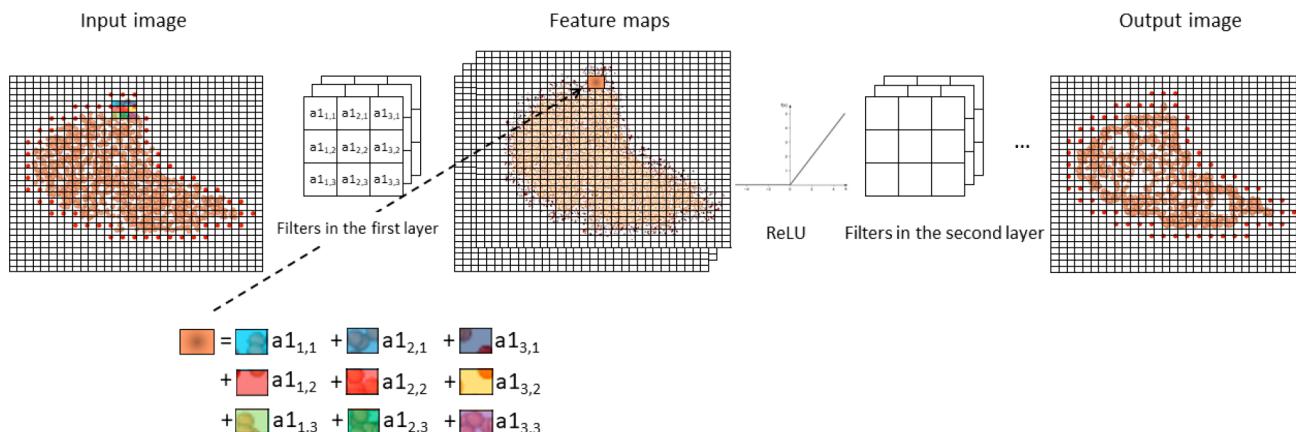


Figure 6. Schematic of a convolutional neural network. The input image is convolved by the three filters in the first layer. The filters have trainable parameters (in the case of the first filter these are labeled as $a_{1,ij}$) that are adjusted to minimize the mismatch between the output of the neural network and the one from the MD. Convolving the input with the first layer results in three feature maps in the next layer. These feature maps are an abstract representation that the machine encoded in a certain way that is useful to the next layers. Finally, these feature maps are passed through a nonlinear activation function (in our case a ReLU or a SoftPlus). This procedure is done subsequently for every layer in the network.

processing approach in ref 60). The last layer of the network is a Softplus function given by

$$\text{Softplus}(x) = \log(1 + e^x) \quad (2)$$

The range of this smooth function is all positive numbers, which matches the range of molecular densities. Further details on network training can be found in section S2 of the Supporting Information.

To assess the optimal complexity and number of trainable parameters of our convolutional neural network model, we test many network architectures; the goal of this exercise to select the best network structure for the modeling gas adsorption. The architecture variables that we tested are the number of convolutional layers and number of filters. We tested networks that ranged from 1 to 9 layers and the number of first layer filters varying from 2 to 64. Five identical networks are trained for each architecture using different random number seeds, and their results were averaged. The average test error of each architecture (with five trained networks) is shown in Figure 7 compared to the baseline generated by linear regression. This error is quantified as follows:

$$\text{error} = \frac{1}{IM} \sum_{im=0}^{IM} \frac{1}{32^2} \sum_{p=0}^P \left| \frac{\rho_{\text{MD}} - \rho_{\text{model}}}{\rho_{\text{MD}}} \right| \times 100 \quad (3)$$

This equation stands for the relative error (compared with the MD solution) in every pixel (P) inside the pore of the image, in every image (IM) of the data set.

Figure 7 shows the results of our tested ConvNet architectures, comparing the error against the number of trainable parameters and noting the time taken to train the network. The primary considerations for selecting an architecture were balancing the error of the network against the computational cost of training. We selected a network architecture (indicated by the cross in Figure 7) that performs accurately (1.6% relative error) in the simple shape test set while minimizing the training time of the network. The mean relative error of five trained networks of the selected architecture on the irregular shape set is 5.3% (compared to 8.8% of the best linear regression model), with a maximum individual pixel error of 96%. Note that this model was trained with just simple shapes (ellipses and rectangles). In our

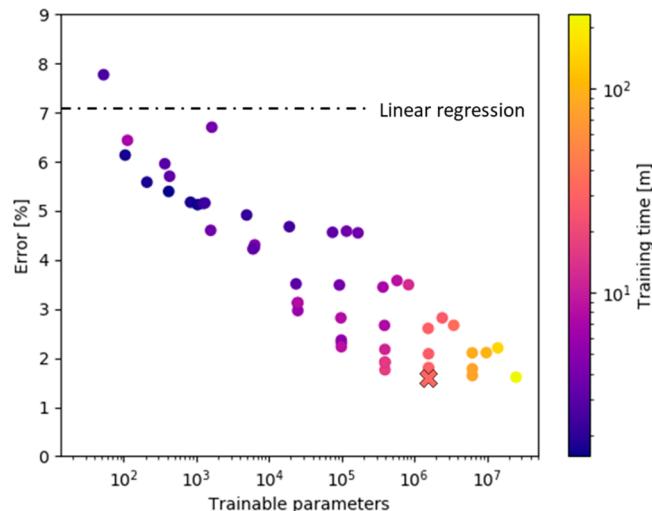


Figure 7. Sensitivity analysis of the network architecture parameters on the simple shapes test set, measured as mean relative error per pixel of the average predictions of five trained models. Each point represents a network architecture; the x (x) highlights the best performing architecture (according to error and training time). This model yields an average error of 1.6% with a training time of 31 min. We selected this as our backbone architecture for the rest of this work.

preliminary studies, we tried more complicated neural network architectures such as the U-Net⁶¹ and the Res-Net,⁶² but they did not improve network performance. We attribute this to the fact the adsorbed layers are a mostly local phenomenon that may be less complex than the image processing tasks for which these architectures were designed.

2.3.3. Specializing ConvNets for Adsorption. Here we present simple, physically motivated improvements of the model architecture to increase the model performance. To ensure that there is no mass of methane leaking outside the pore walls in the prediction of the gas density of our network, we added an additional layer augmented to the end of the selected architecture. This layer is a mask containing the initial gas density inside the pore. The layer scales the results of the ConvNet by the input density, and ensures that the values of gas density at the solid walls remain to be zero. Conceptually,

this can be represented as a chain of nonlinear functions f_i (for each layer i of the network) as follows, where ρ_0 represents the network input image and ρ_{pred} the predicted profile from the network:

$$\rho_{\text{pred}} = \rho_0(f_n(f_{n-1}(f_0(\rho_0)))) \quad (4)$$

Using this new structure, the network learns to create a function that scales the initial homogeneous density to the final one exhibiting the adsorbed profile. The result of this addition is shown in **Table 2**. This improves the performance of the

Table 2. Comparison of Adding a Density Scaling Layer to the Network with the Mass Balance Error of the Prediction

	test set error, % (from the simple shapes subset)	irregular shapes set error, %	max pixel error in the irregular shape set, %	av total mass error in the irregular shape set, %
standard ConvNet	1.6	5.3	96	3.5
density scaling	1.4	4.8	78	2.6

model; the error on irregular shapes decreases from 5.3% to 4.8%, and the total mass error is reduced from 3.5% to 2.6% with this simple additional factor.

Because we average the predictions across five models, we would like to minimize the disagreement between them. To achieve this, we added a regularizing term to our loss function:

$$L = |\rho_{\text{true}} - \rho_{\text{pred}}| + \alpha|\overline{\rho}_{\text{in}} - \overline{\rho}_{\text{pred}}| \quad (5)$$

where ρ_{true} and ρ_{pred} stand for the true density (from the MD simulation) and the predicted one (from the network), respectively. In the second term, $\overline{\rho}_{\text{in}}$ is the mean density (the initial state, before the simulation) and α is a hyperparameter. This additional term can be seen as a penalty in case there is a difference between the initial and the final mass of gas.

A comparison of the networks with the density scaling and this proposed regularization using different α values is displayed in **Table 3**. We stress that these results are from

Table 3. Relative Error and Standard Deviation of the Predictions (across the Five Networks) Using Different α Coefficients That Regularize the Network Training

α	rel error on irregular shape set, %	std dev of the relative errors across the five networks, %
0	4.8	0.13
1	4.6	0.15
2	4.6	0.15
3	4.7	0.13
4	4.6	0.09
5	4.6	0.14

models trained with simple shapes to assess the best possible network architecture and hyperparameters. From **Table 3**, α does not have a strong effect on the performance of the ensemble of networks in aggregate; however, it does affect the disagreement between individual networks in the ensemble. We selected an α value of four since it exhibits the lowest standard deviation throughout 1500 samples.

The accuracy results of linear regression and various convnet models are collected in histogram form in **Figure S3**.

2.4. Active Learning. Next, we describe our active learning strategy to minimize the number of nanoscale simulations required as part of the model training. ConvNets have demonstrated great performance in pattern recognition tasks when very large labeled data sets are provided. Nevertheless, creating comprehensive training sets, especially for physics-based simulations, is an expensive task. For example, to create a set of 1000 labeled pairs of MD data, one would need roughly 1 month of computational time on a desktop computer. Because of this hurdle, we employ an active learning method to label our data in an efficient manner. AL selects images (candidates for training) from a big pool of data to label them (in our case, they are labeled by performing an MD simulation) with the goal of minimizing the relative error of the model.

Parting from the selected network architecture, we trained five identical neural networks that differ only in the initial random number used (which only affects the network initialization and the order of the samples employed per batch of data). The first generation of five networks use the exact same data set to train. At the end of the training process (as explained in the *Supporting Information*), predictions with each trained model are performed throughout the unlabeled pool of images (where only the cross-sectional shape and the initial density are provided). Because the target result (MD simulation) is unknown for this set, we record the disagreement between network predictions. If this disagreement is high, we flag the sample as candidate for labeling (this process is depicted in **Figure 8**). We then select the n samples presenting the highest disagreement, label them (by performing an MD simulation), and use them to train the next generation of networks. By adding these examples to the initial training set, we are reducing the uncertainty space of the models (**Figure 9**).

3. RESULTS AND DISCUSSION

Because MD calculations are relatively expensive, and the space of possible pore geometries is enormous, it is important to develop a model that generalizes to the complexity of real pore geometries from different sources without needing to know in advance what pores it will be applied to. To this end, we constructed an AL workflow to build models, which generalize to the Vaca Muerta pores by training to a limited number of synthetic pore geometries,

The strategy adaptively augments the training set by selecting pore geometries that have not been trained to but are likely to improve the model performance. The method follows the query-by-committee⁴⁵ technique, which takes advantage of the fact that disagreement between different models is a reasonable proxy for model uncertainty and can be used to prioritize additions to the data set. For further details, see **section 2.4**. We start from a pool of examples containing 3000 samples of simple and irregular shapes at different initial density. The initial training set is composed of eight pores drawn from the simple shapes subset (**Figure 1a**). After the training of the five networks is completed (with only these eight examples), predictions with each trained model are performed throughout the unlabeled samples in the pool. To estimate the uncertainty of the predictions, the standard deviation of the predicted density (across the five network predictions) is recorded. The standard deviation per example is normalized using the initial density of the domain, so that the disagreement is dimensionless. **Figure 9**, left panel, shows

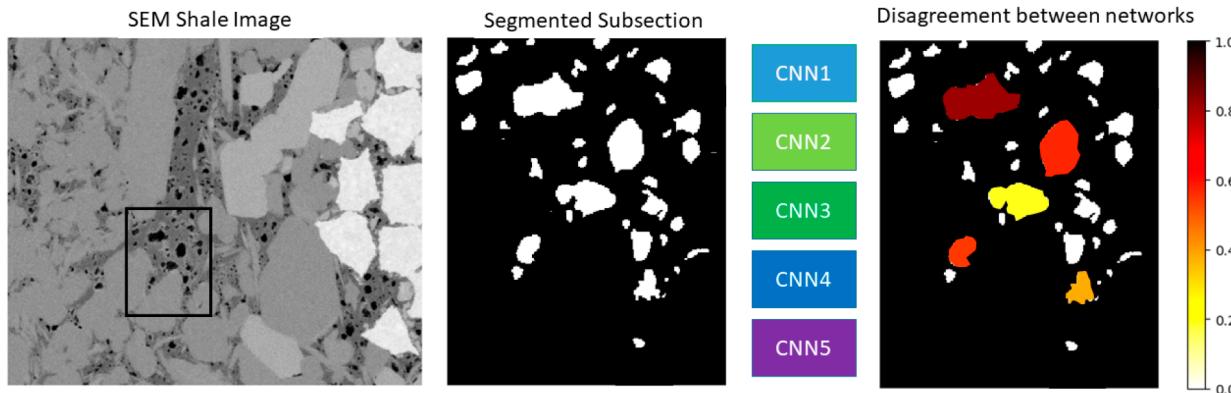


Figure 8. Active learning workflow. We use five trained neural networks (shown as boxes) to make predictions in pores with different shapes. We then flag the locations where these five networks diverge the most. The color bar is in arbitrary units.

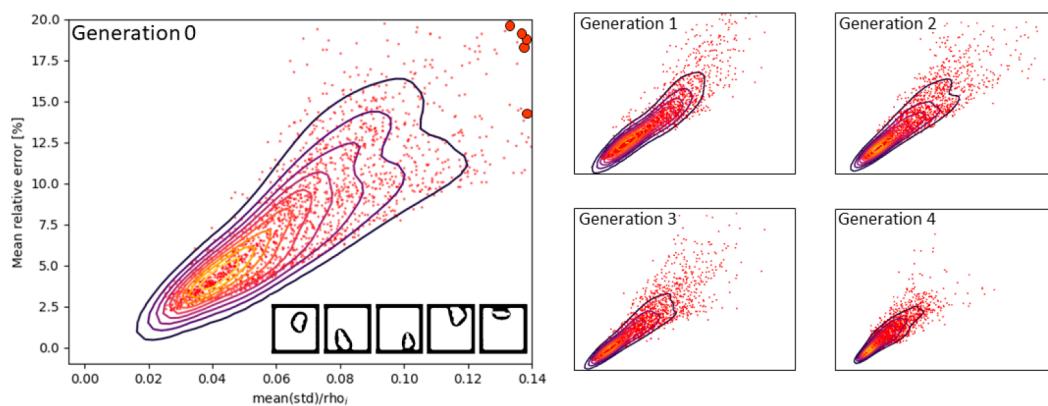


Figure 9. Uncertainty reduction. Mean relative error vs standard deviation of the predictions, where each data point (red dots) corresponds to a simulation case. Each generation depicts an active learning iteration (adding the n -least-certain examples). These n -simulations can be done on the fly by using the AL criterion. In the first plot to the left, we show the five cases (bold red dots) that yield the highest standard deviation of the predictions. The pore shapes of these are shown in the gray squares. The five panels have the same axis limits.

strong relationship between disagreement and model error. The eight unlabeled samples exhibiting the highest value are simulated by using MD, and added to the training set, then the networks are retrained using these. The process is performed repeatedly, yielding generations of models trained to progressively larger data sets. Figure 9, right panels, shows that each progressive generation yields both lower model disagreement and lower model error.

To test the effectiveness of the active learning method, we randomly select 60 pores from the FIB-SEM Vaca Muerta pore data set and run simulations varying the overall gas density and assessed the performance of networks as they selected new data. We tested three different strategies: adding the samples from the pool with the highest mean value (across all the pixels in one image) of standard deviation, adding the samples with the highest maximum value (at a single location), and, for comparison, adding samples randomly, with no active selection strategy.

The results of 90 AL generations (which are defined as every time when more examples are added and the networks retrained) are shown in Figure 10. Dashed lines show comparisons with other methods: linear regression, the ConvNet trained to all of the simple shapes data set, and a ConvNet trained to all of the simple and irregular synthetic shapes (that is, the entire pool of samples available during active learning). From Figure 10, we can see that a targeted

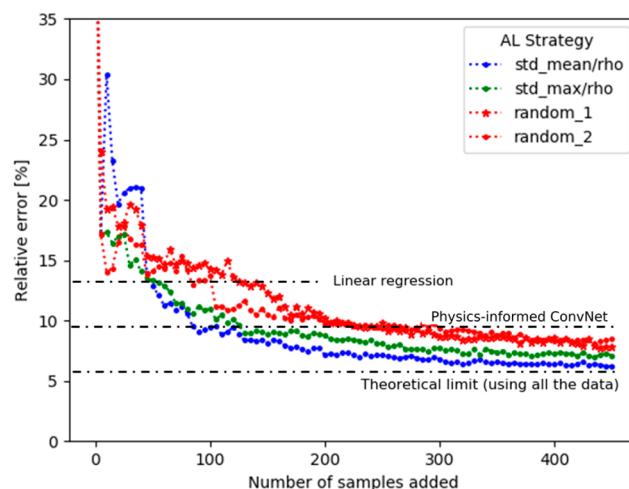


Figure 10. Performance of the different active learning strategies over the Vaca Muerta set. Each AL generation is defined with a marker. Two random sampling runs with different seeds were performed to show the superiority of the other active learning strategies. In black dashed lines, the results of the linear regression and the ConvNet model of section 2.3.3 both trained with 1500 simple shape examples, is shown. The theoretical limit is a network trained with all of the data available (3000 examples of simple and irregular synthetic geometries).

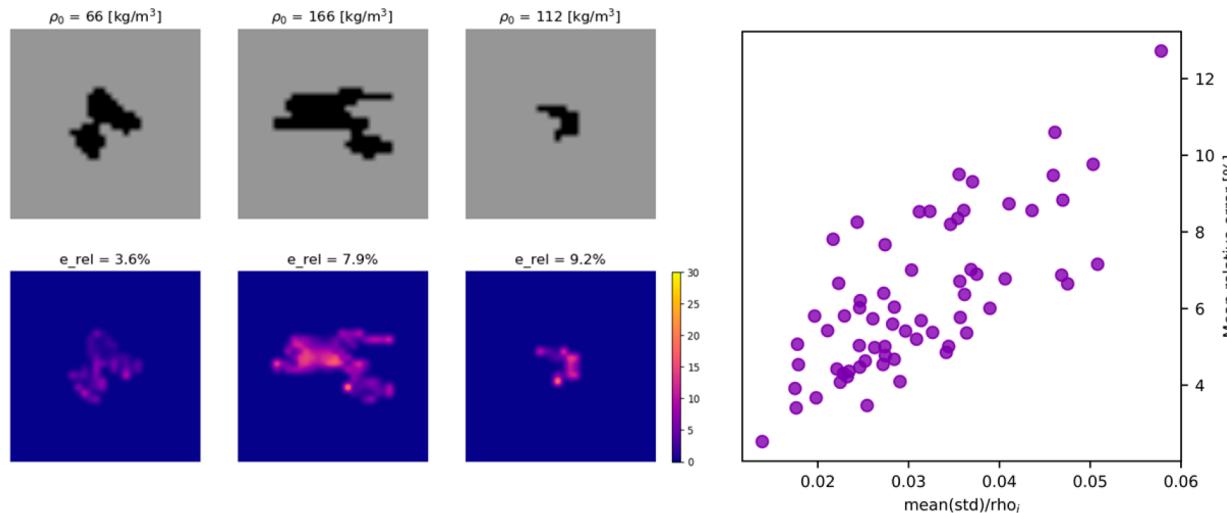


Figure 11. Relative error of the predictions and uncertainty quantification of the Vaca Muerta test set pores using the model trained with AL. To the left, we show a pore with low, medium, and high average relative error. The top panels show the pore geometries, and the bottom panels show the error of the network prediction. The color scheme indicates pixelwise percentage error. To the right, we plot uncertainty vs error on the test set, where each data point corresponds to a domain of the Vaca Muerta FIB-SEM image. There is a strong correlation between uncertainty in the prediction and relative error.

training set containing only 90 examples selected with the AL technique outperforms the simple-shape neural network ConvNet model that is trained 1500 examples. The random sampling approach needs ~ 250 samples, almost 3 times the amount of data, to perform similarly. The curves plateau due to the fact that the pool of candidate samples is limited, but the active selection strategies plateau much more quickly than the random selection strategy. This shows that the AL strategy is more robust and predictive in generalizing to a new, more complex distribution of pore shapes. Figure 11 presents the relative error of the best active learning model for three pores that are examples of low, typical, and high error among the Vaca Muerta pore data set. While it is not visually apparent what geometric features of the pore drive the model errors, the disagreement vs model error is plotted in Figure 11. The strong relationship between these quantities seen in the active learning selection data set (Figure 9) still holds, indicating that when new complex data sets of pore geometries are encountered, the model error can be estimated, and further active learning could be applied.

Multiphase and multicomponent fluid flow through subsurface formations is a challenging research field due to the multiple coupled physical phenomena the affect the fluid behavior. Molecular dynamics simulations offer the ability to capture these complex interactions of fluid and solids at the desired thermodynamically conditions. Nevertheless, the ability to carry out fluid flow simulations of connected porous media at the molecular scale is still a faraway goal due to current computational limitations (physical memory constraints,⁶³ file system limitations,⁶⁴ and accessible simulation time scales⁶⁵). Although there have been some efforts of simulating computationally large domains using parallel software and supercomputer clusters, the number of molecules and the simulation time scales are not sufficient to observe representative behaviors in porous materials. On the other hand, many of these phenomena, for example the adsorption, occur locally. Therefore, the use of a trained ML model is a valid assumption, leads to a significant speed-up, and can be used as a surrogate for simulating larger systems.

In our workflow, we used frozen and a simple united-atom force field. The main difficulty with extending the method to more realistic MD simulations is the construction of realistic pore walls that are stable under MD. Other complexities, such as choice of interaction potential, could be easily changed to analyze a variety of systems. ConvNets are known to be highly flexible models, and by incorporating more input parameters, it may be possible to train a model that simultaneously models different solid surfaces and their affinity with a variety of bulk fluids.

4. CONCLUSIONS

In this paper, we build machine-learning-based surrogate models for predicting the molecular density profile of a gas-filled pore under confinement. We capture adsorption profiles on broad classes of increasingly realistic pore geometries. We present a methodology to select the optimal depth and size of a ConvNet architecture and tune the hyperparameters to carry out physics-based predictions in images. Moreover, we demonstrate additional enhancements to the model to obtain better predictions: by applying physical considerations to the network, we ensure that the predictions we obtain have a physical basis. Finally, we demonstrate an effective way of building data sets with active learning to train robust predictive models efficiently. By using a query-by-committee strategy, we can achieve generality with limited data, only needing simulations to fill targeted areas of the data space where uncertainty is large.

Our results show that active learning for ConvNets can be productively applied to surrogate modeling of a very broad range of pore geometries with relatively few MD calculations. The method that we present can provide predictions of methane density in nanopores in 3 ms (when using five networks) compared to close to an hour with MD, yielding speed-ups factors on the order of 1 million. We show how an active learning strategy gets the same accuracy of a network trained with a comprehensive data set (3000 simulations) using only 15% of this data. By doing this, it is possible to predict the density profile from an SEM image of a specific

formation to enhance the current way of estimating gas reserves.

Adsorption is only one example of the nanoconfinement effects relevant in shale. In this nanoscale regime, all traditional computational fluid dynamics approaches (e.g., LBM) break as a result of being defined in the continuum limit, and more expensive molecular-level calculations must be performed for physics-based modeling. With our workflow, the expense of these molecular calculations can be amortized and prioritized by using a machine learning model that can be rapidly deployed to new pores. A good candidate for future work is to assemble an incredibly broad spectrum of pore geometries from a variety of real pore data and construct an MD workflow utilizing more advanced molecular dynamics techniques.

While having a fast proxy for molecular dynamics in nanopores has value in itself (e.g., for estimating gas reserves), we would like to point toward what we believe to be an important future application of our methods: The use of the surrogate for the parametrization of continuum models of nanoscale porous media—techniques for accomplishing this using machine learning have been recently investigated.⁶⁶ A fast surrogate would enable the upscaling of nanoconfinement effects into, for example, an extended LBM formulation that can treat fluid flow as a coupled problem between many pores of spanning many length scales. While this endeavor comes with many challenges related to workflows and code coupling, such a scale-bridging model would enable a fast computational method to address the role of nanoconfinement in porous media for a variety of applications in engineering, energy sciences, and environmental sciences.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.0c07427>.

Details about the shape factor calculation and the neural network training process ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Javier E. Santos — The Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, Texas 78712, United States;  orcid.org/0000-0002-2404-3975; Email: jesantos@utexas.edu

Authors

Mohammed Mehana — Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States

Hao Wu — Department of Geosciences, Virginia Polytechnic Institute & State University, Blacksburg, Virginia 24061, United States

Maša Prodanović — The Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

Qinjun Kang — Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States

Nicholas Lubbers — Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States

Hari Viswanathan — Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States

Michael J. Pyrcz — The Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcc.0c07427>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded in part by the U.S. Department of Energy through Los Alamos National Laboratory's Laboratory Directed Research and Development program (LANL-LDRD) and the Center for Nonlinear Studies (CNLS) at LANL. Computations were performed in part on the Darwin cluster at LANL, and the authors thank the Darwin team for their support. J.E.S. thanks the Applied Machine Learning (AML) summer program and CNLS at LANL for the mentoring and the fellowship received. We thank the reviewers for their comments towards the improvement of this work.

REFERENCES

- (1) Hughes, J. D. A Reality Check on the Shale Revolution. *Nature* **2013**, *494* (7437), 307–308.
- (2) Gandossi, L. An Overview of Hydraulic Fracturing and Other Formation Stimulation Technologies for Shale Gas Production, 2013.
- (3) Mehana, M.; Callard, J. Reserve Estimation with Unified Production Analysis. *SPE/AAPG/SEG Unconv. Resour. Technol. Conf. 2018, URTC 2018* **2018**, 691–696.
- (4) Valkó, P. P.; Lee, W. J. A Better Way to Forecast Production from Unconventional Gas Wells. *Proc. - SPE Annu. Technol. Conf. Exhib.* **2010**, *3*, 1860–1875.
- (5) McGlade, C.; Speirs, J.; Sorrell, S. Methods of Estimating Shale Gas Resources - Comparison, Evaluation and Implications. *Energy* **2013**, *59*, 116–125.
- (6) Striolo, A.; Cole, D. R. Understanding Shale Gas: Recent Progress and Remaining Challenges. *Energy Fuels* **2017**, *31* (10), 10300–10310.
- (7) Jin, Z.; Firoozabadi, A. Thermodynamic Modeling of Phase Behavior in Shale Media. *SPE J.* **2016**, *21* (1), 190–207.
- (8) Mehana, M.; El-Monier, I. Shale Characteristics Impact on Nuclear Magnetic Resonance (NMR) Fluid Typing Methods and Correlations. *Petroleum* **2016**, *2* (2), 138–147.
- (9) Heller, R.; Zoback, M. Adsorption of Methane and Carbon Dioxide on Gas Shale and Pure Mineral Samples. *J. Unconv. Oil Gas Resour.* **2014**, *8* (C), 14–24.
- (10) Obliger, A.; Ulm, F. J.; Pellenq, R. Impact of Nanoporosity on Hydrocarbon Transport in Shales' Organic Matter. *Nano Lett.* **2018**, *18* (2), 832–837.
- (11) Collell, J.; Galliero, G.; Vermorel, R.; Ungerer, P.; Yianourakou, M.; Montel, F.; Pujol, M. Transport of Multi-component Hydrocarbon Mixtures in Shale Organic Matter by Molecular Simulations. *J. Phys. Chem. C* **2015**, *119* (39), 22587–22595.
- (12) Liu, B.; Qi, C.; Zhao, X.; Teng, G.; Zhao, L.; Zheng, H.; Zhan, K.; Shi, J. Nanoscale Two-Phase Flow of Methane and Water in Shale Inorganic Matrix. *J. Phys. Chem. C* **2018**, *122* (46), 26671–26679.
- (13) Loganathan, N.; Yazaydin, A. O.; Bowers, G. M.; Ngouan-Wakou, B. F.; Kalinichev, A. G.; Kirkpatrick, R. J. Role of Cations in the Methane/Carbon Dioxide Partitioning in Nano- And Mesopores of Illite Using Constant Reservoir Composition Molecular Dynamics Simulation. *J. Phys. Chem. C* **2020**, *124* (4), 2490–2500.

- (14) Sun, H.; Zhao, H.; Qi, N.; Li, Y. Molecular Insights into the Enhanced Shale Gas Recovery by Carbon Dioxide in Kerogen Slit Nanopores. *J. Phys. Chem. C* **2017**, *121* (18), 10233–10241.
- (15) Gubbins, K. E.; Liu, Y. C.; Moore, J. D.; Palmer, J. C. The Role of Molecular Modeling in Confined Systems: Impact and Prospects. *Phys. Chem. Chem. Phys.* **2011**, *13* (1), 58–85.
- (16) Gubbins, K. E.; Long, Y.; Śliwińska-Bartkowiak, M. Thermodynamics of Confined Nano-Phases. *J. Chem. Thermodyn.* **2014**, *74*, 169–183.
- (17) Wang, S.; Javadpour, F.; Feng, Q. Confinement Correction to Mercury Intrusion Capillary Pressure of Shale Nanopores. *Sci. Rep.* **2016**, *6* (1), 1–12.
- (18) Ambrose, R. J.; Hartman, R. C.; Diaz-Campos, M.; Akkutlu, I. Y.; Sondergeld, C. H. Shale Gas-in-Place Calculations Part I: New Pore-Scale Considerations. *SPE J.* **2012**, *17* (1), 219–229.
- (19) Jin, Z.; Firoozabadi, A. Phase Behavior and Flow in Shale Nanopores from Molecular Simulations. *Fluid Phase Equilib.* **2016**, *430*, 156–168.
- (20) Tan, S. P.; Piri, M. Equation-of-State Modeling of Confined-Fluid Phase Equilibria in Nanopores. *Fluid Phase Equilib.* **2015**, *393*, 48–63.
- (21) Dong, X.; Liu, H.; Hou, J.; Wu, K.; Chen, Z. Phase Equilibria of Confined Fluids in Nanopores of Tight and Shale Rocks Considering the Effect of Capillary Pressure and Adsorption Film. *Ind. Eng. Chem. Res.* **2016**, *55* (3), 798–811.
- (22) Pathak, M.; Kweon, H.; Deo, M.; Huang, H. Kerogen Swelling and Confinement: Its Implication on Fluid Thermodynamic Properties in Shales. *Sci. Rep.* **2017**, *7* (1), 1–14.
- (23) Bocquet, L.; Charlaix, E. Nanofluidics, from Bulk to Interfaces. *Chem. Soc. Rev.* **2010**, *39* (3), 1073–1095.
- (24) Mehana, M.; Fahes, M.; Kang, Q.; Viswanathan, H. Molecular Simulation of Double Layer Expansion Mechanism during Low-Salinity Waterflooding. *J. Mol. Liq.* **2020**, *318*, 114079.
- (25) Gruener, S.; Hofmann, T.; Wallacher, D.; Kityk, A. V.; Huber, P. Capillary Rise of Water in Hydrophilic Nanopores. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2009**, *79* (6), 6–9.
- (26) Li, Z. Z.; Min, T.; Kang, Q.; He, Y. L.; Tao, W. Q. Investigation of Methane Adsorption and Its Effect on Gas Transport in Shale Matrix through Microscale and Mesoscale Simulations. *Int. J. Heat Mass Transfer* **2016**, *98*, 675–686.
- (27) Zhong, J.; Zandavi, S. H.; Li, H.; Bao, B.; Persad, A. H.; Mostowfi, F.; Sinton, D. Condensation in One-Dimensional Dead-End Nanochannels. *ACS Nano* **2017**, *11* (1), 304–313.
- (28) Song, W.; Yao, J.; Ma, J.; Li, A.; Li, Y.; Sun, H.; Zhang, L. Grand Canonical Monte Carlo Simulations of Pore Structure Influence on Methane Adsorption in Micro-Porous Carbons with Applications to Coal and Shale Systems. *Fuel* **2018**, *215*, 196–203.
- (29) Ho, T. A.; Wang, Y.; Criscenti, L. J. Chemo-Mechanical Coupling in Kerogen Gas Adsorption/Desorption. *Phys. Chem. Chem. Phys.* **2018**, *20* (18), 12390–12395.
- (30) Jin, Z.; Firoozabadi, A. Methane and Carbon Dioxide Adsorption in Clay-like Slit Pores by Monte Carlo Simulations. *Fluid Phase Equilib.* **2013**, *360*, 456–465.
- (31) Xiong, J.; Liu, K.; Liu, X.; Liang, L.; Zeng, Q. Molecular Simulation of Methane Adsorption in Slit-like Quartz Pores. *RSC Adv.* **2016**, *6* (112), 110808–110819.
- (32) Huang, L.; Ning, Z.; Wang, Q.; Qi, R.; Zeng, Y.; Qin, H.; Ye, H.; Zhang, W. Molecular Simulation of Adsorption Behaviors of Methane, Carbon Dioxide and Their Mixtures on Kerogen: Effect of Kerogen Maturity and Moisture Content. *Fuel* **2018**, *211*, 159–172.
- (33) Chen, L.; Zhang, L.; Kang, Q.; Viswanathan, H. S.; Yao, J.; Tao, W. Nanoscale Simulation of Shale Transport Properties Using the Lattice Boltzmann Method: Permeability and Diffusivity. *Sci. Rep.* **2015**, *5* (1), 8089.
- (34) Xu, R.; Prodanović, M.; Landry, C. J. Study of Subcritical and Supercritical Gas Adsorption Behavior in Different Nanopore Systems in Shale Using Lattice Boltzmann Method. *Int. J. Coal Geol.* **2019**, *212*, 103263.
- (35) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444.
- (36) Santos, J. E.; Xu, D.; Jo, H.; Landry, C. J.; Prodanović, M.; Pyrcz, M. J. PoreFlow-Net: A 3D Convolutional Neural Network to Predict Fluid Flow through Porous Media. *Adv. Water Resour.* **2020**, *138*, 103539.
- (37) Kim, B.; Lee, S.; Kim, J. Inverse Design of Porous Materials Using Artificial Neural Networks. *Sci. Adv.* **2020**, *6* (1), 1–8.
- (38) Chmiela, S.; Sauceda, H. E.; Müller, K. R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, DOI: 10.1038/s41467-018-06169-2.
- (39) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-Gualdrón, D. A. Role of Pore Chemistry and Topology in the CO₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater.* **2018**, *30* (18), 6325–6337.
- (40) Smith, J. S.; Nebgen, B.; Mathew, N.; Chen, J.; Lubbers, N.; Burakovsky, L.; Tretiak, S.; Nam, H. A.; Germann, T.; Fensin, S.; Barros, K. Automated Discovery of a Robust Interatomic Potential for Aluminum. **2020**, 1–11.
- (41) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: 2016.
- (42) Settles, B. Active Learning Literature Survey. *Comput. Sci. Tech. Rep.* **1648**, University Wisconsin—Madison, 2009.
- (43) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148* (24), 241733.
- (44) Dai, C.; Glotzer, S. C. Efficient Phase Diagram Sampling by Active Learning. *J. Phys. Chem. B* **2020**, *124* (7), 1275–1284.
- (45) Seung, H. S.; Opper, M.; Sompolinsky, H. Query by Committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT '92*; ACM Press: New York, 1992; pp 287–294.
- (46) Andrew, M. *Vaca Muerta FIB-SEM*; Digital Rocks Portal, 2019.
- (47) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO—the Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **2010**, *18* (1), 015012.
- (48) Mason, G.; Morrow, N. R. Capillary Behavior of a Perfectly Wetting Liquid in Irregular Triangular Tubes. *J. Colloid Interface Sci.* **1991**, *141* (1), 262–274.
- (49) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (50) Cao, J.; Liang, Y.; Masuda, Y.; Koga, H.; Tanaka, H.; Tamura, K.; Takagi, S.; Matsuoka, T. Molecular Simulation of Methane Adsorption Behavior in Kerogen Nanopores for Shale Gas Resource Assessment. *Int. Pet. Technol. Conf.* **2019**, *IPTC 2019*, 2019.
- (51) Gautam, S.; Le, T.; Striolo, A.; Cole, D. Molecular Dynamics Simulations of Propane in Slit Shaped Silica Nano-Pores: Direct Comparison with Quasielastic Neutron Scattering Experiments. *Phys. Chem. Chem. Phys.* **2017**, *19* (48), 32320–32332.
- (52) Song, W.; Yao, J.; Ma, J.; Li, A.; Li, Y.; Sun, H.; Zhang, L. Grand Canonical Monte Carlo Simulations of Pore Structure Influence on Methane Adsorption in Micro-Porous Carbons with Applications to Coal and Shale Systems. *Fuel* **2018**, *215*, 196–203.
- (53) Cao, J.; Liang, Y.; Masuda, Y.; Koga, H.; Tanaka, H.; Tamura, K.; Takagi, S.; Matsuoka, T. Molecular Simulation of CH₄ Adsorption Behavior in Slit Nanopores: Verification of Simulation Methods and Models. *AICHE J.* **2019**, *65* (11), 1–12.
- (54) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102* (14), 2569–2577.
- (55) Greathouse, J. A.; Hart, D. B.; Bowers, G. M.; Kirkpatrick, R. J.; Cygan, R. T. Molecular Simulation of Structure and Diffusion at Smectite-Water Interfaces: Using Expanded Clay Interlayers as Model Nanopores. *J. Phys. Chem. C* **2015**, *119* (30), 17126–17136.
- (56) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2012**, *60*, 1097–1105.

- (57) Güneş Baydin, A.; Pearlmutter, B. A.; Andreyevich Radul, A.; Mark Siskind, J. Automatic Differentiation in Machine Learning: A Survey. *J. Mach. Learn. Res.* **2018**, *18*, 1–43.
- (58) Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *32nd Int. Conf. Mach. Learn. ICML 2015* **2015**, *1*, 448–456.
- (59) Nair, V.; Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning; ICML'10*; Omnipress: Madison, WI, 2010; pp 807–814.
- (60) Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* **2015**, *1*–14.
- (61) Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2015**, *9351*, 234–241.
- (62) He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2016**, *9908*, 630–645.
- (63) Kadau, K.; Germann, T. C.; Lomdahl, P. S. Large-Scale Molecular-Dynamics Simulation of 19 Billion Particles. *Int. J. Mod. Phys. C* **2004**, *15* (1), 193–201.
- (64) Jung, J.; Nishima, W.; Daniels, M.; Bascom, G.; Kobayashi, C.; Adedoyin, A.; Wall, M.; Lappala, A.; Phillips, D.; Fischer, W.; Tung, C. S.; Schlick, T.; Sugita, Y.; Sanbonmatsu, K. Y. Scaling Molecular Dynamics beyond 100,000 Processor Cores for Large-Scale Biophysical Simulations. *J. Comput. Chem.* **2019**, *40* (21), 1919–1930.
- (65) Klein, M. L.; Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science (Washington, DC, U. S.)* **2008**, *321* (5890), 798–800.
- (66) Lubbers, N.; Agarwal, A.; Chen, Y.; Son, S.; Mehana, M.; Kang, Q.; Karra, S.; Junghans, C.; Germann, T. C.; Viswanathan, H. S. Modeling and Scale-Bridging Using Machine Learning: Nanoconfinement Effects in Porous Media. *Sci. Rep.* **2020**, *10*, 1–13.