UNIVERSITY OF CALIFORNIA
RIVERSIDE

Inverse Design of Nanoporous Materials for Gas Separation and Energy Storage via Physics-Based and Data-Driven Modeling

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Musen Zhou

September 2022

Dissertation Committee:
Dr. Jianzhong Wu, Chairperson
Dr. Juchen Guo
Dr. Bryan M. Wong

The Dissertation of Musen Zhou is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

This dissertation is the outcome of my five exciting and rewarding years' Ph. D. journey. It would not be completed without the support and guidance from my supervisor Professor Jianzhong Wu. He is generous in sharing his wisdom and aspiring ideas, and I enjoyed my many inspiring discussions with him. He is also a role model to me in many aspects, and in particular I learned from him how to think of research and solve problems from the logical and fundamental perspective.

I would like to express my thanks to Prof. Juchen Guo and Prof. Bryan M. Wong for their time serving on my dissertation committee. I'm also grateful to Prof. Pingyun Feng and Prof. Kandis Abdul-Aziz for serving on my qualifying examination committee. Although none of my collaboration work is included in this dissertation, I would like to thank all my collaborators, Dr. Jian Zhang, Dr. Tao Wang, Mr. Hao Li, Prof. Frank Gao and Prof. Yan Qiao, to educate me and open up my eyes of research and scientific problems outside the field of my Ph. D. work. In Wu Group, I'm particularly grateful to my fellow graduate student and friend, Mr. Alejandro Gallegos, for his support and share of wisdom and knowledge in research and life. Special thanks are also given to Mrs. Jingqi Wang, Mr. Anthony Vassallo and Mr. Jose Garcia, and working with them helped me become a better partner and mentor. I would also like to thank many former members in Wu Group for their valuable discussion: Prof. Yun Tian, Dr. Kun Liu, Dr. Xiaoyu Hu, Mr. Gary Ong.

To all my friends and roommates, I'm grateful to your support and companion, and you all delight my life at Riverside.

Last but certainly not least, I would like to express my deepest and most sincere thanks to my parents and family for their unconditional love to support and believe in me to pursue things of my interest and passion. On a personal note, this dissertation is dedicated to my grandfather, who taught me how to face and confront difficulties in life, and I will never forget the time we shared and spent together.

The content of this dissertation, in part of in full, is a reprint of the materials as it appears in *Carbon* (Volume 157, Page 147-152, Feb 2020), *Journal of Membrane Science* (Volume 598, Page 117675, Feb 2020), *ACS Applied Energy Materials* (Volume 3, Issue 6, Page 5993-6000, Jun 2020), *The Journal of Chemical Physics* (Volume 153, Issue 7, Page 074101, Aug 2020), *ACS Applied Nano Materials* (Volume 4, Issue 3, Page 5394-5403, May 2021) and *Industrial & Engineering Chemistry Research* (Volume 60, Issue 47, Page 17016-17025, Dec 2021).

The work completed in this dissertation is financially supported by the U.S. National Science Foundation under Grant No. NSF-CBET-1404046, the U.S. National Science Foundation Harnessing the Data Revolution (HDR) Big Ideas Program under Grant No. NSF 1940118 and the Fluid Interface Reactions, Structures, and Transport (FIRST) Center, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences. Parts of computation in this dissertation used resources of the National Energy Research Science Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy, under Contract DE-AC02-05CH11231. I would also like to

thank the NVIDIA corporation for the donation of NVIDIA A100 GPUs through NVIDIA Academic Hardware Grant Program.

ABSTRACT OF THE DISSERTATION

Inverse Design of Nanoporous Materials for Gas Separation and Energy Storage via Physics-Based and Data-Driven Modeling

by

Musen Zhou

Doctor of Philosophy, Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, September 2022
Dr. Jianzhong Wu, Chairperson

Nanoporous materials (e.g., activated carbon, zeolite and metal-organic framework) have attracted great interest in recent years because of their excellent structural and chemical properties including large specific surface area, good mechanic strength and tailorable pore size and local chemical environment. Such materials are promising to serve as better alternatives to conventional materials such as porous electrodes in energy storage or adsorbents and membranes in separating molecules with similar properties. Because of the almost infinite design space, not only is the identification of best nanoporous materials with target performance practically infeasible with traditional experimental trail-and-error methods, but also it imposes theoretical and computational challenges for the computational modeling of nanoporous materials in gas separation and energy storage.

To accomplish the inverse design of nanoporous materials in gas separation and energy storage, this dissertation aims to establish physics-based and data-driven models that can be used to fast and accurately evaluate the performance of nanoporous materials. Toward that end, I developed classical density functional theory (cDFT) to predict the adsorption of multicomponent gas mixtures in nanoporous materials. The adsorption isotherms predicted by cDFT were in excellent agreement with grand canonical Monte Carlo simulation and experimental measurement. In addition, I extended the simplified string method to calculate the minimum energy path (MEP) of rigid polyatomic molecules in nanoporous materials. The diffusion coefficients predicted from MEP via the transition-state theory agreed quantitatively well with those from molecular dynamics simulation. Furthermore, I implemented the physics-based models with massively parallel GPU-acceleration, which leaded to orders of magnitude speedup compared with conventional molecular simulation. Moreover, I combined the data-driven models and evolutionary algorithm with physics-based models to case study the inverse design of nanoporous materials for the separation of D_2/H_2 and of C_2H_4/C_2H_6. For energy storage, I established excellent correlations between the structural and chemical features of nanoporous materials and their *in-operando* electrochemical performance in supercapacitors using data-driven models and proposed useful guidelines for the inverse design. The computational framework developed in this dissertation demonstrated the feasibility for the inverse design of nanoporous materials for gas separation and energy storage with the combination of physics-based and data-driven models.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Scope of Research

The recent development and advancement of nanoporous materials have attracted tremendous research interest in the applications of gas storage, separation, drug delivery and energy storage.[1-3] While experimental approach is capable of discovering new nanoporous materials with distinct behavior performance such as negative thermal expansion and reversible water adsorption at low pressure,[4-6] it is practically infeasible to experimentally trail-and-error all the nanoporous materials in the almost infinite chemical space of nanoporous materials to find the best candidates with target performance. Whereas computational methods not only can provide atomistic details to understand and design materials for the applications of interest,[4, 7] but can also provide fast and accurate evaluations of materials to identify best candidates in the large structural databases, via high-throughput screening, before experimental synthesis and testing.[8-10] Compared with forward process that identify the best candidates based on given structural database via physics-based models, the backward procedure, viz., inverse design, is more of interest but also much more difficult at the same time.[11-13] With inverse design, the recognition of excellent material candidates can be much faster and more efficient than high-throughput screening in a large chemical space. However, to accomplish the inverse design, not only performance properties of materials need to be evaluated fast and accurately, but also the mapping from the performance properties to molecular structure is needed.

This dissertation is focused on the development of new physics-based models and its integration with data-driven models to achieve the inverse design of nanoporous materials for gas separation and energy storage. For gas separation, we are in particular interested in adsorption-based and membrane-based separation process because they are much more energy efficient than conventional separation process such as distillation separation and have been widely used in industrial separation.[14] In adsorption and membrane separation, the adsorption amount and diffusivity coefficient are the most essential and important properties to evaluate the performance of nanoporous materials.[15-18] Whereas most previous work are only concerned with using conventional grand canonical Monte Carlo (GCMC) simulation and molecular dynamics (MD) simulation to calculate those properties,[10, 19, 20] one key contribution of this work is the development of GPU-accelerated physics-based theoretical models which predict adsorption amount and diffusivity coefficient with similar accuracy and orders of magnitude speedup compared with conventional GCMC and MD simulation. The excellent accuracy and significant speedup of developed physics-based model empowers the inverse the design of nanoporous materials, especially for membrane separation, with the target performance. Although the developed physics-based models are generally applicable to all kinds of nanoporous materials, the discussion of this dissertation is mostly based on one particular type of nanoporous materials – metal-organic framework (MOF). For energy storage, this dissertation is focused on the supercapacitor due to its better power density compared to lithium ion battery and higher energy density compared to conventional capacitor.[21-23] While most previous work used physics-based models

to investigate the capacitive behavior, redox reaction and quantum capacitance at the electrode-electrolyte interface in supercapacitors, those physics-based models are mostly only concerned with equilibrium properties due to the limitations in the time and length scale.[24-26] Another key contribution of this dissertation is that, for the first time, we demonstrated data-driven models can be used to understand and design the structural and chemical features of electrode materials in supercapacitors to optimize the *in-operando* performance that is far away from equilibrium. The combination of physics-based and data-driven models demonstrated in this dissertation opens up opportunity to customize nanoporous materials for gas separation and energy storage system of academic and industrial interest.

1.2 Recent Progress in Molecular Modeling and Inverse Design

Many attempts have been made in recent years to achieve faster and more accurate modeling of modeling system via physics-based and data-driven modeling. While physics-based models are developed to consider more complicated molecular systems, data-driven models are often used to correlate the features with the physical properties of interest and give fast prediction. The databases of molecular properties built by either physics-based or data-driven models allow the exploration of a much larger chemical space to inverse design molecules with target performance.

1.2.1 Physics-Based Molecular Modeling

For gas separation, physics-based models are often used to evaluate the performance, such as adsorption amount and diffusivity coefficient, of nanoporous materials. In terms of adsorption amount, open source software helps grand canonical

Monte Carlo (GCMC) simulation now become a standard tool to study the sorption properties for specific material of interest or high-throughput screening. Despite the widespread application of GCMC simulation, the theoretical framework of GCMC simulation is rather complete[27-30] and most recent advancement of GCMC simulation is focused on the extension of developed sampling method to the general applications. For example, Witman et al. demonstrated that 'flat-histogram' sampling in combination with temperature extrapolation of the free energy landscape can be used to efficiently provide useful thermodynamic properties in adsorption,[31] while Datar et al. found 'flat-histogram' sampling can also help better predict water adsorption isotherm in nanoporous materials.[32] The recent advancement in GCMC simulation has also leveraged the progress in computational hardware such as graphic processing unit (GPU). Many open source GCMC simulation software has incorporated the GPU-accelerated implementation while the speedup factor can be up to 30 compared with the conventional serial computing.[33, 34]

Similar to GCMC simulation in predicting sorption properties, molecular dynamics simulation has become a computational routine in predicting the transport properties such as diffusion coefficient. In MD simulation, the diffusion coefficient can be calculated from the molecular mean square displacement according to Einstein equation

$$D = \frac{1}{n}\lim_{t\to\infty}\frac{\left|\mathbf{r}(t)-\mathbf{r}(0)\right|^2}{2t} \tag{1.1}$$

where n represents the number of dimensions, $\mathbf{r}$ stands for the cartesian coordinate and t

is the simulation time. One drawback of predicting diffusion coefficient with brute-force MD simulation is, in order to have reasonably accurate prediction of diffusion coefficient, the computational cost is directly related to the scale of diffusion coefficient. For slow diffusion at the scale of 1×10^{-15} m^2/s, it is computationally prohibitive for conventional MD simulation to predict the diffusion coefficient. Although the incorporation of GPU-accelerated algorithm in many MD simulation software makes it possible to consider relatively slow diffusion ($\sim 1\times10^{-12}$ m^2/s) with acceptable computational cost,[35-37] it is still computationally challenging for MD simulation to high-throughput screen and construct large database of the transport properties of gas molecules in nanoporous materials. To overcome the difficulty in the computational inefficiency of MD simulation, transition path sampling was introduced to MD simulation by Boulfelfel et al. to predict slow diffusion in zeolites[38] and Verploegh et al. applied transition-state theory (TST) along with umbrella sampling and weighted histogram analysis method in MD simulation to predict the slow diffusion of hydrocarbons in zeolite.[39] It is worth mentioning that these methods are applicable not only in rigid nanoporous materials, but flexible nanoporous materials as well because of the versatile computational ensemble in MD simulation. In addition to MD simulations, other computational methods have also been attempted to provide fast and accurate prediction of diffusion coefficient in nanoporous materials. While Tian et al. combined TST and simplified string method to efficiently predict the gas diffusion coefficients in nanoporous materials,[40] Mace et al. used TST along with tunnels and transition states search algorithm to predict the diffusion coefficient of simple gas molecules.[41] Although these methods are promising in the

high-throughput screening of nanoporous materials for gas separation and construction of large database of diffusion coefficient, rigorous tests are still needed in their applicability for nanoporous materials with various topologies and physiochemical environments.

Aside from molecular simulation methods, theoretical methods such as density functional theory (DFT) have also been widely used to investigate nanoporous materials for gas storage and separation.[42, 43] While quantum DFT (qDFT) is often used to investigate the open-metal site in metal-organic framework (MOF) and provide a more accurate force field, classical DFT (cDFT) is often used to evaluate the sorption and thermodynamic properties of gas molecules in nanoporous materials. For example, Kulkarni and Sholl used quantum DFT derived force field to study the adsorption of C_2 and C_3 olefins and paraffins in CuBTC where open-metal site exists,[44] Jia et al. used cDFT to high-throughput screen nanoporous materials for the potential candidates in hydrogen storage,[45] Yun et al. used to entropy calculated from cDFT in combination with excess entropy scaling to predict the diffusion coefficient of gas molecules in nanoporous materials at finite-loading.[46] Compared with conventional molecular simulation methods, quantum DFT are more accurate but more computationally expensive because it does not rely on the empirical force field but the electron distribution instead, while cDFT is much more computationally efficient with similar accuracy and contains much more thermodynamic information.

1.2.2 Data-Driven Molecular Modeling

While physics-based modeling has demonstrated its capability to understand the physical and chemical mechanism of specific nanoporous materials and even high-

throughput screen large structural database for gas separation and energy storage, it is still practically infeasible for physics-based models to explore the almost infinite structural and chemical space of nanoporous materials. Whereas data-driven modeling is able to not only offer insights on the design of nanoporous materials with excellent performance, but also rapidly predict the performance and inverse design nanoporous materials.[47, 48] In general, data-driven modelling can be used for either regression or classification. While regression models can provide quantitative prediction of the materials' properties, classification models can help find the common features of materials with top performance.

For gas separation, different data-driven models and molecular fingerprints have been used to approach the accurate prediction of sorption and diffusion properties of gas molecules in nanoporous materials. Intuitively, structural features of nanoporous materials (e.g., pore size, surface area, density and void fraction) and molecular properties of gas molecules (e.g., size and energy parameter) can be used to predict the adsorption amount of simple gas molecules.[49, 50] Alternatively, the crystal structure of nanoporous material have been converted into the SMILES text string and used as an input to include more information of nanoporous materials for the prediction of gas adsorption.[51] In addition, more detailed information such as energy histogram that describes gas-materials interaction at infinite dilution was also used to predict the adsorption amount at finite loading.[52] While the better representation/input of molecular systems helps improve the prediction accuracy of data-driven models such as deep neural networks, the choice of data-driven model would lead to different design

insights from the extensive data generated from physics-based modeling. For example, random forest model not only can correlate the sorption and transport properties with the structural features of nanoporous materials, it can also specify the exact structural criteria for the nanoporous materials to achieve the excellent separation performance.[53] Aside from the applications of conventional data-driven models in the gas storage and separation, advanced data-driven models have also been used for gas storage and separation in nanoporous materials. Sun et al. attempted to use the autoencoder to represent the comprehensive adsorption space of hydrogen adsorption which allows the interpolation/prediction of hydrogen adsorption in nanoporous materials.[54] Although the recent development in data-driven models has made them promising alternatives to physics-based models for evaluating nanoporous materials in gas separation, there are still challenges awaiting for the practical applications of data-driven models for the nanoporous materials in gas separation. While the performance of most nanoporous materials can be predicted by the data-driven models with the acceptable accuracy, some nanoporous materials, especially those with exceptional performance, still suffer a significant underestimate from data-driven models, which makes the application of data-driven models less compelling in the discovery of best nanoporous materials for gas separation. Besides, the transferability of trained data-driven models is still limited, which hampers their applications in the systems involved new gas molecules and nanoporous materials.

In energy storage, data-driven models have also been used to guide the design of electrode materials, electrolytes and fast-charging algorithm.[55, 56] Similar to gas

separation, data-driven models are used in energy storage to correlate the structural and chemical features of electrode and electrolyte with their electrochemical performance. For example, Joshi et al. applied deep neural network to correlate the chemical properties and properties derived from element constituents in crystal structures with their electrode voltage in metal-ion battery,[57], Zhang and Xu showed that Gaussian process regression can be used to well correlate the molecular structure of electrolyte additive with the redox potential of electrolyte,[58] and Dave et al. optimize the composition of salt in aqueous electrolyte for a better stability via Bayesian optimization and Robotic experiments.[59] Different from gas separation, data-driven models, especially machine learning potentials, have also been used in energy storage to accelerate physics-based modeling and provide a better understanding of the microscopic mechanism, which ultimately leads to the optimal design of electrode and electrolyte. While Houchins and Viswanathan used machine learning potentials to accelerate the evaluation of open-circuit voltage and optimize the cathode composition,[60] Hajibabaei and Kim investigated the diffusivity of lithium ion in hundreds of ternary crystals as solid state electrolytes with sparse Gaussian process regression (sGPR)-accelerated *ab initio* molecular dynamics simulation.[61] In addition, leveraging the vast amount of battery testing data, Severson et al. and Attia et al. showed data-driven methods can be used to make early predictions of failure for fast-charged batteries and optimize the fast-charging algorithm to balance the charging rate and lifespan of batteries.[62, 63] Although the direct mapping from the microscopic structure of electrode and electrolyte to the electrochemical performance of energy storage device is still missing in data-driven models, the current advancement in data-

driven models has significantly improved the understanding and accelerated design of electrode, electrolyte and charging protocols for energy storage device.

1.2.3 Inverse Design

While conventional material design and discover is a forward process where materials are modified to achieve target performance after initial synthesis and evaluation, the backward process, a.k.a. inverse design, that design the structure based on the target performance is of paramount interest from academic and industrial perspective because it not only provides significant insights on the structure-property correlation, but also significantly cuts the time and financial cost to find the materials with better target performance.[11, 12, 64]

The recent advancement of inverse design in materials science is often inspired by the generative models developed in computer vision.[65-67] For example, generative adversarial network (GAN) was initially developed to generate ‘fake’ photos.[65] However, its application has also now been extended to generate and inverse design nanoporous material in gas storage. While Lee et al. first demonstrated GAN can be used generate energy landscape in zeolite,[68] Kim et al. further showed GAN framework is also capable of being extended to directly design the zeolite for the optimal methane storage.[69] However, unfortunately, the computationally complexity of GAN significantly increases and it even fails to converge for complicated crystal materials such as MOFs with the increasing number of elements considered in the crystal which was not a big issue for its original application in computer vision because photos are usually only made from red, green and blue (RGB). Whereas variational autoencoder (VAE) can well

accommodate the complex topology and molecular structure of the secondary building blocks (SBUs) by compressing the MOF structure into a text string and projecting it into the latent space.[70] However, VAE requires accurate projection (*viz*., encode and decoder) between the crystal structure and a latent space, and the VAE training would become infeasible when a vast number of secondary building blocks are considered for the MOF design. Alternatively, evolutionary algorithms, such as the genetic algorithm, are promising for the inverse design of complicated nanoporous materials such as MOFs because they can accommodate not only a large number of SBUs for MOF design, but also find the solution in a nonlinear space consisted of the material topology and SBUs, which can be used to *de novo* design nanoporous materials for gas separation and energy storage.[71, 72] Aside from evolutionary algorithm, end-to-end differentiable neural networks have also shown its capability to *de novo* inverse design nanoporous materials with target performance with proper descriptors to represent nanoporous materials.

To sum up, in order to accomplish the inverse design of nanoporous materials for gas separation and energy storage, not only physics-based and data-driven models need to be developed to provide fast and accurate evaluations for the properties of interest, but also the deliberate integration of physics-based/data-driven models and generative/searching models is required to identify the structure of nanoporous materials with target performance in gas separation and energy storage.

1.3 Dissertation Organization

This dissertation is focused on the development of physics-based models and utilization of data-driven models to fast and accurately evaluate the performance of

nanoporous materials in gas separation and energy storage, and ultimately the inverse design nanoporous materials with target performance. There are in total 10 Chapters included in this dissertation.

Following the introduction in Chapter 1, Chapter 2 describes the basic concepts and formulations of physics-based used in this dissertation. The emphasis is given to the construction of both classical density functional theory (cDFT) formulation, especially for the multicomponent gas mixture system, and transition-state theory (TST) with simplified string method for the simple gas molecules and rigid polyatomic gas molecules.

Chapter 3 demonstrates an efficient implementation of graphic processing unit (GPU)-accelerated 3D cDFT calculation for gas adsorption. Compared with serial central processing unit (CPU) implementation, the GPU-accelerated implementation reduces the computational cost of cDFT by more than two orders of magnitude for variety types of calculation related in gas adsorption. The GPU-accelerated implementation of 3D cDFT for gas adsorption in this work paves the road to inverse design nanoporous materials for gas adsorption by providing an efficient computational tool to construct large databases of adsorption isotherm with high fidelity.

Chapter 4 extends the cDFT formulations to accurately predict multicomponent gas mixture adsorption in nanoporous materials. Compared with alternative methods such as grand canonical Monte Carlo (GCMC) simulation, ideal adsorbed solution theory (IAST) and experimental measurements, the developed cDFT formulations give excellent prediction of adsorption isotherms for multicomponent gas mixtures over a wide range of pressure. This work provides the theoretical basis for the computational design of

nanoporous materials in adsorption-based separation process as well as for screening and data-driven inverse design of nanoporous materials.

Chapter 5 presents a fast and accurate computational method to predict the diffusivity coefficient of rigid polyatomic gas molecules in nanoporous materials. In this work, the simplified string method is extended to calculate the minimum energy path of rigid polyatomic molecules in nanoporous materials with the massively parallel GPU-accelerated implementation. The diffusion coefficients predicted by TST with the found minimum energy path agree well with that calculated from MD simulation while reducing the computational cost by several orders of magnitude. This work, therefore, opens up opportunities for high-throughput screening and inverse design of nanoporous materials in the applications that diffusion coefficient is of interest.

Chapter 6 demonstrates the computational workflow via developed physics-based models to identify the ideal material candidates and provide design insights of nanoporous materials for the separation of simple gas molecules, more specifically hydrogen isotopes. Compared with state-of-art literature, top MOFs identified in this work via high-throughput screening with physics-based models have a much higher separation selectivity at higher temperature. In addition, the membrane performance score is introduced to identify top ranked MOF membranes with the best selectivity and permeability. With the extensive data generated from physics-based modeling, this work also showcases how physics-based and data-driven models can be integrated to provide insights for the inverse design of nanoporous materials in gas separation.

Chapter 7 presents the computational framework using the developed physics-based models to high-throughput screen and inverse design nanoporous materials for the separation rigid polyatomic gas molecules. In this work, the separation of C_2H_4/C_2H_6 is used as the case study because of its paramount importance in the petrochemical industry. While high-throughput screening of the computational-ready, experimental (CoRE) MOF database leads to materials with exceptionally high ethane-selective adsorption selectivity and ethene-selective membrane selectivity, the inverse design enables the exploration of a broader chemical space and identification of MOFs with even higher membrane selectivity and permeability. In addition, relative membrane performance score (rMPS) is formulated to evaluate the overall membrane performance relative to the Robeson boundary. The computational framework illustrated in this work is generically applicable to materials discovery for gas storage and separation.

Chapter 8 and Chapter 9 examine the idea of using data-driven models to interpret the relationship between structural and chemical features of activated carbon electrodes and their *in-operando* performance in supercapacitors. Not only quantitative correlations can be found, via machine learning models, between the pore structures and surface chemistry of activated carbon electrodes and their experimental *in-operando* behavior in supercapacitors, but also important characteristics of activated carbon electrodes can be identified and used to optimize their efficiency in supercapacitor. This work demonstrates that machine learning models are promising alternatives to understand and design electrode materials with better performance in energy storage devices.

Finally, Chapter 10 summarizes the key conclusions from this dissertation and provides perspectives for future work.

Bibliography

1. Lawson, H.D., S.P. Walton, and C. Chan, *Metal–organic frameworks for drug delivery: a design perspective.* ACS applied materials & interfaces, 2021. **13**(6): p. 7004-7020.

2. Rangnekar, N., et al., *Zeolite membranes - a review and comparison with MOFs.* Chem Soc Rev, 2015. **44**(20): p. 7128-54.

3. Li, J.R., J. Sculley, and H.C. Zhou, *Metal-organic frameworks for separations.* Chemical Reviews (Washington, D. C.), 2012. **112**(2): p. 869-932.

4. Lock, N., et al., *Elucidating negative thermal expansion in MOF-5.* The Journal of Physical Chemistry C, 2010. **114**(39): p. 16181-16186.

5. Kim, H., et al., *Water harvesting from air with metal-organic frameworks powered by natural sunlight.* Science, 2017. **356**(6336): p. 430-434.

6. Xu, W. and O.M. Yaghi, *Metal–organic frameworks for water harvesting from air, anywhere, anytime.* ACS central science, 2020. **6**(8): p. 1348-1354.

7. Dubbeldam, D., et al., *Exceptional negative thermal expansion in isoreticular metal–organic frameworks.* Angewandte Chemie, 2007. **119**(24): p. 4580-4583.

8. Liu, Y., J. Fu, and J. Wu, *High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory.* The Journal of Physical Chemistry Letters, 2013. **4**(21): p. 3687-3691.

9. Fu, J., et al., *Density Functional Methods for Fast Screening of Metal Organic Frameworks for Hydrogen Storage.* J Phys Chem C, 2015. **119**(10): p. 5374-5385.

10. Tang, H. and J. Jiang, *In silico screening and design strategies of ethane-selective metal–organic frameworks for ethane/ethylene separation.* AIChE Journal, 2021. **67**(3): p. e17025.

11. Bhowmik, A., et al., *A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning.* Energy Storage Materials, 2019. **21**: p. 446-456.

12. Zunger, A., *Inverse design in search of materials with target functionalities.* Nature Reviews Chemistry, 2018. **2**(4): p. 1-16.

13. Lee, S., et al., *Computational Screening of Trillions of Metal–Organic Frameworks for High-Performance Methane Storage.* ACS Applied Materials & Interfaces, 2021. **13**(20): p. 23647-23654.

14. Sholl, D.S. and R.P. Lively, *Seven chemical separations to change the world.* Nature, 2016. **532**(7600): p. 435-7.

15. Wijmans, J.G. and R.W. Baker, *The solution-diffusion model: a review.* Journal of membrane science, 1995. **107**(1-2): p. 1-21.

16. Leperi, K.T., et al., *Development of a general evaluation metric for rapid screening of adsorbent materials for postcombustion CO2 capture.* ACS Sustainable Chemistry & Engineering, 2019. **7**(13): p. 11529-11539.

17. Krishna, R., *Metrics for evaluation and screening of metal–organic frameworks for applications in mixture separations.* ACS omega, 2020. **5**(28): p. 16987-17004.

18. Krishna, R., *Screening metal–organic frameworks for mixture separations in fixed-bed adsorbers using a combined selectivity/capacity metric.* RSC advances, 2017. **7**(57): p. 35724-35737.

19. Rungta, M., et al., *Membrane-based ethylene/ethane separation: The upper bound and beyond.* AIChE Journal, 2013. **59**(9): p. 3475-3489.

20. Kang, M., et al., *High-Throughput Discovery of Ni (IN) 2 for Ethane/Ethylene Separation.* Advanced Science, 2021. **8**(11): p. 2004940.

21. Mathis, T.S., et al., *Energy Storage Data Reporting in Perspective—Guidelines for Interpreting the Performance of Electrochemical Energy Storage Systems.* Advanced Energy Materials, 2019. **9**(39).

22. Pomerantseva, E., et al., *Energy storage: The future enabled by nanomaterials.* Science, 2019. **366**(6468): p. eaan8285.

23. Gür, T.M., *Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage.* Energy & Environmental Science, 2018. **11**(10): p. 2696-2767.

24. Wang, L., et al., *Origin of theoretical pseudocapacitance of two-dimensional supercapacitor electrodes Ti3C2T2 (T = bare, O, S).* Journal of Materials Chemistry A, 2019. **7**(27): p. 16231-16238.

25. Zhan, C., et al., *Enhancing graphene capacitance by nitrogen: effects of doping configuration and concentration.* Phys Chem Chem Phys, 2016. **18**(6): p. 4668-74.

26. Beguin, F., et al., *Carbons and electrolytes for advanced supercapacitors.* Adv Mater, 2014. **26**(14): p. 2219-51, 2283.

27. Rosenbluth, M.N. and A.W. Rosenbluth, *Monte Carlo calculation of the average extension of molecular chains.* The Journal of Chemical Physics, 1955. **23**(2): p. 356-359.

28. Siepmann, J.I. and D. Frenkel, *Configurational bias Monte Carlo: a new sampling scheme for flexible chains.* Molecular Physics, 1992. **75**(1): p. 59-70.

29. Frenkel, D., G. Mooij, and B. Smit, *Novel scheme to study structural and thermal properties of continuously deformable molecules.* Journal of Physics: Condensed Matter, 1992. **4**(12): p. 3053.

30. Shi, W. and E.J. Maginn, *Continuous fractional component Monte Carlo: an adaptive biasing method for open system atomistic simulations.* Journal of chemical theory and computation, 2007. **3**(4): p. 1451-1463.

31. Witman, M., N.A. Mahynski, and B. Smit, *Flat-histogram Monte Carlo as an efficient tool to evaluate adsorption processes involving rigid and deformable molecules.* Journal of chemical theory and computation, 2018. **14**(12): p. 6149-6158.

32. Datar, A., M. Witman, and L.-C. Lin, *Improving computational assessment of porous materials for water adsorption applications via flat histogram methods.* The Journal of Physical Chemistry C, 2021. **125**(7): p. 4253-4266.

33. Nejahi, Y., et al., *GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids.* SoftwareX, 2019. **9**: p. 20-27.

34. Barhaghi, M.S., et al., *py-MCMD: Python Software for Performing Hybrid Monte Carlo/Molecular Dynamics Simulations with GOMC and NAMD.* Journal of Chemical Theory and Computation, 2022.

35. Lee, T.-S., et al., *GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features.* Journal of chemical information and modeling, 2018. **58**(10): p. 2043-2050.

36. Zhu, Y.L., et al., *GALAMOST: GPU-accelerated large-scale molecular simulation toolkit.* J Comput Chem, 2013. **34**(25): p. 2197-211.

37. Phillips, J.C., et al., *Scalable molecular dynamics on CPU and GPU architectures with NAMD.* The Journal of chemical physics, 2020. **153**(4): p. 044130.

38. Boulfelfel, S.E., P.I. Ravikovitch, and D.S. Sholl, *Modeling diffusion of linear hydrocarbons in silica zeolite LTA using transition path sampling.* The Journal of Physical Chemistry C, 2015. **119**(27): p. 15643-15653.

39. Verploegh, R.J., S. Nair, and D.S. Sholl, *Temperature and loading-dependent diffusion of light hydrocarbons in ZIF-8 as predicted through fully flexible molecular simulations.* Journal of the American Chemical Society, 2015. **137**(50): p. 15760-15771.

40. Tian, Y., X. Xu, and J. Wu, *Thermodynamic Route to Efficient Prediction of Gas Diffusivity in Nanoporous Materials.* Langmuir, 2017. **33**(42): p. 11797-11803.

41. Mace, A., S. Barthel, and B. Smit, *Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field.* J Chem Theory Comput, 2019. **15**(4): p. 2127-2141.

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

43. Banerjee, D., et al., *Xenon gas separation and storage using metal-organic frameworks.* Chem, 2018. **4**(3): p. 466-494.

44. Kulkarni, A.R. and D.S. Sholl, *Screening of copper open metal site MOFs for olefin/paraffin separations using DFT-derived force fields.* The Journal of Physical Chemistry C, 2016. **120**(40): p. 23044-23054.

45. Fu, J., Y. Tian, and J.Z. Wu, *Classical density functional theory for methane adsorption in metal-organic framework materials.* AIChE J, 2015. **61**(9): p. 3012-3021.

46. Liu, Y., J. Fu, and J. Wu, *Excess-entropy scaling for gas diffusivity in nanoporous materials.* Langmuir, 2013. **29**(42): p. 12997-3002.

47. Jablonka, K.M., et al., *Big-data science in porous materials: materials genomics and machine learning.* Chemical reviews, 2020. **120**(16): p. 8066-8129.

48. Moosavi, S.M., K.M. Jablonka, and B. Smit, *The role of machine learning in the understanding and design of materials.* Journal of the American Chemical Society, 2020. **142**(48): p. 20273-20287.

49. Anderson, R., A. Biong, and D.A. Gómez-Gualdrón, *Adsorption isotherm predictions for multiple molecules in MOFs using the same deep learning model.* Journal of chemical theory and computation, 2020. **16**(2): p. 1271-1283.

50. Anderson, R. and D.A. Gómez-Gualdrón, *Deep learning combined with IAST to screen thermodynamically feasible MOFs for adsorption-based separation of multiple binary mixtures.* The Journal of Chemical Physics, 2021. **154**(23): p. 234102.

51. Gurnani, R., et al., *Interpretable Machine Learning-Based Predictions of Methane Uptake Isotherms in Metal–Organic Frameworks.* Chemistry of Materials, 2021. **33**(10): p. 3543-3552.

52. Li, Z., et al., *Machine learning using host/guest energy histograms to predict adsorption in metal–organic frameworks: Application to short alkanes and Xe/Kr mixtures.* The Journal of Chemical Physics, 2021. **155**(1): p. 014701.

53. Bai, X., et al., *Machine-Learning-Assisted High-Throughput computational screening of Metal–Organic framework membranes for hydrogen separation.* Chemical Engineering Journal, 2022. **446**: p. 136783.

54. Sun, Y., et al., *Fingerprinting diverse nanoporous materials for optimal hydrogen storage conditions using meta-learning.* Science Advances, 2021. **7**(30): p. eabg3983.

55. Barrett, D.H. and A. Haruna, *Artificial intelligence and machine learning for targeted energy storage solutions.* Current Opinion in Electrochemistry, 2020. **21**: p. 160-166.

56. Aykol, M., P. Herring, and A. Anapolsky, *Machine learning for continuous innovation in battery technologies.* Nature Reviews Materials, 2020. **5**(10): p. 725-727.

57. Joshi, R.P., et al., *Machine learning the voltage of electrode materials in metal-ion batteries.* ACS applied materials & interfaces, 2019. **11**(20): p. 18494-18503.

58. Zhang, Y. and X. Xu, *Machine learning properties of electrolyte additives: A focus on redox potentials.* Industrial & Engineering Chemistry Research, 2020. **60**(1): p. 343-354.

59. Dave, A., et al., *Autonomous discovery of battery electrolytes with robotic experimentation and machine learning.* Cell Reports Physical Science, 2020. **1**(12): p. 100264.

60. Houchins, G. and V. Viswanathan, *An accurate machine-learning calculator for optimization of Li-ion battery cathodes.* The Journal of Chemical Physics, 2020. **153**(5): p. 054124.

61. Hajibabaei, A. and K.S. Kim, *Universal machine learning interatomic potentials: surveying solid electrolytes.* The Journal of Physical Chemistry Letters, 2021. **12**(33): p. 8115-8120.

62. Attia, P.M., et al., *Closed-loop optimization of fast-charging protocols for batteries with machine learning.* Nature, 2020. **578**(7795): p. 397-402.

63. Severson, K.A., et al., *Data-driven prediction of battery cycle life before capacity degradation.* Nature Energy, 2019. **4**(5): p. 383-391.

64. Molesky, S., et al., *Inverse design in nanophotonics.* Nature Photonics, 2018. **12**(11): p. 659-670.

65. Creswell, A., et al., *Generative adversarial networks: An overview.* IEEE signal processing magazine, 2018. **35**(1): p. 53-65.

66. Pu, Y., et al., *Variational autoencoder for deep learning of images, labels and captions.* Advances in neural information processing systems, 2016. **29**.

67. Bagal, V., et al., *Molgpt: Molecular generation using a transformer-decoder model.* Journal of Chemical Information and Modeling, 2021. **62**(9): p. 2064-2076.

68. Lee, S., B. Kim, and J. Kim, *Predicting performance limits of methane gas storage in zeolites with an artificial neural network.* Journal of Materials Chemistry A, 2019. **7**(6): p. 2709-2716.

69. Kim, B., S. Lee, and J. Kim, *Inverse design of porous materials using artificial neural networks.* Sci Adv, 2020. **6**(1): p. eaax9324.

70. Yao, Z.P., et al., *Inverse design of nanoporous crystalline reticular materials with deep generative models.* Nature Machine Intelligence, 2021. **3**(1): p. 76-86.

71. Gustafson, J.A. and C.E. Wilmer, *Intelligent selection of metal–organic framework arrays for methane sensing via genetic algorithms.* ACS sensors, 2019. **4**(6): p. 1586-1593.

72. Chung, Y.G., et al., *In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm.* Science advances, 2016. **2**(10): p. e1600909.

Chapter 2. Basic Formulations of Physics-Based Models

In this dissertation, physics-based models are developed to provide fast and accurate evaluation of molecular properties in nanoporous materials for gas separation. In this chapter, the basic concepts and formulations of physics-based models used in this dissertation would be introduced. While classical density functional theory (cDFT) provides rapid prediction of sorption properties of gas molecules in nanoporous materials, transition-state theory (TST) along with the minimum energy path (MEP) calculated from simplified string method is an efficient computational framework to predict the kinetic properties (viz., diffusion coefficient) of gas molecules in nanoporous materials. In essence, both cDFT and TST are able to provide quantitatively accurate prediction compared with molecular simulation methods but with orders of magnitudes reduction in computational cost. The significant speedup and excellent accuracy make them appealing alternatives compared with conventional methods when constructing large database of molecular properties in nanoporous materials for gas separation.

2.1 Classical Density Functional Theory (cDFT)

Density functional theory (DFT) is an efficient computational method to study the microscopic behavior and thermodynamics properties of electronic properties and inhomogeneous fluids via the one-body density profile.[1, 2] While quantum DFT (qDFT) use density profiles of electrons to solve Schrödinger equation and investigate the electronic properties, density profiles of classical particles (e.g., gas molecules) are used in classical DFT (cDFT) to study the properties of fluid. Although the fame of DFT is now mostly attributed by the public to the Hohenberg-Kohn theorem and Kohn-Sham

equation,[3, 4] the idea of DFT can be traced back to as early as late 19th when van der Waals proved the functional minimization of free energy can be used as the criterion of liquid-gas equilibrium. In Hohenberg-Kohn-Sham theorem, it states that: 1. The one-body density profiles is uniquely determined by the one-body external potential. 2. The ground state energy is the global minimum of the energy functional, which paves the mathematical ground for the application of DFT.

2.1.1 Density Profile

In DFT, almost all the properties can be represented in terms of the functionals of the molecule/site/electron density profiles. The density profile describes the distribution of molecules in the many-body system via the one-body format. For a system containing *N* identical particles, the instantaneous particle density $\hat{\rho}(\mathbf{r})$, which accounts the number of particles at the position **r**, can be described as

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \tag{2.1}$$

where δ is the Dirac-Delta function defined as

$$\delta(\mathbf{r} - \mathbf{r}') = \begin{cases} 0, & \mathbf{r} \neq \mathbf{r}' \\ \infty, & \mathbf{r} = \mathbf{r}' \end{cases}. \tag{2.2}$$

The Dirac-Delta function is also subject to the normalization condition

$$\int d\mathbf{r}\delta(\mathbf{r} - \mathbf{r}_i) = 1. \tag{2.3}$$

In equilibrium, the density profile is defined as the ensemble average of the instantaneous density

$$\rho(\mathbf{r}) = \langle \hat{\rho}(\mathbf{r}) \rangle = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle. \tag{2.4}$$

For single component open system at the temperature T and volume V, the grand partition function can be written as

$$\Xi = \sum_N \frac{1}{N!\Lambda^{3N}} \int d\mathbf{r}^N \exp\left\{-\beta\left[\Gamma\left(\mathbf{r}^N\right) + N\mu + \sum_{i=1}^{N} V^{ext}\left(\mathbf{r}_i\right)\right]\right\}, \tag{2.5}$$

where Λ represents the thermal wavelength, $\beta = 1/\left(k_B T\right)$, k_B stands for the Boltzmann constant, $\Gamma\left(\mathbf{r}^N\right)$ is the total interaction of N particles at configuration $\mathbf{r}^N = \left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\right)$, and V^{ext} is the external potential.

The one-body density profile can be related to the grand partition function by substituting eq (2.5) into eq (2.4)

$$\rho\left(\mathbf{r}\right) = \frac{1}{\Xi}\sum_N \frac{1}{N!\Lambda^{3N}} \int d\mathbf{r}^N \sum_{i=1}^{N} \delta\left(\mathbf{r} - \mathbf{r}_i\right) \exp\left\{-\beta\left[\Gamma\left(\mathbf{r}^N\right) + N\mu + \sum_{i=1}^{N} V^{ext}\left(\mathbf{r}_i\right)\right]\right\}. \tag{2.6}$$

The right-hand side of the above equation is related to the functional derivative of the grand partition function with respect the external potential. The one-body density profile can also be written as, after rearrangement,

$$\rho\left(\mathbf{r}\right) = -\frac{1}{\beta\Xi}\frac{\delta\Xi}{\delta V^{ext}\left(\mathbf{r}\right)} = -\frac{\delta \ln \Xi}{\beta \delta V^{ext}\left(\mathbf{r}\right)}. \tag{2.7}$$

The grand potential Ω is defined in terms of grand partition function

$$\beta\Omega \equiv -\ln \Xi. \tag{2.8}$$

By substituting eq (2.8) into eq (2.7), the one-body density profile can be expressed by

$$\rho\left(\mathbf{r}\right) = \frac{\delta\Omega}{\delta V^{ext}\left(\mathbf{r}\right)}. \tag{2.9}$$

For one-component system, the intrinsic Helmholtz free energy is defined as

$$F \equiv A - \int V^{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{2.10}$$

where A represents the Helmholtz free energy. The intrinsic Helmholtz free energy is related to grand potential via the following the derivation

$$\begin{aligned}\Omega = A - N\mu &= F + \int V^{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} - N\mu \\ &= F + \int \left[V^{ext}(\mathbf{r}) - \mu\right]\rho(\mathbf{r})d\mathbf{r}\end{aligned}. \tag{2.11}$$

The intrinsic Helmholtz free energy and grand potential are then the direct functionals of the one-body density profiles of molecule/site/electron. This allows one to find equilibrium density profiles by minimizing the grand potential

$$\frac{\delta\Omega\left[\rho(\mathbf{r})\right]}{\delta\rho(\mathbf{r})} = 0 \tag{2.12}$$

where the derivative of intrinsic Helmholtz free energy corresponds to the contribution of chemical potential and external potential

$$\frac{\delta F\left[\rho(\mathbf{r})\right]}{\delta\rho(\mathbf{r})} = \mu - V^{ext}(\mathbf{r}). \tag{2.13}$$

The functional derivative of grand potential with respect to the one-body density profile offers the mathematical framework to investigate and understand the molecular structure and relevant thermodynamic properties.

2.1.2 Intrinsic Helmholtz Free Energy

While the basic concepts of density profile and DFT are presented in the previous section, the following sections discuss the detailed formulations of intrinsic Helmholtz free energy, used in the dissertation, for multi-component simple gas molecules in

nanoporous materials. For multi-component system, we use $\boldsymbol{\rho}(\mathbf{r}) = \left[\rho_1(\mathbf{r}), ..., \rho_N(\mathbf{r})\right]$ as the shorthand notation to represent the density profiles of different types of gas molecules, and $\mathbf{r} = (x, y, z)$ is the center-of-mass position for each gas molecule.

Conventionally, the intrinsic Helmholtz free energy is split into an ideal term and excess term

$$F[\boldsymbol{\rho}(\mathbf{r})] = F^{id}[\boldsymbol{\rho}(\mathbf{r})] + F^{ex}[\boldsymbol{\rho}(\mathbf{r})] \tag{2.14}$$

The ideal part corresponds to the intrinsic Helmholtz free energy functional of an inhomogeneous ideal gas

$$F^{id}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \sum_{i=1}^{N} \int \{\ln[\rho_i(\mathbf{r})\Lambda_i^3] - 1\} \rho_i(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{2.15}$$

where T represents the system temperature and N stands for number of molecules considered in the system. While the ideal part can be analytical derived from statistical mechanics, no exact solution is available for the excess part which often requires approximation via perturbation theory. One key advantage of DFT is that the excess Helmholtz free energy can be explicitly customized at different level of accuracy to represent the contributions of different types of intermolecular interactions based on the specific system.

2.1.3 Excess Helmholtz Free Energy

One essential task of all cDFT calculations is to formulate an excess Helmholtz energy functional that is reliable for the specific system under consideration. According to the Lennard-Jones (LJ) model which is often used to represent the intermolecular interactions between simple gas molecules, the excess Helmholtz energy can be split into

contributions due to short-range repulsion and long-range attraction. The former is often represented by the hard-sphere model, F^{hs}, and a perturbation term F^{attr} can be applied to account for van der Waals attractions

$$F^{ex}[\boldsymbol{\rho}(\mathbf{r})] = F^{hs}[\boldsymbol{\rho}(\mathbf{r})] + F^{attr}[\boldsymbol{\rho}(\mathbf{r})] \tag{2.16}$$

2.1.3.1 Short-Range Repulsion

As well documented, the excess Helmholtz energy of a hard-sphere system can be accurately described by the modified fundamental measure theory (MFMT)[5, 6]

$$F^{hs}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \int \Phi^{hs}[n_\alpha(\mathbf{r})]\, \mathrm{d}\mathbf{r} \tag{2.17}$$

where

$$\Phi^{hs} = -n_0 \ln(1-n_3) + \frac{n_0 n_2 - \mathbf{n}_{V1}\cdot\mathbf{n}_{V1}}{1-n_3} + \frac{n_3 + (1-n_3)^2 \ln(1-n_3)}{36\pi n_3^2 (1-n_3)^2}\left[(n_2)^3 - 3n_2 \mathbf{n}_{V2}\cdot\mathbf{n}_{V2}\right] \tag{2.18}$$

with

$$\begin{aligned} & n_\alpha(\mathbf{r}) = \sum_{i=1}^{N} \int \rho_i(\mathbf{r}')\, w_i^{(\alpha)}\left(|\mathbf{r}-\mathbf{r}'|\right) \mathrm{d}\mathbf{r}' \\ & \alpha \in \{0, 1, 2, 3, V1, V2\} \end{aligned} \tag{2.19}$$

In eq (2.19), $w_i^{(\alpha)}$ are a set of weight functions characterizing the differential geometry of each spherical particle:

$$\begin{cases} w_i^{(2)}(r) = \pi d_i^2 w_i^{(0)}(r) \\ \qquad\quad = 2\pi d_i w_i^{(1)}(r) \\ \qquad\quad = \delta(d_i/2 - r) \\ w_i^{(3)}(r) = \theta(d_i/2 - r) \\ \mathbf{w}_i^{(V2)}(r) = 2\pi d_i \mathbf{w}_i^{(V1)}(r) = w_i^{(2)}(r)\dfrac{\mathbf{r}}{r} \end{cases} \tag{2.20}$$

where δ denotes the Dirac-delta function, θ is the Heaviside step function, and d_i is the hard-sphere diameter for component *i*. For all gas molecules considered in this dissertation, the Barker-Henderson theory is used to calculate the hard-sphere diameter from the LJ parameters[7, 8]

$$d_i = \sigma_i \left(\frac{1 + 0.2977 T_i^*}{1 + 0.33163 T_i^* + 0.0010477 T_i^{*2}} \right) \tag{2.21}$$

where $T_i^* = k_B T / \varepsilon_i$, ε_i and σ_i stand for the LJ energy and size parameters of the gas molecule *i*, respectively.

2.1.3.2 Mean-Field Approximation

For the attraction part of the excess Helmholtz energy, one convenient choice is that from the mean-field approximation (MFA)

$$F^{MFA} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \iint \rho_i(\mathbf{r}) \rho_j(\mathbf{r}') u_{ij}^{attr}\left(|\mathbf{r} - \mathbf{r}'|\right) \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}' \tag{2.22}$$

where

$$u_{ij}^{attr}(r) = \begin{cases} 0 & r < \sigma_{ij} \\ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] & r > \sigma_{ij} \end{cases} \tag{2.23}$$

and the cross parameters are calculated from the Lorentz-Berthelot mixing rule. Whereas MFA is commonly used in cDFT calculations including characterization of porous materials by gas adsorption, it reduces to an equation of state for bulk systems similar to the van der Waals equation. While more accurate formulations are available for one-component LJ fluids,[9] extension of those formulations to multicomponent systems is

theoretically challenging due to the lack of analytical expressions for the bulk correlation functions.

2.1.3.3 Weighted-Density Approximation

Alternatively, the attraction part of excess Helmholtz energy can be more accurately described with the consideration of correlation effect via weighted density approximation (WDA)[10, 11]

$$F^{attr}[\boldsymbol{\rho}(\mathbf{r})] = F^{MFA}[\boldsymbol{\rho}(\mathbf{r})] + F^{cor}[\boldsymbol{\rho}(\mathbf{r})] \tag{2.24}$$

where F^{cor} corresponds to the local correlation Helmholtz energy

$$F^{cor}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \sum_{i=1}^{N} \int \Phi^{cor}[\bar{\rho}(\mathbf{r})]\, \mathrm{d}\mathbf{r} \,. \tag{2.25}$$

The reduced local correlation Helmholtz energy per volume, Φ^{cor}, is approximated by that corresponding to the bulk phase at weighted density

$$\bar{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \frac{3}{4\pi d_i^3} \int \rho_i(\mathbf{r}')\, \theta\left(d_i - |\mathbf{r} - \mathbf{r}'|\right) d\mathbf{r}' \tag{2.26}$$

In the bulk phase, Φ^{cor} can be written in the following form

$$\Phi^{cor}(\rho) = \frac{F_{bulk}^{LJ}(\rho) - F_{bulk}^{hs}(\rho) - F_{bulk}^{MFA}(\rho)}{V} \beta \tag{2.27}$$

where $F_{bulk}^{LJ}(\rho)$ stands for the excess Helmholtz energy of a bulk LJ fluid calculated from the MBWR equation of state,[12] $F_{bulk}^{hs}(\rho)$ denotes the hard-sphere Helmholtz energy according to the Carnahan-Starling equation of state,[13] $F_{bulk}^{MFA}(\rho)$ represents the mean-field Helmholtz energy for the one-component fluid. Explicit expressions are available

for the hard-sphere and mean-field excess Helmholtz energies shown in eq (2.28) and eq (2.29), respectively:

$$\beta \frac{F_{bulk}^{hs}[\bar{\rho}(\mathbf{r})]}{V} = \frac{4y - 3y^2}{(1-y)^2} \bar{\rho}(\mathbf{r}) \tag{2.28}$$

$$\beta \frac{F_{bulk}^{MFA}[\bar{\rho}(\mathbf{r})]}{V} = -\frac{16}{9} \pi \beta \varepsilon_x \bar{\rho}(\mathbf{r})^2 \sigma_x^3 \tag{2.29}$$

where $y = \frac{\bar{\rho}(\mathbf{r}) \pi d_x^3}{6}$. To take advantage of the well-established bulk properties, van der Waals one-fluid theory (vdw1) is used to estimate the d_x, ε_x and x_i which stands for the hard-sphere diameter, the LJ energy and size parameters for the mixture, respectively:

$$d_x^3 = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} d_{ij}^3 \tag{2.30}$$

$$\sigma_x^3 = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \sigma_{ij}^3 \tag{2.31}$$

$$\varepsilon_x = \frac{1}{\sigma_x^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \varepsilon_{ij} \sigma_{ij}^3 \tag{2.32}$$

Combing eqs (2.12) ~ (2.15) leads to the following Euler-Lagrange equations

$$\rho_i(\mathbf{r}) = \rho_{bulk,i} \exp\left[\beta \mu_i^{ex,mixture} - \beta V_i^{ext} - \beta \frac{\delta F^{ex}}{\delta \rho_i(\mathbf{r})} \right] \qquad (i = 1, 2, ..., N) \tag{2.33}$$

From eq (2.33), one can calculate the density profiles of individual species for an *N*-component gas mixture adsorption in nanoporous materials. With an explicit expression for the excess Helmholtz energy functional as given by eqs (2.16) ~ (2.32), we can solve the density profiles using conjugate gradient descent method.[14] In comparison with

cDFT for single-component systems, the computational cost scales linearly with the number of chemical species in the gas mixture. With the massively parallel algorithm via graphic processing unit (GPU)-acceleration, the computational cost of multi-component gas mixture adsorption in nanoporous materials using cDFT can be drastically reduced to the scale of few seconds for each cDFT calculation.

2.1.4 Molecular System

Although the above discussion is focused on the formulations of cDFT for simple gas, the applications of cDFT can also be extended to molecular system with proper formulations for the ideal and excess Helmholtz energy term. For example, for rigid polyatomic gas molecules, the external potential depends on the position of molecular center-of-mass and orientation and the external potential at **r** is determined by

$$V^{ext}(\mathbf{r}) = \frac{1}{2\pi}\int V^{ext}(\mathbf{r},\omega)d\omega \tag{2.34}$$

where ω represents molecular orientation. For the adsorption at relatively low pressure, the adsorption amount would linearly increase with the increase of pressure according to Henry's constant. The Henry's constant for rigid polyatomic molecules can be expressed as:

$$K_h = \frac{1}{2\pi k_B TV}\int d\mathbf{r}\int d\omega \exp\left[-\beta V^{ext}(\mathbf{r},\omega)\right]. \tag{2.35}$$

For the contribution of ideal Helmholtz energy, the formulation depends on the choice for the excess Helmholtz energy. For small polyatomic molecules, the intermolecular potential can still be fitted into single Lennard-Jones potential. And the formulations discussed above are directly applicable for the ideal and excess Helmholtz energy terms.

For more complicated molecules, the excess Helmholtz energy can be accurately described by well-developed equation of state for molecular systems such as variations of statistical associating fluid theory (SAFT) while the formulations for the ideal Helmholtz energy need to be tailored to fit the equation of state of choice.[15, 16]

2.2 Transition-State Theory (TST)

Transition-state theory (TST) has been widely used in predicting the essential properties in chemical reaction and transport phenomenon.[17] As stated by transition-state theory (TST)[18], the self-diffusion coefficient for a guest molecule inside a nanoporous materials can be calculated by

$$D_0 = \frac{1}{2} k a^2 \tag{2.36}$$

where D_0 is the self-diffusion coefficient, k is the hopping rate (i.e., transmission rate), and a is the hopping distance between two neighboring cages. At infinite dilution, the hopping rate can be obtained from the minimum energy path (MEP) for the molecular diffusion following the Bennett-Chandler approach[19]

$$k = \sqrt{\frac{k_B T}{2\pi m}} \frac{\exp\left[-\beta V^{ext}\left(s^*\right)\right]}{\int_0^1 \exp\left[-\beta V^{ext}\left(s\right)\right] ds} \tag{2.37}$$

where m represents the molecular mass, V^{ext} is the potential energy due to the interaction of the guest molecule with the porous material. In the TST, the minimum energy path is described in terms of a dimensionless variable s, which represents the normalized reaction coordinate for the molecular transition between neighboring cages. In general, s

depends on the molecular configuration and the center-of-mass (COM) position for the guest molecule.

2.2.1 Simplified String Method for Simple Gas Molecules

For simple gas molecules, the molecule is often described as single Lennard-Jones (LJ) site for its van der Waals (vdW) interaction and its interaction with the nanoporous material can be described as

$$V^{ext} = \sum_{j=1}^{N_f} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.38}$$

where N_f is the number of atoms in nanoporous materials, ε stand for the LJ energy parameter, σ represents the LJ size parameter and r_{ij} is the interatomic distance between atom i and atom j. Because the external potential is only the function of molecular COM position for simple gas molecules, in this case the dimensionless variable can be represented by $s = s(\mathbf{r})$ where $\mathbf{r} = (x, y, z)$ represents the position of the molecular COM. Therefore, the original simplified string method is directly applicable to find the MEP for simple gas molecules. In each iteration, the string is evolved according to the gradient of external potential with respect to cartesian coordinate

$$\frac{ds_i}{dt} = -\nabla V^{ext}(s_i) = -\left[\frac{\partial V^{ext}(s_i)}{\partial x} + \frac{\partial V^{ext}(s_i)}{\partial y} + \frac{\partial V^{ext}(s_i)}{\partial z}\right] \tag{2.39}$$

where t is the fictitious time used in the iteration to search for the MEP, and s_i represents the image I on the string. The images along the string can then be evolved according to

$$\mathbf{r}_{i,k}^{\parallel}(t) = \mathbf{r}_{i,k}(t) - \Delta t \frac{\partial V^{ext}(s_i)}{\partial \mathbf{r}_k}\Big|_t \tag{2.40}$$

where the superscript ‖ represents the evolved string. After the string evolution, images along the string are redistributed via linear interpolation

$$\mathbf{r}_i\left(t+\Delta t\right)=\mathbf{r}_{k-1}^{\|}\left(t\right)+\left[\frac{i-1}{N-1}L_{\mathbf{r}}^{\|}\left(t\right)-l_{\mathbf{r},k-1}^{\|}\left(t\right)\right]\frac{\mathbf{r}_k^{\|}\left(t\right)-\mathbf{r}_{k-1}^{\|}\left(t\right)}{\left|\mathbf{r}_k^{\|}\left(t\right)-\mathbf{r}_{k-1}^{\|}\left(t\right)\right|} \tag{2.41}$$

where $l_{\mathbf{r},i}^{\|}=\sum_{j=2}^{i}\left|\mathbf{r}_j^{\|}-\mathbf{r}_{j-1}^{\|}\right|$ is the string arc length for spatial coordinates (*x*, *y*, *z*) at image *i* after evolution, $L_{\mathbf{r}}^{\|}$ is the entire string arc length for molecular COM position after evolution, and *N* is the number of images used in the string. It is worth mentioning although simplified version of string method is used, it is more stable and accurate compared with the original string method and nudged elastic band method.

2.2.2 Simplified String Method for Rigid Polyatomic Gas Molecules

While the original simplified string method can be directly applied to find the MEP for simple gas molecule, modifications are needed in the simplified string method for rigid polyatomic gas molecules. For the rigid polyatomic gas molecules, the reaction coordinate needs to be defined by six collective variables $s\left(\theta_1,...,\theta_6\right)$. The first three variables are related to the molecular position and the other three variables represent the Euler angles of the guest molecule, i.e., $s\left(\theta_1,...,\theta_6\right)=s\left(\mathbf{r},\omega\right)$, where $\mathbf{r}=\left(x,y,z\right)$ represents the position for the molecular center of mass (COM), and $\omega=\left(\alpha,\beta,\gamma\right)$ describes how a polyatomic molecule is oriented relative to its original input structure. At each point/image, the dimensionless variables *s* can be evaluated in terms of **r** and ω

$$s_i = \sqrt{\left(\frac{l_{\mathbf{r},i}}{L_{\mathbf{r}}}\right)^2 + \left(\frac{l_{\omega,i}}{L_{\omega}}\right)^2} \tag{2.42}$$

where $l_{\mathbf{r},i} = \sum_{j=2}^{i} \left|\mathbf{r}_j - \mathbf{r}_{j-1}\right|$ and $l_{\omega,i} = \sum_{j=2}^{i} \left|\omega_j^* - \omega_{j-1}^*\right|$ are the string arc lengths for the spatial and rotational variables at image *i*, while $L_{\mathbf{r}}$ and L_{ω} are the arc lengths of the entire string for **r** and ω, respectively.

For polyatomic molecules, the non-bonded intermolecular potential often involves short-range repulsion, vdW attraction and electrostatic interactions due to atomic partial charges, which can be accurately captured by the combination of LJ potential and Coulomb potential. As a result, given the position and configuration of the guest molecule, the external potential describing its interaction with nanoporous material is given by

$$V^{ext} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_f} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.43}$$

where N_g and N_f are the number of atoms in each guest molecule and that from the nanoporous material, ε and σ stand for the LJ energy and size parameters, respectively, r_{ij} is the distance between atom *i* and atom *j*, ε_0 stands for the vacuum permeability and q_i represents charge for atom *i*.

According to the simplified string method,[20] the evolution of the normalized reaction coordinate is driven by the full gradient of the external potential

$$\frac{ds_i}{dt} = -\nabla V^{ext}\left(s_i\right) = -\left[\frac{\partial V^{ext}\left(s_i\right)}{\partial x} + \frac{\partial V^{ext}\left(s_i\right)}{\partial y} + \frac{\partial V^{ext}\left(s_i\right)}{\partial z} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \alpha} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \beta} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \gamma}\right] \tag{2.44}$$

where t is a fictitious time used in the iteration to search for the minimum energy path, and s_i represents image i on the string. During each iteration, the guest molecule is first updated according to

$$\theta_{i,k}^{\|}\left(t\right) = \theta_{i,k}\left(t\right) - \Delta t \frac{\partial V^{ext}\left(s_i\right)}{\partial \theta_k}|_t, \qquad k = 1,...,6 \tag{2.45}$$

where θ_i represents a molecular coordinate (position or angle) corresponding to image i, the partial derivative is evaluated at fictious time t, and the superscript $^{\|}$ represents the updated string.

After string evolution in each step, interpolation/reparameterization is needed to retain the continuous shape of string through the nanoporous materials. Different from simple gas molecules, when the rigid polyatomic molecule diffuses through a nanoporous material, the preferred molecular orientation depends on the position of the molecular center of mass (COM). As a result, different images have orientations independent from each other, and only the molecular COM positions need to be redistributed/interpolated according to the arc length

$$\mathbf{r}_i\left(t+\Delta t\right) = \mathbf{r}_{k-1}^{\|}\left(t\right) + \left[\frac{i-1}{N-1} L_{\mathbf{r}}^{\|}\left(t\right) - l_{\mathbf{r},k-1}^{\|}\left(t\right)\right] \frac{\mathbf{r}_k^{\|}\left(t\right) - \mathbf{r}_{k-1}^{\|}\left(t\right)}{\left|\mathbf{r}_k^{\|}\left(t\right) - \mathbf{r}_{k-1}^{\|}\left(t\right)\right|} \tag{2.46}$$

where $l_{\mathbf{r},i}^{\|} = \sum_{j=2}^{i}\left|\mathbf{r}_j^{\|} - \mathbf{r}_{j-1}^{\|}\right|$ is the string arc length for spatial coordinates (x, y, z) at image i

after evolution, $L_{\mathbf{r}}^{\|}$ is the entire string arc length for molecular COM position after evolution, and *N* is the number of images used in the string. To prevent the abrupt change of the molecular orientation, a smooth function is used for angular variables between neighboring images[21]

$$\theta_i = (1-\delta)\theta_i + \frac{\delta}{2}(\theta_{i-1} + \theta_{i+1}) \tag{2.47}$$

where δ is the parameter to control the degree of smoothness. A small smooth parameter, such as $\delta = 1\times10^{-4}$, can ensure the accuracy of molecular orientation that minimizes the external potential and give accurate prediction of diffusion coefficient via TST with the converged MEP.

2.2.3 Simplified String Method for Flexible Gas Molecules

For flexible gas molecules, the straightforward way for the string evolution is by directly differentiating the molecular coordinate with respect to the combination of external potential and intramolecular potential

$$\frac{ds_i}{dt} = -\nabla V^{ext}(s_i) = -\left[\frac{\partial V^{ext}(s_i)}{\partial x} + \frac{\partial V^{ext}(s_i)}{\partial y} + \frac{\partial V^{ext}(s_i)}{\partial z} + \frac{\partial V^{ext}(s_i)}{\partial \alpha} + \frac{\partial V^{ext}(s_i)}{\partial \beta} + \frac{\partial V^{ext}(s_i)}{\partial \gamma}\right] \tag{2.48}$$

where *j* stands for atom *j* in the molecule. However, this formulation often leads to extremely slow converge of string because the strong intramolecular interaction often holds atoms together from moving towards the converged position. To speed up the string evolution, the movement of molecular COM and molecular flexibility could be split.

More specifically, in each iteration the molecular COM would be updated according to the external potential assuming the gas molecule to be rigid

$$\frac{ds_i}{dt} = -\nabla V^{ext}\left(s_i\right) = -\left[\frac{\partial V^{ext}\left(s_i\right)}{\partial x} + \frac{\partial V^{ext}\left(s_i\right)}{\partial y} + \frac{\partial V^{ext}\left(s_i\right)}{\partial z}\right]. \tag{2.49}$$

Then, the molecular structure of each image would be updated according to the external potential and intramolecular interaction as eq (2.48). It is also worth mentioning here because the energy barriers between molecular conformations often lead to molecular structures constrained within local minima within the framework of simplified string method. In order to find the molecular conformation minimizing the external potential and intramolecular potential, different molecular conformations need to be tested in each iteration

$$\min\left[V^{ext}\left(x,y,z,\left[\mathbf{r}_1,...,\mathbf{r}_N\right]_M\right) + V^{intra}\left(x,y,z,\left[\mathbf{r}_1,...,\mathbf{r}_N\right]_M\right)\right]V \tag{2.50}$$

where M represents the number of conformations for the gas molecule and N stands for the number of atoms in the gas molecules.

Bibliography

1. Wu, J., *Variational Methods in Molecular Modeling*. 2016: Springer.

2. Wu, J. and Z. Li, *Density-functional theory for complex fluids.* Annu Rev Phys Chem, 2007. **58**: p. 85-112.

3. Hohenberg, P. and W. Kohn, *Inhomogeneous electron gas.* Physical review, 1964. **136**(3B): p. B864.

4. Kohn, W. and L.J. Sham, *Self-consistent equations including exchange and correlation effects.* Physical review, 1965. **140**(4A): p. A1133.

5. Yu, Y.X. and J.Z. Wu, *Structures of hard-sphere fluids from a modified fundamental-measure theory.* J Chem Phys, 2002. **117**(22): p. 10156-10164.

6. Roth, R., et al., *Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version.* J Phys: Condens Matter, 2002. **14**(46): p. 12063.

7. Henderson, D., *Monte Carlo and perturbation theory studies of the equation of state of the two-dimensional Lennard-Jones fluid.* Mol Phys, 1977. **34**(2): p. 301-315.

8. Cotterman, R.L., B.J. Schwarz, and J.M. Prausnitz, *Molecular thermodynamics for fluids at low and high densities. Part I: Pure fluids containing small or large molecules.* AIChE J, 1986. **32**(11): p. 1787-1798.

9. Fu, J., et al., *Density Functional Methods for Fast Screening of Metal Organic Frameworks for Hydrogen Storage.* J Phys Chem C, 2015. **119**(10): p. 5374-5385.

10. Yu, Y.X., *A novel weighted density functional theory for adsorption, fluid-solid interfacial tension, and disjoining properties of simple liquid films on planar solid surfaces.* J Chem Phys, 2009. **131**(2): p. 024704.

11. Liu, Y., et al., *Development of a density functional theory in three-dimensional nanoconfined space: H2 storage in metal-organic frameworks.* J Phys Chem B, 2009. **113**(36): p. 12326-31.

12. Johnson, J.K., J.A. Zollweg, and K.E. Gubbins, *The Lennard-Jones Equation of State Revisited.* Mol Phys, 1993. **78**(3): p. 591-618.

13. Carnahan, N.F. and K.E. Starling, *Equation of state for nonattracting rigid spheres.* J Chem Phys, 1969. **51**(2): p. 635-636.

14. Hager, W.W. and H.C. Zhang, *Algorithm 851: CG DESCENT, a conjugate gradient method with guaranteed descent.* ACM Transactions on Mathematical Software, 2006. **32**(1): p. 113-137.

15. Chapman, W.G., et al., *SAFT: Equation-of-state solution model for associating fluids.* Fluid Phase Equilibria, 1989. **52**: p. 31-38.

16. Gross, J. and G. Sadowski, *Application of the perturbed-chain SAFT equation of state to associating systems.* Industrial & engineering chemistry research, 2002. **41**(22): p. 5510-5515.

17. Truhlar, D.G., B.C. Garrett, and S.J. Klippenstein, *Current status of transition-state theory.* The Journal of physical chemistry, 1996. **100**(31): p. 12771-12800.

18. Truhlar, D.G. and B.C. Garrett, *Variational transition state theory.* Annual Review of Physical Chemistry, 1984. **35**(1): p. 159-189.

19. Kärger, J., D.M. Ruthven, and D.N. Theodorou, *Diffusion in Nanoporous Materials*. 2012.

20. E, W., W. Ren, and E. Vanden-Eijnden, *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events.* J Chem Phys, 2007. **126**(16): p. 164103.

21. Maragliano, L., et al., *String method in collective variables: minimum free energy paths and isocommittor surfaces.* J Chem Phys, 2006. **125**(2): p. 24106.

Chapter 3. A GPU Implementation of Classical Density Functional Theory for Rapid Prediction of Gas Adsorption in Nanoporous Materials

Nanoporous materials are promising as the next generation of adsorbents for gas storage and separation with ultrahigh capacity and selectivity. The recent advent of data-driven approaches in materials modeling provides alternative routes to tailor nanoporous materials for customized applications. Typically, a data-driven model requires a large amount of training data that cannot be generated solely by experimental methods or molecular simulations. In this chapter, we propose an efficient implementation of classical density functional theory with a graphic processing unit (GPU) for the fast yet accurate prediction of gas adsorption isotherms in nanoporous materials. In comparison to serial computing with the central processing unit, the massively parallel GPU-accelerated implementation reduces the computational cost by more than two orders of magnitude. The proposed algorithm renders new opportunities not only for the efficient screening of a large materials database for gas adsorption but it may also serve as an important stepping stone towards the inverse design of nanoporous materials tailored to desired applications.

3.1 Introduction

Nanoporous materials such as zeolites, metal-organic frameworks and covalent organic frameworks are excellent candidates as the next generation of absorbents to achieve ultrahigh adsorption amount and separation selectivity.[1-3] To harness the data revolution and inverse design of nanoporous materials for specific applications, we need large databases to gauge materials performance which can hardly be met only by

experiment and molecular simulations.[4-7] Semi-analytical methods, such as classical density functional theory (cDFT), offer an alternative for rapid prediction of gas adsorption in nanoporous materials.[8-10]

Recent increase in computational power can be attributed not only to the enhanced performance of arithmetic logic units (ALUs) but also to advances in more efficient and convenient programming interface for parallel computations. The library of parallel computing, such as message passenger interface (MPI), open multi-processing (openMP), compute unified device architecture (CUDA) and open graphics library (OpenGL), saves significant effort for software development in scientific computation and allows for the computational power to go beyond the performance limit of single threads. Intuitively, a large-scale parallel-computing system may be considered as many CPUs running at the same time.[11] Although supercomputers equipped with massive CPUs have been widely used for parallel computing, conventional CPUs are not designed for high-throughput data processing but for fast serial processing. As the number of CPUs increases, the performance speedup by parallel computing quickly reaches a plateau because more computer time would be needed to communicate among CPUs in order to avoid race condition. In comparison to CPU, a graphic processing unit (GPU) is designed to have many more threads and higher floating-point operations per second (FLOPS) for parallel computing and high-throughput data processing. However, GPU applications are often limited by the hardware capacity, especially the memory size.[12]

GPUs have found broad applications in computational science. Many molecular simulation packages (e.g. LAMMPS, NAMD and GROMACS) now include GPU

versions to support parallel computing.[13-16] Impressive speedups are often reported by implementing various first principles methods with GPU. For example, Jia et al. found that GPU can accelerate the computational speed by 20 times for evaluating plane-wave pseudopotentials in the Kohn-Sham density functional theory (DFT).[17] Nitsche et al. achieved similar improvements in hybrid QM/MM calculations.[18] Stopper and Roth demonstrated that the GPU implementation can speed up the cDFT calculation up to 70 times for systems of hard disks in 2D and 60 times for hard spheres in 3D.[19]

We demonstrated in previous work that cDFT facilitates fast and accurate prediction of gas adsorption in nanoporous materials.[8, 9] Here we propose an efficient implementation of cDFT with parallel computation on GPU and test its efficiency for predicting gas adsorption in nanoporous materials. With the cDFT functions fully parallelized and a minimal memory transfer between the host (CPU) and device (GPU), the new algorithm can further speed up the calculation by more than two orders of magnitudes (with the speedup factor up to 350) while an accurate prediction of the adsorption amount is maintained in comparison to the serial implementation executed on CPU. Besides, we find that the conjugate gradient descent method allows for faster and more robust convergence for solving the cDFT equation compared to the Picard iteration, which also accelerates the computational speed. The excellent performance empowered by the GPU implementation highlights the potential application of cDFT to the inverse design of nanoporous materials for gas storage and separation.

3.2 Methods and Models

3.2.1 Classical Density Functional Theory (cDFT)

Classical density functional theory (cDFT) is applicable to gas adsorption in one-, two- and three-dimensional systems but the computational cost scales exponentially with the dimensionality[20]. Efficient numerical algorithms are essential for its applications to high-dimensional systems in particular for the adsorption of polyatomic gas molecules. From a practical perspective, the numerical performance of cDFT is mostly evaluated for its prediction of gas adsorption in three-dimensional nanoporous materials.

In this work, we consider the adsorption of small gas molecules such as methane and hydrogen in nanoporous materials that have a rigid porous structure. Approximately, the interaction between gas molecules and that between a gas molecule and each atom from the nanostructure material can be represented by the Lennard-Jones (LJ) 12-6 model. Given temperature and pressure of the gas phase in the bulk, the total adsorption can be fully specified by the local density, $\rho(\mathbf{r})$, of gas molecules inside the nanoporous material:

$$\Gamma = \frac{1}{V}\int \rho(\mathbf{r})\, d\mathbf{r} \tag{3.1}$$

where $\mathbf{r} = (x, y, z)$ stands for the center-of-mass position for a gas molecule and V is the system volume. Whereas adsorption amount is conventionally expresed in the units of adsorbate volume at the standard condition (1 atm and 25 °C) per adsorbent volume, here we use the moles of adsorbed gas per adsorbent volume for simplicity.

The cDFT equations for calculating gas adsorption have been reported before.[8, 9] Here we recapitulate the key equations that are pertinent to numerical implementations. According to cDFT, the thermodynamic properties of gas molecules in the micropores of a nanoporous material can be derived from the grand potential

$$\Omega[\rho(\mathbf{r})] = F[\rho(\mathbf{r})] + \int [V^{ext}(\mathbf{r}) - \mu_{bulk}]\rho(\mathbf{r})\, d\mathbf{r} \tag{3.2}$$

where F represents the intrinsic Helmholtz energy functional, V^{ext} denotes the external potential, i.e., the potential energy for each gas molecule due to its interaction with the nanoporous material, and μ_{bulk} stands for the chemical potential of the gas molecule in the bulk phase. Throughout this work, the external potential is calculated from the universal force field (UFF)[21]. Meanwhile, the modified Benedict-Webb-Rubin (MBWR) equation of state is used to calculate the chemical potential of the bulk fluid.[22]

Conventionally, the intrinsic Helmholtz energy functional is split into an ideal term and the excess:

$$F[\rho(\mathbf{r})] = F^{id}[\rho(\mathbf{r})] + F^{ex}[\rho(\mathbf{r})] \tag{3.3}$$

The ideal part corresponds to the intrinsic Helmholtz energy functional of an inhomogeneous ideal gas at the system temperature

$$F^{id}[\rho(\mathbf{r})] = k_B T \int \{\ln[\rho(\mathbf{r})\Lambda^3] - 1\}\rho(\mathbf{r})\, d\mathbf{r} \tag{3.4}$$

where T is absolute temperature, k_B represents the Boltzmann constant and Λ denotes the thermal wavelength of the gas molecule. Within the LJ model, the excess Helmholtz energy can be further divided in terms of a hard-sphere contribution, F^{hs}, which accounts

for the molecular excluded volume effect, and a perturbation term, F^{attr}, to describe intermolecular attractions[23-26]

$$F^{ex}(\mathbf{r}) = F^{hs}(\mathbf{r}) + F^{attr}(\mathbf{r}) \tag{3.5}$$

The modified fundamental measure theory (MFMT) provides an accurate description of the hard-sphere free energy[27, 28]

$$F^{hs} = k_B T \int \Phi^{hs}[n_\alpha(\mathbf{r})]\, \mathrm{d}\mathbf{r} \tag{3.6}$$

where

$$\Phi^{hs} = -n_0 \ln(1-n_3) + \frac{n_1 n_2 - \mathbf{n}_{V1}\cdot\mathbf{n}_{V2}}{1-n_3} + \frac{n_3 + (1-n_3)^2 \ln(1-n_3)}{36\pi n_3^2 (1-n_3)^2}\left(n_2^3 - 3n_2 \mathbf{n}_{V2}\cdot\mathbf{n}_{V2}\right) \tag{3.7}$$

$$\begin{aligned} & n_\alpha(\mathbf{r}) = \int \rho(\mathbf{r}')\, w^{(\alpha)}\left(|\mathbf{r}-\mathbf{r}'|\right) \mathrm{d}\mathbf{r}' \\ & \alpha \in \{0, 1, 2, 3, V1, V2\} \end{aligned} \tag{3.8}$$

and $w^{(\alpha)}$ are the weight functions

$$\begin{cases} w^{(2)}(r) = \pi d^2 w^{(0)}(r) \\ \qquad\quad = 2\pi d w^{(1)}(r) \\ \qquad\quad = \delta(d/2 - r) \\ w^{(3)}(r) = \theta(d/2 - r) \\ \mathbf{w}^{(V2)}(r) = 2\pi d \mathbf{w}^{(V1)}(r) = w^{(2)}(r)\dfrac{\mathbf{r}}{r} \end{cases} \tag{3.9}$$

where δ stands for the Dirac-delta function, θ denotes the Heaviside step function, and d is the hard-sphere diameter. For gas systems considered in this work, the hard-sphere diameter is calculated from the LJ model according to the Barker-Henderson (BH) theory[29, 30]

$$d = \sigma\left(\frac{1+0.2977T^*}{1+0.33163T^*+0.0010477T^*}\right) \tag{3.10}$$

where $T^*=k_BT/\varepsilon$, ε and σ represent the LJ energy and size parameters of the gas molecule, respectively.

For the attraction part of the excess Helmholtz energy, one popular choice is by using the mean-field approximation (MFA)[31]

$$F^{MFA} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int\int \rho_i(\mathbf{r})\rho_j(\mathbf{r}')u_{ij}^{attr}\left(|\mathbf{r}-\mathbf{r}'|\right)\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{3.11}$$

where

$$u^{attr}(r) = \begin{cases} 0 & r<\sigma \\ 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right] & r>\sigma \end{cases} \tag{3.12}$$

Although MFA is numerically convenient, it underestimates the adsorption amount in comparison with the grand canonical Monte Carlo simulation (GCMC). To account for correlation effects, we include an additional term by taking the weighted density approximation (WDA)[32]

$$F^{attr}\left[\rho(\mathbf{r})\right] = F^{MFA}\left[\rho(\mathbf{r})\right] + F^{cor}\left[\rho(\mathbf{r})\right] \tag{3.13}$$

where F^{cor} is a local correlation Helmholtz energy

$$F^{cor}\left[\rho(\mathbf{r})\right] = k_BT\int \Phi^{cor}\left[\overline{\rho}(\mathbf{r})\right]\mathrm{d}\mathbf{r} \tag{3.14}$$

The local reduced Helmholtz energy per volume, Φ^{cor} , is approximated by that corresponding to the bulk phase at a weighted density $\overline{\rho}(\mathbf{r})$ defined as

$$\bar{\rho}(r) = \frac{3}{4\pi d^3} \int \rho(\mathbf{r}')\theta\left(d - |\mathbf{r} - \mathbf{r}'|\right) \mathrm{d}\mathbf{r}' \tag{3.15}$$

For the bulk system, Φ^{cor} can be written in terms of an analytical form

$$\Phi^{cor}(\rho) = \frac{F_{bulk}^{LJ}(\rho) - F_{bulk}^{hs}(\rho) - F_{bulk}^{MFA}(\rho)}{V}\beta \tag{3.16}$$

where $F_{bulk}^{LJ}(\rho)$ is the excess Helmholtz energy of the bulk fluid calculated from the MBWR equation of state, $F_{bulk}^{hs}(\rho)$ represents the hard-sphere contribution as calculated from the Carnahan-Starling equation of state, and $F_{bulk}^{MFA}(\rho)$ denotes the corresponding mean-field Helmholtz energy for the bulk fluid. Compared with MFA, WDA predicts a more accurate adsorption isotherm for gas adsorption by nanoporous materials. However, it is also computationally more expensive because WDA takes an extra convolution in evaluating the weighted density and the local correlation Helmholtz energy.

At equilibrium, the density profile of gas molecules is found by minimizing the grand potential

$$\frac{\delta\Omega\left[\rho(r)\right]}{\delta\rho(r)} = 0 \tag{3.17}$$

leading to the self-consistent Euler-Lagrange equation:

$$\rho(\mathbf{r}) = \frac{1}{\Lambda^3}\exp\left[\beta\mu^{bulk} - \beta V^{ext} - \beta\frac{\delta F^{ex}}{\delta\rho(\mathbf{r})}\right] = \rho_{bulk}\exp\left[\beta\mu^{ex} - \beta V^{ext} - \beta\frac{\delta F^{ex}}{\delta\rho(\mathbf{r})}\right] \tag{3.18}$$

where $\beta = 1/(k_B T)$.

The equations above demonstrate how the adsorption amount can be calculated from the intermolecular potential and the interaction energy between each gas molecule

and the nanoporous material. Even for a simple fluid, different levels of accuracy and numerical efficiency can be achieved by formulating the excess Helmholtz energy. In principle, the above procedure is applicable to more complicated gas molecules by using either coarse-grained or atomistic models. In that case, cDFT must account for the configurations of polyatomic molecules that involve a higher dimensionality and often requires additional term to describe thermodynamic nonideality (viz., excess Helmholtz energy) due to both intra- and intermolecular interactions.[33] While adsorption of polyatomic molecules is beyond the scope of this work, we expect that parallel implementation of cDFT would have even more advantages in comparison to conventional CPU-based numerical methods. At present, direct implementation of molecular cDFT for polyatomic systems is still computationally prohibitive to generate extensive thermodynamic data that can be used for training machine-learning algorithms as required for the inverse design of nanoporous materials.

3.2.2 GPU Implementation

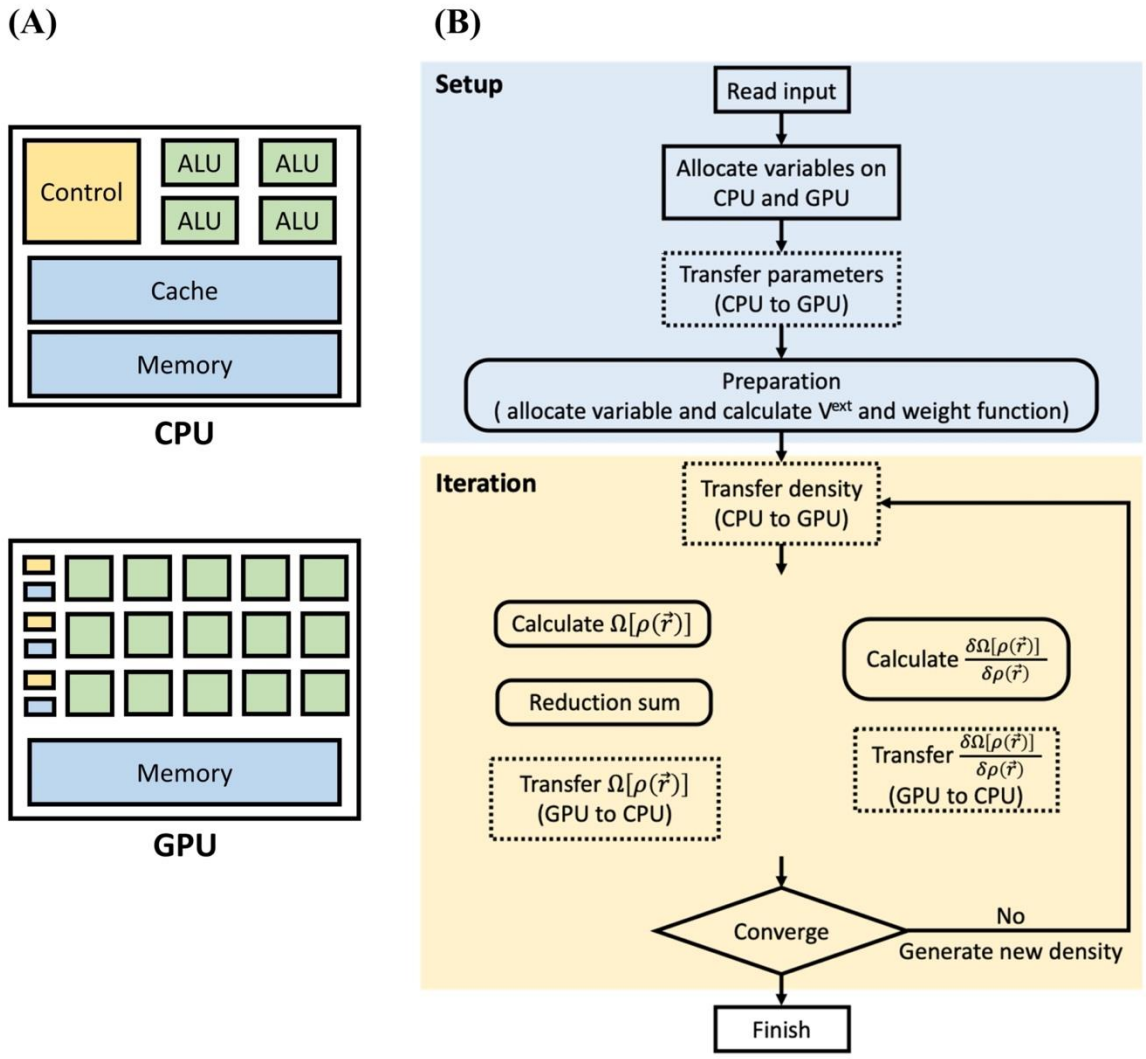

Figure 3.1. (A) Schematic computer architectures for CPU and GPU. (B) Calculation steps for the GPU implementation of classical density functional theory (cDFT). Solid rectangles represent operations on CPU, rounded rectangles stand for operations executed with GPU, and the dashed rectangles are for the data transfer between CPU and GPU.

CPU (central processing unit) and GPU (graphic processing unit) have different computer architectures, which dictate how they handle computational tasks and what types of computation they are particularly good at. As shown in Figure 3.1(A), CPU is designed for low latency (viz. low response time) operations and is equipped with a few

arithmetic-logic units (ALUs) to process mathematical and logical operations. By contrast, GPU is designed for high-throughput operations. In comparison to CPU, GPU often has many more ALUs but requires higher response time. As a result, CPU is good at serial computations while GPU has advantages in parallel computations. For gas adsorption in nanoporous materials, we need construct three-dimensional grids and calculate the numerical values for a large number of cDFT functions (Section 2.1) at each grid point. Because evaluation of these functions can be executed independently during the cDFT iterations (*viz*., in solving eq [3.6]), we can significantly reduce the computational cost by a massively parallelized computational scheme. In our implementation of cDFT, we maximize the GPU performance by parallel calculations of all local properties.

In applications of cDFT to gas adsorption in nanoporous materials, one repeated operation is the evaluation of the convolution integrals. Both the hard-sphere and attractive components of the excess Helmholtz energy involve of several weighted densities. The functional derivatives of these Helmholtz energies need to be calculated in terms of the convolution integrals with different weight functions. For numerical integration in three dimensions, the computational cost for each convolution scales as $O\left(N^2\right)$, where N represent the number of total points. By utilizing the Fast Fourier Transformation (FFT), we can express the convolution as

$$\int f_1(\mathbf{r})\, f_2(\mathbf{r}-\mathbf{r}')\, d\mathbf{r}' = \mathcal{F}^{-1}\left\{\mathcal{F}\left[f_1(\mathbf{r})\right]\mathcal{F}\left[f_2(\mathbf{r})\right]\right\} \tag{3.19}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ stand for the forward and backward Fourier transforms, respectively. Application of FFT reduces the computational cost to $O(N \log N)$.

In this work, we have implemented the FFT convolution for both CPU and GPU versions of cDFT calculations. Specifically, fftw3 is used for the CPU implementation while *cuFFT* library, which is designed for high-performance FFT on Nvidia GPU, is used in GPU implementation.[34, 35] All nanoporous materials considered in this work are crystals with known atomic composition and configuration. As a result, the density profiles satisfy periodic boundary conditions and can be easily implemented with fast Fourier transformation. The computational domain includes several unit cells of the crystalline material such as the length of each axis is more than two times the cutoff distance. The single-site Lennard-Jones (LJ) 12-6 model is used to describe the intermolecular potential between gas molecules and their interactions with individual atoms of the nanoporous materials.

All nanoporous materials investigated in this work are crystalline solids. Each material has known atomic composition and the configuration is assumed to be rigid. The Lorentz-Berthelot mixing rule is used to calculate the binary parameters for different species. Table 3.1 lists the Lennard-Jones (LJ) parameters for different gas molecules considered in this work. For all atoms from the nanoporous materials, the force field parameters are obtained from the universal force field (UFF).[36] A cutoff of 12.9 Å is applied to the calculation of all intermolecular interactions. The intermolecular potential is truncated and shifted to zero at the cutoff distance. The computational domain consists

of several duplicated crystal unit cells such that each axis is more than two times the cutoff distance.

In all cDFT calculations, the grid size (step length) is set to be around 0.2σ. As demonstrated in our earlier publications[8-10], the cDFT predictions showed good agreement with the results from the grand canonical Monte Carlo (GCMC) simulation.

Table 3.1. The Lennard-Jones parameters and grid size for gases considered in this work

adsorbate	ε/ k_BT (K)	σ (Å)	grid size (Å)
H_2	36.7	2.96	0.6
N_2	93.98	3.572	0.7
CH_4	148	3.73	0.75

3.2.3 Conjugate Gradient Descent Method

Most cDFT calculations are implemented by using the Picard iteration to solve the self-consistent Euler-Lagrange equation or for the direct minimization of the grand potential.[37, 38] While the Picard iteration is straightforward to use in serial computing, the computational efficiency depends on the judicious choice of the mixing parameter, which is used to update the density profile. The mixing parameter can be tricky to optimize, depending on the grid resolution and formulation of the Helmholtz energy functional. A large mixing parameter can help the grand potential rapidly descent at the beginning of iteration but may never lead to convergence due to bouncing back and forth around the minimum. While a small mixing parameter would almost guarantee the convergence, it would sacrifice the computational efficiency because it requires many iterations. The conjugate gradient method provides an alternative to minimize the grand potential. Because the density profile is updated according to its gradient at each step, the conjugate gradient method is often more efficient than the Picard iteration albeit the

evaluation of the gradients requires additional calculations. In this work, we use CG_DESCENT in both CPU and GPU implementations of cDFT.[39]

Figure 3.1(B) shows the overall computational flow chart in our new implementation of cDFT. As the memory on GPU (device) is physically separated from CPU (host), certain variables need to be transferred from CPU to GPU in order for the GPU kernels to conduct calculations. In general, the bandwidth of data transfer between the device and the host is much smaller than that within the device or the host.[4, 19] As a result, excellent GPU performance can be achieved by minimizing host-device data transferring, i.e., only for the essential input variables, the density profile, the grand potential and its functional derivative. The input variables must be transferred between the host and the device as GPU is not able to access the input file on the hard disk directly. The density profile, the grand potential, and the derivative of grand potential must be transferred in each iteration step because these variables are used in the CG_DESCENT package. The time consuming step could be avoided if CG_DESCENT could be executed on GPU. To achieve better performance, we allocate pinned memory for those variables that are frequently transferred between the host and the device. It is worth mentioning that a reduction sum kernel can be implemented for the grand potential. Because only the sum instead of its local value at every grid point needs to be transferred, the sum kernel method also helps achieve a better performance. All the variables are declared as double precision to achieve a better numerical accuracy.

In the GPU implementation of cDFT, we set the maximum threads per block (1024 for Nvidia Tesla P100 GPU). The initial density profile for all cDFT calculations is

generated from the ideal Boltzmann distribution. For the conjugate gradient descent method, the convergence criterion is set as the gradient to be 1×10^{-3} for both CPU and GPU implementations. For the Picard iteration, we assume that the procedure is converged when the difference between two density values at all grid point is smaller than 1×10^{-6} (#/Å^{-3}). The mixing parameter is set to be 0.01 for the Picard iteration, which is the largest mixing parameter would guarantee the convergence for all materials over the range of pressure considered in this work.

3.3 Results and Discussion

In this section, we compare the CPU and GPU implementations of classical density function theory (cDFT) for predicting gas adsorption in terms of both computational performance and numerical quality. Here the computational performance is evaluated by the time cost of calculation, while the numerical quality is referred to the accuracy of the adsorption amount predicted by CPU and GPU calculations. The speedup factor is calculated from the GPU time divided by the CPU time. For simplicity, we use the mean-field approximation (MFA) for the excess Helmholtz energy due to intermolecular attraction unless other specifications are given. Whereas cDFT calculations yield the same results for the same implementation, the initialization time for the CUDA program can vary significantly thus affect the speedup factor. Except for cases related to high-throughput screening, all calculations are repeated five times to minimize the variance of GPU initialization latency. For most results reported in this work, the standard deviation is much smaller than the mean value, and the error bar is within the

symbol size. For the case study of hydrogen and nitrogen adsorption, we use one particular metal-organic framework (MOF-5) as adsorbent material.[40]

3.3.1 Case I: H_2 Adsorption in MOF-5

We first benchmark the numerical results for predicting hydrogen adsorption in nanoporous materials as hydrogen storage represents an important topic of ongoing research. Despite extensive computational and experimental investigations over the past decade, no material has been found to meet the hydrogen-storage target set by the US Department of Energy.[41-44] The inverse design of nanoporous materials requires efficient computational tools for accurate and rapid evaluation of hydrogen adsorption. Here we use MOF-5 as an example because it is a model nanoporous material that has been extensively studied in the literature.[4]

Figure 3.2 shows the adsorption isotherms for hydrogen adsorption in MOF-5 at cryogenic and room temperatures and the related computational time for GPU and CPU implementation of cDFT. In both conditions, the GPU implementation shows excellent numerical quality in comparison to the CPU implementation. In our earlier work, we demonstrated that the CPU implementation of cDFT is much faster than conventional simulation methods (e.g. Monte Carlo) for predicting gas adsorptions.[8] The GPU implementation can further raise the computational speed by about 40 times. As shown in Figure 3.1(C) and (D), it takes less than 2 seconds for a Tesla P100 GPU to calculate the adsorption amount in MOF-5 for hydrogen in a given bulk state over a broad range of pressure at both 77 K and 298 K. Interestingly, the computational time is virtually

independent of pressure in GPU implementation of cDFT but it fluctuates and slightly increases with pressure in CPU implementation.

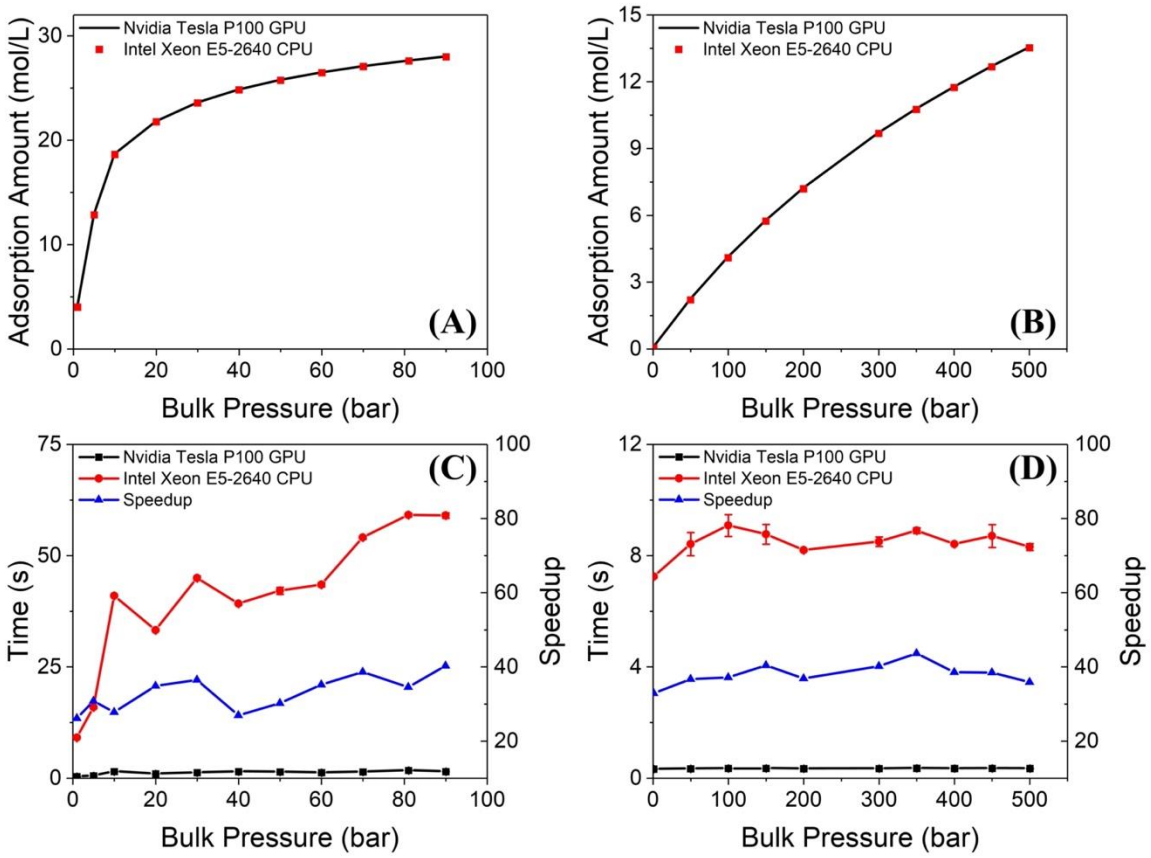

Figure 3.2. Hydrogen adsorption in MOF-5 calculated with Nvidia Tesla P100 GPU and Intel Xeon E5-2640 CPU versus the gas pressure in the bulk at (A) 298 K and (B) 77K. The computational time with Nvidia Tesla P100 GPU and Intel Xeon E5-2640 CPU versus pressure at (C) 298 K and (D) 77K. Also shown here are the GPU speedups compared with CPU.

Nanoporous materials cover a broad range of unit cell sizes and pore volumes. With a fixed grid distance, the computational cost increases linearly with the system volume. In addition, the system size would also depend on different kinds of adsorbates and different types of gas-solid interactions. For systems involving long-range interactions, we may consider a supercell that contains multiple unit cells such that the

long-range interactions can be accurately evaluated. Besides, a large supercell is often required for amorphous systems with crystalline materials with defects. To compare how the GPU and CPU implementations of cDFT scale with the system size, we consider hydrogen adsorption in MOF-5 by changing the grid distance at pressure of 1 bar and two temperatures discussed above (77 K and 298 K). As shown in Figure 3.3, the time cost increases cubically with the system size (*viz*., the number of grid points) for both CPU and GPU. With the increase of system size, the parallel overhead becomes insignificant compares to the time used for calculations.

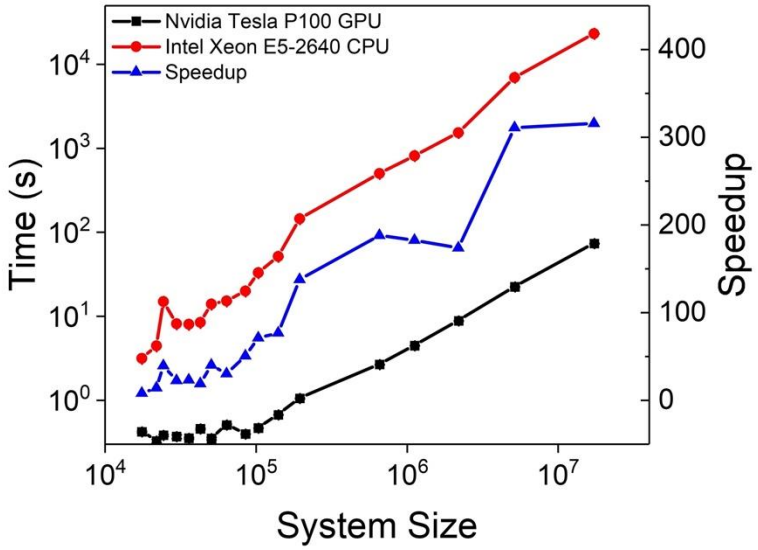

Figure 3.3. Computational time for predicting hydrogen adsorption in MOF-5 calculated with Nvidia Tesla P100 GPU and Intel Xeon E5-2640 CPU versus the system size (i.e., the total number of grid points). In call cases, the pressure is fixed at 1 bar and temperature is (A) 298 K and (B) 77K. The blue lines correspond to the speedup of GPU compared to CPU.

Massive parallelization has no obvious advantage when the system size is small because not all the threads in GPU are put into operation. For a small system, not only is the time saved in calculation of convolutions via massive parallel on GPU not significant,

GPU implementation also comes with the extra burden to transfer the memory, which slows down the speed of computation. Figure 3.3 shows that the speedup factor increases monotonically with the system size and reaches almost 300 when the number of grid points is around 5×10^6. This number is typically used in our cDFT prediction of gas adsorption in nanoporous materials. The speedup factor at 77 K is higher than that at 298 K because the density profile deviates much further from the ideal Boltzmann distribution at lower temperature, requiring more iterations toward convergence. When we consider a more complicated functional for the excess Helmholtz energy, such as the weighted-density approximation (WDA) for van der Waals attraction, the GPU implementation leads to a slightly higher speedup factor at different system sizes due to the adoption of more complicated mathematical operations in WDA (shown in Figure 3.4).

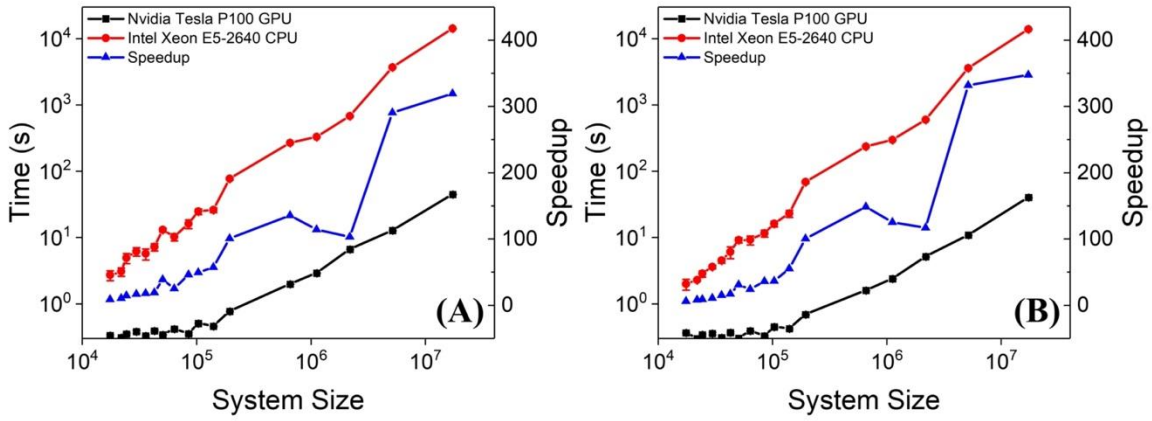

Figure 3.4. Computational time for hydrogen adsorption in MOF-5 at 77 K and 1 bar calculated with Nvidia Tesla P100 GPU and Intel Xeon E5-2640 CPU. The system size is defined in terms of the total number of grid points. Here cDFT is formulated with the weighted-density approximation, and the GPU speedup is compared to the computational time with CPU.

Figure 3.5 shows the percentage of computational time used in the system setup and in iteration for CPU and GPU implementations of cDFT at pressure of 1 bar and

temperatures 77 K and 298 K. At both the cryogenic temperature and room temperature, the percentage of time used in iteration and memory usage (shown in Figure 3.6) increases with the system size. Within each iteration step, the most expensive operation is devoted to calculating the convolutions affiliated with the local density and the local excess chemical potential. As shown in Figure 3.5, the computational cost of calculating the convolution integrals rises rapidly when the system size increases. With the help of massive parallelization on GPU, the computational cost is significantly reduced. For systems pertinent to gas adsorption in nanoporous materials, the GPU implementation results in a factor of more than 300 speedup on Nvidia Tesla P100 GPU in comparison with the serial implementation on Intel E5-2640 CPU.

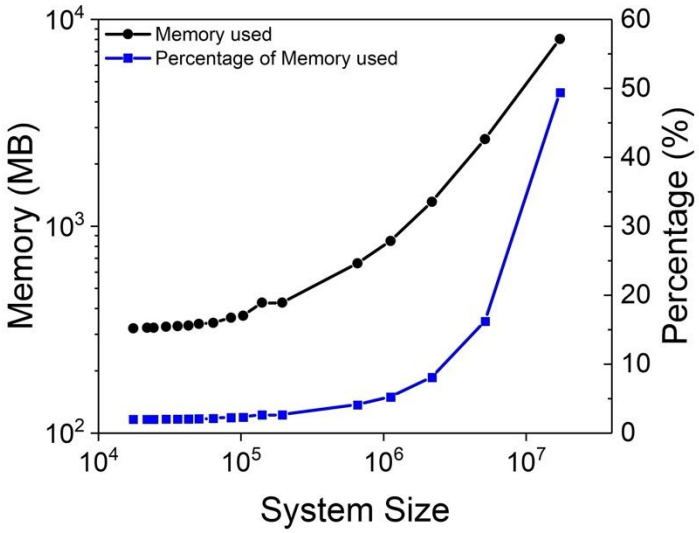

Figure 3.5. Percentage of computer time used for initialization and iteration during the cDFT calculation of hydrogen adsorption in MOF-5 versus the system size. In all cases, the gas pressure is 1 bar and temperature is (A) 298 K and (B) 77 K.

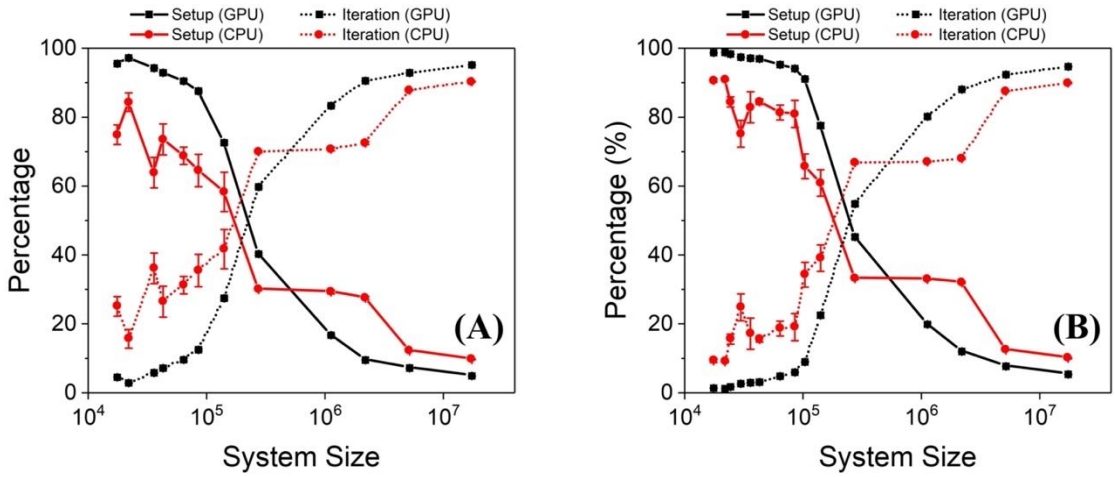

Figure 3.6. Memory usage for hydrogen adsorption in MOF-5 at 77 K and 1 bar calculated with Nvidia Tesla P100 GPU. The system size is defined in terms of the total number of grid points.

3.3.2 Case II: N_2 Adsorption at 77K

Nitrogen adsorption at 77 K is a routine procedure for experimental characterization of porous materials in terms of both the surface area and pore size distribution, and cDFT calculation is instrumental for the data analysis.[45] From the computational perspective, the existence of micropores (< 2 nm) and mesopores (2 to 50 nm) would induce vapor-liquid transition which makes the cDFT calculation much more expensive and challenging to find the density profiles at equilibrium.[38, 46, 47] Therefore, nitrogen adsorption at 77 K provides another good system to benchmark the advantages of GPU implementation in comparison to CPU.

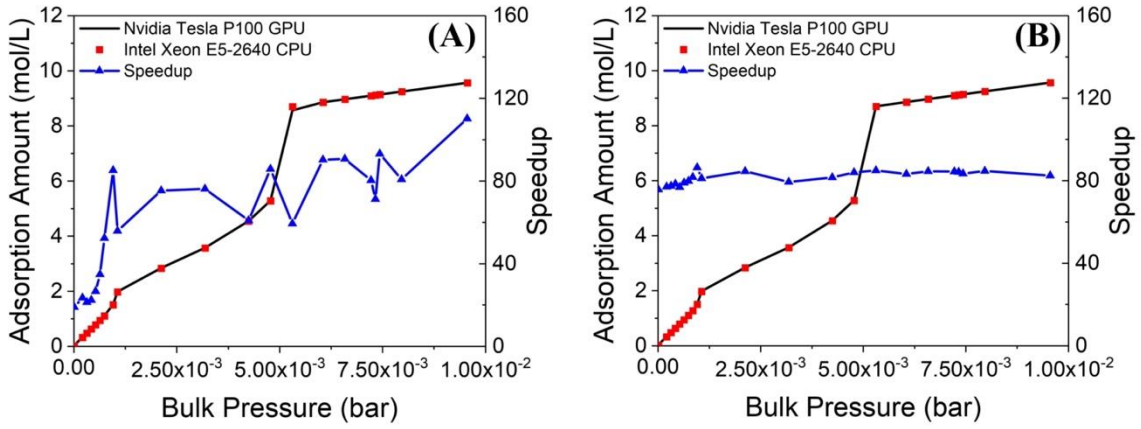

Figure 3.7. The nitrogen adsorption isotherm for MOF-5 calculated with Nvidia Tesla P100 GPU and with Intel Xeon E5-2640 CPU. The blue lines show the GPU speedup compared to CPU. Different numerical schemes are used in grand potential minimization (A) the conjugate gradient descent method, and in (B) the Picard iteration.

Figure 3.7 shows nitrogen adsorption isotherm in MOF-5 at 77 K where the coexistence of vapor-liquid is expected. Figure 3.7 (A) and (B) are calculated with the conjugate gradient descent and Picard iterations, respectively. In terms of the adsorption amount, the numerical quality of GPU implementation is as good as CPU implementation for both minimization schemes. However, the Picard iteration takes many more iterations to converge than the conjugate gradient descent iteration (shown in Figure 3.8). For the Picard iteration, more numerical iterations are needed when the system is near the vapor-liquid transition inside the pores and high loadings of adsorbent. The speedup factor between GPU and CPU for the Picard iteration is almost a constant (around 80) because, within each iteration step, the same calculation is executed and the speedup factor only depends on the system size. For the conjugate gradient descent method, CG_DESCENT has the options to evaluate the grand potential or the derivatives of grand potential or both variables within each iteration step, which results in the fluctuations of the speedup

factor. Evaluation of the derivatives of the grand potential requires more time for memory transfer than that only for the grand potential. As the conjugate gradient descent method takes a smaller number of iteration steps to converge, the parallelization overhead diminishes the advantages of the GPU implementation at low loading. With the increase of the gas pressure, GPU implementation may lead to a speedup more than two orders of magnitude (Fig. 5A).

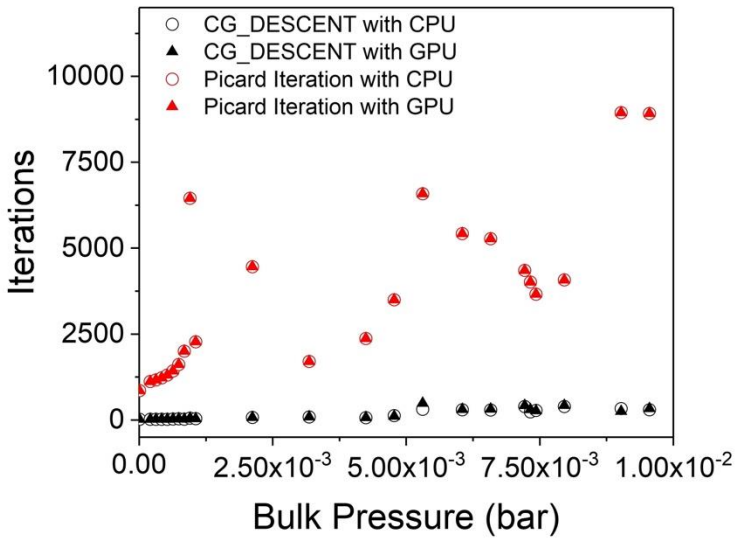

Figure 3.8. The number of iterations to predict nitrogen adsorption in MOF-5 at 77 K with the conjugate gradient descent method and with the Picard iteration.

3.3.3 Case III: High-Throughput Screening

In the final case study, we compare the difference between GPU and CPU implementations of cDFT for massive computations of gas adsorption in porous materials. In recent years, high-throughput screening has been widely used to identify the best candidates in a given structural database for a specific application.[48-50] In comparison to trial-and-error or the combinatory approach, high-throughput computational modeling

could save enormous experimental effort for materials discovery. For gas adsorption in nanoporous materials, high-throughput screening requires efficient predictions of the adsorption amount for a large set of materials over a broad range of conditions. The task could be computationally prohibitive if one considers a wide variety of thermophysical properties important for practical applications.

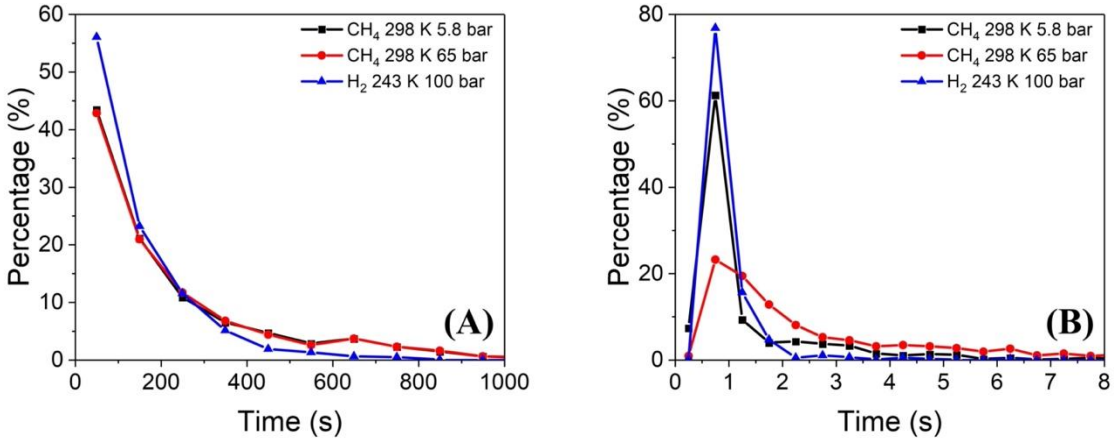

Figure 3.9. Distributions of computational time for the cDFT predictions of CH_4 and H_2 adsorption in 800 CoRE metal-organic frameworks. (A) Intel Xeon E5-2640 CPU, and (B) Nvidia Tesla P100 GPU.

Figure 3.9 shows the distributions of computational time for the cDFT predictions of methane and hydrogen adsorptions in 800 nanoporous materials from the Computation-Ready, Experimental (CoRE) metal-organic frameworks Database.[51] Compared with the CPU implementation of cDFT, implementation with GPU can significantly reduce the computational cost. For all three cases tested here, more than 50% of calculation can be accomplished within two seconds with GPU. For methane adsorption, the calculation becomes more computationally expensive when the system is further away from ideality (5.8 bar versus 65 bar), which is consistent with our previous

report.[8] However, the scale of difference for the computational cost in different conditions is almost negligible for the GPU implementation of cDFT. Empowered with the massively paralleled implementation of cDFT on GPU, we expect that the computational procedure will be valuable for more comprehensive evaluation of the diverse properties of nanoporous materials for gas adsorption.

3.4 Conclusions

In this work, we have implemented a massively parallelized algorithm for classical density functional theory (cDFT) calculations with graphic processing unit (GPU) and tested its performance for predicting gas adsorption in nanoporous materials. The GPU implementation leads to significant improvements in terms of both numerical quality and performance speedup in comparison to the serial implementation with central processing unit (CPU). For all the cases considered in this work, the GPU implementation shows outstanding accuracy for predicting the adsorption amount compared to the CPU implementation. For most calculations, the speed of cDFT calculation with GPU is more than two orders of magnitude higher than that with the CPU implementation. The acceleration becomes more significant as the system size increases independent of the specific formalism of the excess Helmholtz energy functional. We expect similar accelerations for more complicated systems such as multicomponent mixtures, polymeric fluids, and systems with electrostatic interactions that are typically much more time consuming for cDFT calculations.

In addition, we showed that the conjugate gradient descent method is more efficient (less iterations) and more robust (less dependent on the choice of mixing factor)

than the Picard iteration for the minimization of grand potential. The difference is evident especially when gas adsorption is accompanied by vapor-liquid transition within the micro and mesopores. For high-throughput screening, the massively parallelized GPU implementation of cDFT is able to predict gas adsorption at different conditions (*viz*., temperature and pressure) within seconds for most nanoporous materials tested. The GPU algorithm renders accurate and extremely fast prediction of gas adsorption in nanoporous materials thus accelerates martials screening and construction of more comprehensive properties database. Development of the more efficient computational methods will eventually pave the way for the inverse design of nanoporous materials for gas storage and separation.

Bibliography

1. Li, J.R., R.J. Kuppler, and H.C. Zhou, *Selective gas adsorption and separation in metal-organic frameworks.* Chem Soc Rev, 2009. **38**(5): p. 1477-504.

2. Feng, X., X. Ding, and D. Jiang, *Covalent organic frameworks.* Chem Soc Rev, 2012. **41**(18): p. 6010-6022.

3. Morris, R.E. and P.S. Wheatley, *Gas storage in nanoporous materials.* Angew Chem Int Ed Engl, 2008. **47**(27): p. 4966-81.

4. Kim, B., S. Lee, and J. Kim, *Inverse design of porous materials using artificial neural networks.* Sci Adv, 2020. **6**(1): p. eaax9324.

5. Lee, S., B. Kim, and J. Kim, *Predicting performance limits of methane gas storage in zeolites with an artificial neural network.* Journal of Materials Chemistry A, 2019. **7**(6): p. 2709-2716.

6. Thornton, A.W., et al., *Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage.* Chem Mater, 2017. **29**(7): p. 2844-2854.

7. Boyd, P.G., Y. Lee, and B. Smit, *Computational development of the nanoporous materials genome.* Nature Reviews Materials, 2017. **2**(8): p. 1-15.

8. Fu, J., et al., *Density Functional Methods for Fast Screening of Metal Organic Frameworks for Hydrogen Storage.* J Phys Chem C, 2015. **119**(10): p. 5374-5385.

9. Fu, J., Y. Tian, and J.Z. Wu, *Classical density functional theory for methane adsorption in metal-organic framework materials.* AIChE J, 2015. **61**(9): p. 3012-3021.

10. Liu, Y., et al., *Density functional theory for adsorption of gas mixtures in metal-organic frameworks.* J Phys Chem B, 2010. **114**(8): p. 2820-7.

11. Meng, H. and C.Y. Wang, *Large-scale simulation of polymer electrolyte fuel cells by parallel computing.* Chemical Engineering Science, 2004. **59**(16): p. 3331-3343.

12. Stone, J.E., et al., *GPU-accelerated molecular modeling coming of age.* J Mol Graph Model, 2010. **29**(2): p. 116-25.

13. Brown, W.M., et al., *Implementing molecular dynamics on hybrid high performance computers - short range forces.* Computer Physics Communications, 2011. **182**(4): p. 898-911.

14. Brown, W.M., et al., *Implementing molecular dynamics on hybrid high performance computers - Particle-particle particle-mesh.* Computer Physics Communications, 2012. **183**(3): p. 449-459.

15. Acun, B., et al., *Scalable Molecular Dynamics with NAMD on the Summit System.* IBM J Res Dev, 2018. **62**(6): p. 1-9.

16. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

17. Jia, W.L., et al., *Fast plane wave density functional theory molecular dynamics calculations on multi-GPU machines.* Journal of Computational Physics, 2013. **251**: p. 102-115.

18. Nitsche, M.A., et al., *GPU Accelerated Implementation of Density Functional Theory for Hybrid QM/MM Simulations.* J Chem Theory Comput, 2014. **10**(3): p. 959-67.

19. Stopper, D. and R. Roth, *Massively parallel GPU-accelerated minimization of classical density functional theory*. J Chem Phys, 2017. **147**(6): p. 064508.

20. Wu, J., *Density functional theory for liquid structure and thermodynamics*. 2008.

21. Rappe, A.K., et al., *Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations.* J Am Chem Soc, 1992. **114**(25): p. 10024-10035.

22. Johnson, J.K., J.A. Zollweg, and K.E. Gubbins, *The Lennard-Jones Equation of State Revisited.* Mol Phys, 1993. **78**(3): p. 591-618.

23. Ye, Z., et al., *Density functional theory of homopolymer mixtures confined in a slit.* J Chem Phys, 2006. **125**(12): p. 124705.

24. Li, Z., D. Cao, and J. Wu, *Density-functional theory and Monte Carlo simulation for the surface structure and correlation functions of freely jointed Lennard-Jones polymeric fluids.* J Chem Phys, 2005. **122**(17): p. 174708.

25. Liu, Y., et al., *Development of a density functional theory in three-dimensional nanoconfined space: H2 storage in metal-organic frameworks.* J Phys Chem B, 2009. **113**(36): p. 12326-31.

26. Siderius, D.W. and L.D. Gelb, *Predicting Gas Adsorption in Complex Microporous and Mesoporous Materials Using a New Density Functional Theory of Finely Discretized Lattice Fluids.* Langmuir, 2009. **25**(3): p. 1296-1299.

27. Yu, Y.X. and J.Z. Wu, *Structures of hard-sphere fluids from a modified fundamental-measure theory.* J Chem Phys, 2002. **117**(22): p. 10156-10164.

28. Roth, R., et al., *Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version.* Journal of Physics-Condensed Matter, 2002. **14**(46): p. 12063-12078.

29. Henderson, D., *Monte Carlo and perturbation theory studies of the equation of state of the two-dimensional Lennard-Jones fluid.* Mol Phys, 1977. **34**(2): p. 301-315.

30. Cotterman, R.L., B.J. Schwarz, and J.M. Prausnitz, *Molecular Thermodynamics for Fluids at Low and High-Densities .1. Pure Fluids Containing Small or Large Molecules.* Aiche Journal, 1986. **32**(11): p. 1787-1798.

31. Tarazona, P. and R. Evans, *A Simple Density Functional Theory for Inhomogeneous Liquids - Wetting by Gas at a Solid Liquid Interface.* Molecular Physics, 1984. **52**(4): p. 847-857.

32. Yu, Y.X., *A novel weighted density functional theory for adsorption, fluid-solid interfacial tension, and disjoining properties of simple liquid films on planar solid surfaces.* J Chem Phys, 2009. **131**(2): p. 024704.

33. Wu, J. and Z. Li, *Density-functional theory for complex fluids.* Annu Rev Phys Chem, 2007. **58**: p. 85-112.

34. Frigo, M. and S.G. Johnson, *The design and implementation of FFTW3.* Proceedings of the Ieee, 2005. **93**(2): p. 216-231.

35. NVIDIA, C., *Fast Fourier Transform Library (cuFFT).* 2013: Accessed.

36. Rappe, A.K., et al., *UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations.* Journal of the American Chemical Society, 1992. **114**(25): p. 10024-10035.

37. Roth, R., *Fundamental measure theory for hard-sphere mixtures: a review.* J Phys Condens Matter, 2010. **22**(6): p. 063102.

38. Edelmann, M. and R. Roth, *A numerical efficient way to minimize classical density functional theory.* J Chem Phys, 2016. **144**(7): p. 074105.

39. Hager, W.W. and H.C. Zhang, *Algorithm 851: CG DESCENT, a conjugate gradient method with guaranteed descent.* ACM Transactions on Mathematical Software, 2006. **32**(1): p. 113-137.

40. Rosi, N.L., et al., *Hydrogen storage in microporous metal-organic frameworks.* Science, 2003. **300**(5622): p. 1127-1129.

41. Morris, R.E. and P.S. Wheatley, *Gas storage in nanoporous materials.* Angewandte Chemie International Edition, 2008. **47**(27): p. 4966-4981.

42. Basdogan, Y. and S. Keskin, *Simulation and modelling of MOFs for hydrogen storage.* CrystEngComm, 2015. **17**(2): p. 261-275.

43. Broom, D., et al., *Outlook and challenges for hydrogen storage in nanoporous materials.* Applied Physics A, 2016. **122**(3): p. 151.

44. Gao, Y., X.J. Wu, and X.C. Zeng, *Designs of fullerene-based frameworks for hydrogen storage.* Journal of Materials Chemistry A, 2014. **2**(16): p. 5910-5914.

45. Landers, J., G.Y. Gor, and A.V. Neimark, *Density functional theory methods for characterization of porous materials.* Colloids and Surfaces a-Physicochemical and Engineering Aspects, 2013. **437**: p. 3-32.

46. He, S., et al., *Competitive coordination strategy for the synthesis of hierarchical-pore metal–organic framework nanostructures.* Chemical science, 2016. **7**(12): p. 7101-7105.

47. Makowski, W., et al., *Unusual adsorption behavior of volatile hydrocarbons on MOF-5 studied using thermodesorption methods.* Thermochimica Acta, 2014. **587**: p. 1-10.

48. Zhou, M., A. Vassallo, and J. Wu, *Toward the inverse design of MOF membranes for efficient D2/H2 separation by combination of physics-based and data-driven modeling.* Journal of Membrane Science, 2020. **598**.

49. Canepa, P., et al., *High-throughput screening of small-molecule adsorption in MOF.* Journal of Materials Chemistry A, 2013. **1**(43): p. 13597-13604.

50. Tong, M., et al., *High-throughput computational screening and design of nanoporous materials for methane storage and carbon dioxide capture.* Green Energy & Environment, 2018. **3**(2): p. 107-119.

51. Chung, Y.G., et al., *Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019.* Journal of Chemical and Engineering Data, 2019. **64**(12): p. 5985-5998.

Chapter 4. Modeling Multicomponent Gas Adsorption in Nanoporous Materials with Two Versions of Nonlocal Classical Density Functional Theory

In this chapter, two versions of nonlocal classical density functional theory (cDFT) have been proposed to predict multicomponent gas adsorption in nanoporous materials by using the Lennard-Jones model for gas mixture and the universal force field for the adsorbents. With the modified fundamental measure theory to describe the short-range repulsions or volume-exclusion effects, one version of cDFT adopts the mean-field approximation for van der Waals attraction (here referred to as cDFT-MFA) as commonly used in porous materials characterization, the other version accounts for long-range correlations through a weighted-density approximation (cDFT-WDA). For a number of gas mixtures in MOF-5 (without sub-pores inaccessible to gas molecules), the adsorption isotherms predicted from cDFT-WDA are quantitatively consistent with results from grand canonical Monte Carlo simulation, while cDFT-MFA systematically underestimates the adsorption due the neglect of correlation effects. Nevertheless, both versions of cDFT outperform the ideal adsorbed solution theory (IAST) at high pressure. Because IAST predicts mixture adsorption using only single-component data, it fails to capture the selective behavior arising from asymmetric interactions among different chemical species. The cDFT calculations are implemented with massively parallel GPU-accelerated algorithms to achieve rapid yet accurate predictions of multicomponent adsorption isotherms with full atomistic details of the adsorbent materials. This work thus provides a theoretical basis for the computational design of adsorption-based separation process as well as for screening and data-driven inverse design of nanoporous materials.

4.1 Introduction

Selective adsorption represents one of the most important industrial separation processes, with widespread applications such as desulfurization in petroleum refining, H_2 and CH_4 purification and CO_2 capture.[1-5] As chemical separation takes about half of the industrial energy use in the US, development of more efficient separation processes is desired to reduce the energy cost.[6] Conventionally, adsorption-based separation processes are categorized into pressure swing adsorption (PSA) and temperature swing adsorption (TSA). In both cases, the adsorbent material is a dominating factor determining the separation efficiency and operation cost. Therefore, adsorbent design is of great importance in optimizing industrial separation processes.

Nanoporous materials, such as metal-organic frameworks (MOFs) and covalent-organic frameworks (COFs), are promising for adsorption separation because of their excellent mechanic stability, large surface area and adjustable pore geometry.[7-9] More importantly, both adsorption capacity and selectivity can be tailored for specific applications by altering the secondary building blocks of such materials.[10, 11] Compared with trial-and-error syntheses, a data-driven approach enables the inverse design of nanoporous materials over a much larger chemical space.[12-14] However, data-driven methods require a large set of data for multicomponent adsorption isotherms. Such data are scarce due to the time-consuming experimental procedures and technical difficulties in controlling adsorption over a wide range of thermodynamic conditions.[15, 16]

Thermodynamic modeling of multicomponent adsorption isotherms is mostly based on the ideal adsorbed solution theory (IAST).[17-20] Although IAST adopts ideal gas and ideal solution models for the bulk and adsorbed phases, respectively, it is able to predict the adsorption isotherms of many gas mixture systems (e.g., CH_4/C_2H_6, C_2H_4/CO_2 and CO/O_2) with pure component adsorption isotherms up to moderate pressures. Many attempts have been made to modify IAST by considering surface heterogeneity and the distinct properties of gas molecules such as polarity, size and interaction energy.[21-23] Although the modified IAST models account for more realistic gas-absorbent interactions and improve the numerical performance for specific systems, they often entail extra fitting parameters thus hamper transferability. Moreover, the modified models do not provide a consistent description of the thermodynamic non-ideality of gas mixtures in the bulk and inside the adsorbent materials, which also hinders their applications at high pressure.

Molecular simulation and statistical-mechanical methods, such as grand canonical Monte Carlo (GCMC) simulation and classical density functional theory (cDFT), are main alternates to IAST for predicting mixture adsorption in nanoporous materials.[24-27] For single-component adsorption, adsorption isotherms predicted by GCMC and cDFT are generally in good agreement with each other.[28-31] In contrast to the IAST models, GCMC and cDFT are able to predict adsorption isotherms for single- and multi-component systems alike. More importantly, they offer atomistic details useful for the computational design of adsorbent materials. From a practical perspective, neither GCMC nor cDFT is perfect. In addition to intrinsic errors introduced by the atomistic

models, GCMC is not suitable for massive calculations due to the excessive computational burden. Conversely, cDFT is often less accurate because it relies on the formulation of the excess Helmholtz energy functional.[32-34] While highly reliable models for the excess Helmholtz energy functional have been developed over the past few decades to describe single-component adsorption isotherms,[35] application of cDFT to adsorption in gas mixtures is mostly limited to slit pores[36, 37] and virtually nonexistent for three dimensional structures and materials even though they are most relevant to practical applications. To bridge the gap, this work introduces two versions of cDFT for gas mixtures represented by the Lennard-Jones (LJ) model. The LJ model is often used to represent the thermodynamic properties of gas mixtures and relevant to industrial applications such as adsorption-based separation of noble gases and methane purification.[26, 38] Both versions of cDFT are based on the modified fundamental measure theory (MFMT) that is naturally applicable to systems containing particles of different sizes.[39] With the short-range repulsions or volume-exclusion effects described by MFMT, one version of cDFT adopts the mean-field approximation (herein referred to as cDFT-MFA) to represent the excess Helmholtz energy due to van der Waals attractions, and the other accounts for the correlation effects with a weighted density approximation (herein referred to as cDFT-WDA). For single-component systems, cDFT-MFA is essentially equivalent to nonlocal DFT (NLDFT) that is conventionally used for the characterization of porous materials by gas adsorption. Both versions of cDFT are able to generate mixture adsorption isotherms in good agreement with GCMC simulation for a wide variety of systems, while cDFT-WDA yields slightly better numerical

performance because it considers correlation effects. Compared with IAST, both versions of cDFT predict more accurate adsorption isotherms especially at high pressure where gas-gas interactions and correlation effects become more significant. Because cDFT contains atomistic details for adsorbent materials and can be implemented through massively paralleled GPU programming, it empowers the rapid construction of large database potentially useful for the inverse design of nanoporous materials for gas separation.

4.2 Methods and Models

4.2.1 Classical Density Functional Theory (cDFT)

In principle, classical density functional theory (cDFT) is able to predict the thermodynamic properties of any macroscopic system at equilibrium.[40-42] In its application to multi-component gas adsorption with the Lennard-Jones (LJ) model, the grand potential is minimized with respect to the density profiles

$$\frac{\delta\Omega[\boldsymbol{\rho}(\mathbf{r})]}{\delta\rho_i(\mathbf{r})} = 0 \qquad \left(i = 1, 2, ..., N\right) \tag{4.1}$$

where N is the number of chemical species in the gas mixture, $\boldsymbol{\rho}(\mathbf{r}) = \left[\rho_1(\mathbf{r}), \rho_2(\mathbf{r}), ..., \rho_N(\mathbf{r})\right]$ is a short notation for the density profiles of gas molecules, and $\mathbf{r} = (x, y, z)$ is the center-of-mass position for each gas molecule. More specifically, the grand potential is often written as

$$\Omega[\boldsymbol{\rho}(\mathbf{r})] = F[\boldsymbol{\rho}(\mathbf{r})] + \sum_{i=1}^{N} \int [V_i^{ext}(\mathbf{r}) - \mu_i^{bulk,mixture}] \rho_i(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{4.2}$$

where F is the intrinsic Helmholtz energy functional, V_i^{ext} stands for the external potential for component i, i.e., the potential energy on a molecule of species i due to its interaction with the absorbent, and $\mu_i^{bulk,mixture}$ represents the chemical potential for component i in the bulk phase. In this work, we use the universal force field (UFF) to describe gas interaction with nanoporous materials.[43] For the gas molecules considered in this work, the force-field parameters are provided in Support Information (SI). The modified Benedict-Webb-Rubin (MBWR) equation of state is used to calculate the chemical potentials of all species for the gas mixture in the bulk phase.[44] The intrinsic Helmholtz energy can be divided into an ideal part and an excess:

$$F[\boldsymbol{\rho}(\mathbf{r})] = F^{id}[\boldsymbol{\rho}(\mathbf{r})] + F^{ex}[\boldsymbol{\rho}(\mathbf{r})] \tag{4.3}$$

With the gas molecules represented by the LJ model, the ideal part is exactly known

$$F^{id}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \sum_{i=1}^{N} \int \{\ln[\rho_i(\mathbf{r})\Lambda_i^3] - 1\}\rho_i(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{4.4}$$

where k_B stands for the Boltzmann constant, T is the absolute temperature, and Λ_i represents the thermal wavelength of component i.

One essential task of all cDFT calculations is to formulate an excess Helmholtz energy functional that is reliable for the specific system under consideration. According to the LJ model, the excess Helmholtz energy can be split into contributions due to short-range repulsion and long-range attraction. The former is often represented by the hard-sphere model, F^{hs}, and a perturbation term F^{attr} is applied to account for van der Waals attractions

$$F^{ex}[\boldsymbol{\rho}(\mathbf{r})] = F^{hs}[\boldsymbol{\rho}(\mathbf{r})] + F^{attr}[\boldsymbol{\rho}(\mathbf{r})] \tag{4.5}$$

As well documented, the excess Helmholtz energy of a hard-sphere system can be accurately described by the modified fundamental measure theory (MFMT)[32, 45]

$$F^{hs}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \int \Phi^{hs}[n_\alpha(\mathbf{r})]\, \mathrm{d}\mathbf{r} \tag{4.6}$$

where

$$\Phi^{hs} = -n_0 \ln(1-n_3) + \frac{n_0 n_2 - \mathbf{n}_{V1}\cdot\mathbf{n}_{V1}}{1-n_3} + \frac{n_3 + (1-n_3)^2 \ln(1-n_3)}{36\pi n_3^2 (1-n_3)^2}\left[(n_2)^3 - 3n_2 \mathbf{n}_{V2}\cdot\mathbf{n}_{V2}\right] \tag{4.7}$$

with

$$\begin{aligned} & n_\alpha(\mathbf{r}) = \sum_{i=1}^{N} \int \rho_i(\mathbf{r}')\, w_i^{(\alpha)}\left(|\mathbf{r}-\mathbf{r}'|\right) \mathrm{d}\mathbf{r}' \\ & \alpha \in \{0, 1, 2, 3, V1, V2\} \end{aligned} \tag{4.8}$$

In eq (4.8), $w_i^{(\alpha)}$ are a set of weight functions characterizing the differential geometry of each spherical particle:

$$\begin{cases} w_i^{(2)}(r) = \pi d_i^2 w_i^{(0)}(r) \\ \qquad\quad = 2\pi d_i w_i^{(1)}(r) \\ \qquad\quad = \delta(d_i/2 - r) \\ w_i^{(3)}(r) = \theta(d_i/2 - r) \\ \mathbf{w}_i^{(V2)}(r) = 2\pi d_i \mathbf{w}_i^{(V1)}(r) = w_i^{(2)}(r)\dfrac{\mathbf{r}}{r} \end{cases} \tag{4.9}$$

where δ denotes the Dirac-delta function, θ is the Heaviside step function, and d_i is the hard-sphere diameter for component i. For all gas molecules considered in this work, the Barker-Henderson theory is used to calculate the hard-sphere diameter from the LJ parameters[46, 47]

$$d_i = \sigma_i \left(\frac{1 + 0.2977 T_i^*}{1 + 0.33163 T_i^* + 0.0010477 T_i^{*2}} \right) \tag{4.10}$$

where $T_i^* = k_B T / \varepsilon_i$, ε_i and σ_i stand for the LJ energy and size parameters of the gas molecule *i*, respectively.

For the attraction part of the excess Helmholtz energy, one convenient choice is that from the mean-field approximation (MFA)

$$F^{MFA} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int\int \rho_i(\mathbf{r}) \rho_j(\mathbf{r}') u_{ij}^{attr}\left(|\mathbf{r} - \mathbf{r}'|\right) \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{4.11}$$

where

$$u_{ij}^{attr}(r) = \begin{cases} 0 & r < \sigma_{ij} \\ 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] & r > \sigma_{ij} \end{cases} \tag{4.12}$$

and the cross parameters are calculated from the Lorentz-Berthelot mixing rule. In this work, the excess Helmholtz energy given by eqs (4.5) to (4.12) is referred to as cDFT-MFA. Whereas MFA is commonly used in cDFT calculations including characterization of porous materials by gas adsorption, it reduces to an equation of state for bulk systems similar to the van der Waals equation. While more accurate formulations are available for one-component LJ fluids,[48] extension of existing formulations to multicomponent systems is theoretically challenging due to the lack of analytical expressions for the bulk correlation functions.

In this work, we account for the correlation effects using van der Waals one-fluid theory (vdW1) and the weighted density approximation (WDA)[33, 34]

$$F^{attr}[\boldsymbol{\rho}(\mathbf{r})] = F^{MFA}[\boldsymbol{\rho}(\mathbf{r})] + F^{cor}[\boldsymbol{\rho}(\mathbf{r})] \tag{4.13}$$

where F^{cor} corresponds to the local correlation Helmholtz energy

$$F^{cor}[\boldsymbol{\rho}(\mathbf{r})] = k_B T \sum_{i=1}^{N} \int \Phi^{cor}[\bar{\rho}(\mathbf{r})]\, \mathrm{d}\mathbf{r} \tag{4.14}$$

The reduced local correlation Helmholtz energy per volume, Φ^{cor}, is approximated by that corresponding to the bulk phase at weighted density

$$\bar{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \frac{3}{4\pi d_i^3} \int \rho_i(\mathbf{r}')\, \theta\left(d_i - |\mathbf{r} - \mathbf{r}'|\right) d\mathbf{r}' \tag{4.15}$$

In the bulk phase, Φ^{cor} can be written in the following form

$$\Phi^{cor}(\rho) = \frac{F_{bulk}^{LJ}(\rho) - F_{bulk}^{hs}(\rho) - F_{bulk}^{MFA}(\rho)}{V} \beta \tag{4.16}$$

where $F_{bulk}^{LJ}(\rho)$ stands for the excess Helmholtz energy of a bulk LJ fluid calculated from the MBWR equation of state,[44] $F_{bulk}^{hs}(\rho)$ denotes the hard-sphere Helmholtz energy according to the Carnahan-Starling equation of state,[49] $F_{bulk}^{MFA}(\rho)$ represents the mean-field Helmholtz energy for the one-component fluid, and $\beta = 1/(k_B T)$. Explicit expressions are available for the hard-sphere and MF excess Helmholtz energies shown in eq (4.16):

$$\beta \frac{F_{bulk}^{hs}[\bar{\rho}(\mathbf{r})]}{V} = \frac{4y - 3y^2}{(1-y)^2} \bar{\rho}(\mathbf{r}) \tag{4.17}$$

$$\beta \frac{F_{bulk}^{MFA}[\bar{\rho}(\mathbf{r})]}{V} = -\frac{16}{9} \pi \beta \varepsilon_x \bar{\rho}(\mathbf{r})^2 \sigma_x^3 \tag{4.18}$$

where $y = \frac{\bar{\rho}(\mathbf{r})\pi d_x^3}{6}$. According to the vdw1 approximation, d_x, ε_x and x_i stand for the hard-sphere diameter, the LJ energy and size parameters for the mixture:

$$d_x^3 = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} d_{ij}^3 \tag{4.19}$$

$$\sigma_x^3 = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \sigma_{ij}^3 \tag{4.20}$$

$$\varepsilon_x = \frac{1}{\sigma_x^3}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\bar{\rho}_i(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \frac{\bar{\rho}_j(\mathbf{r})}{\bar{\rho}(\mathbf{r})} \varepsilon_{ij}\sigma_{ij}^3 \tag{4.21}$$

Combing eqs (4.1) ~ (4.4) leads to the following Euler-Lagrange equations

$$\rho_i(\mathbf{r}) = \rho_{bulk,i} \exp\left[\beta\mu_i^{ex,mixture} - \beta V_i^{ext} - \beta\frac{\delta F^{ex}}{\delta\rho_i(\mathbf{r})}\right] \qquad (i = 1, 2, ..., N) \tag{4.22}$$

From eq (4.22), we can calculate the density profiles of individual species for an *N*-component gas mixture adsorption in nanoporous materials. In comparison with cDFT for single-component systems, the computational cost scales linearly with the number of chemical species in the gas mixture. With an explicit expression for the excess Helmholtz energy functional as given by eqs (4.5) ~ (4.21), we can solve the density profiles using conjugate gradient descent method.[50] As explained in our previous work,[51] the cDFT calculations can be implemented with massively paralleled algorithm through graphic processing unit (GPU). The GPU-accelerated parallel implementation drastically reduces the computational cost thereby empowering potential industrial applications. In this work, all cDFT calculations are carried out with an Nvidia Tesla P100 graphic card. The time

cost is at the scale of few seconds for each cDFT calculation. More information on the computational details is given in Support Information.

4.2.2 Ideal Adsorbed Solution Theory (IAST)

The basic concepts and numerical procedure of ideal adsorbed solution theory (IAST) have been well documented.[17-19] Here, we recapitulate only the key equations for easy reference. With an *N*-component gas mixture inside an adsorbent represented by a two-dimensional system, pure reference states are defined for all chemical species that share the same spreading pressure of the mixture

$$\pi_1 = \pi_2 = \cdots = \pi_N \tag{4.23}$$

For each pure component *i* in equilibrium with its pure bulk phase, the spreading pressure at pressure p_i^0 can be written as:

$$\frac{\pi_i A}{RT} = \int_0^{p_i^0} \frac{n_i^0\left(P_i\right)}{P_i} dP_i \tag{4.24}$$

where *A* is the surface area, *R* stands for the gas constant, *T* represents temperature, and n_i^0 is the adsorption amount for pure component *i* at its own reference state. In analogy with the Lewis fugacity rule, IAST assumes an ideal solution for the two-dimensional system of adsorbed gas molecules. The adsorbed phase would be in equilibrium with a gas mixture following the adsorption analog of Raoult's law

$$P_i = y_i P = x_i p_i^0 \tag{4.25}$$

where *y* is the composition in the bulk phase, *x* represents the composition of the two-dimensional system (viz., the adsorbed gas mixture). With adsorption isotherms for single-component systems as the input, eqs (4.23) ~ (4.25) can be used to determine the

spreading pressure and, consequently, the corresponding adsorption amount for each pure reference state as well as the compositions of the adsorbed phase. If the adsorption isotherm is fitted into a specific adsorption model (e.g. Langmuir adsorption model), eqs (4.23) ~ (4.25) can be solved analytically. Otherwise, it requires iteration to find p_i^0 for each chemical species such that eqs (4.23) and (4.25) are satisfied. An interpolation scheme would be needed to approximate the adsorption amount between the existing points on the adsorption isotherm. The total adsorption amount n_T for gas mixture is finally calculated from the ideal solution assumption

$$\frac{1}{n_T} = \sum_{i=1}^{N} \frac{x_i}{n_i^0} \tag{4.26}$$

At low pressure, Henry's law predicts the adsorption selectivity

$$\frac{x_i}{x_j} = \frac{K_i}{K_j} \frac{y_i}{y_j} \tag{4.27}$$

where K represents Henry's constant. In this work, single-component adsorption isotherms from GCMC simulations are used as the input for the IAST prediction of multi-component adsorption isotherms. Instead of fitting the adsorption isotherm into a specific adsorption model, the direct interpolation of single-component adsorption isotherm (shown in Figure 4.1) is used when calculating adsorption isotherm for mixtures in order to preserve the accuracy of adsorption isotherm at low pressure.

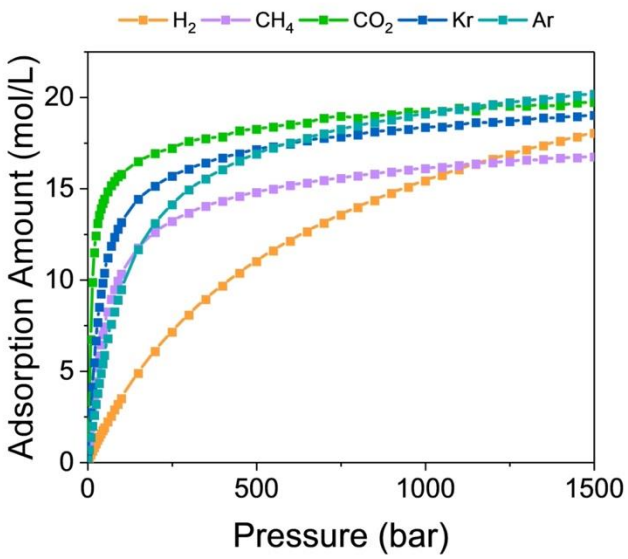

Figure 4.1. Adsorption isotherm for individual gases calculated from GCMC in MOF-5 for Kr and Ar at 297 K and H_2, CH_4, CO_2 at 313.15 K.

4.3 Results and Discussion

In this section, we first compare adsorption isotherms for gas mixtures and the selectivity predicted by the two versions of cDFT along with those from IAST and GCMC simulations. Table 4.1 lists the Lennard-Jones (LJ) parameters for gas molecules considered in this work.[52-56] The grid size in all cDFT calculations is 0.5 Å. The potential is truncated and shifted to zero at 12.9 Å. One unit cell of MOF-5 with periodic boundary condition is used in the calculation. CG_DESCNT package is used to minimize the grand potential with the convergence criteria set as the gradient to be 1×10^{-3}.[50] For the calibration of the excess Helmholtz energy functions, these comparisons are discussed in the context of one particular metal-organic framework (MOF) material – MOF-5. Different from other MOFs, MOF-5 does not contain sub-pores inaccessible to small gas molecules, which may lead to an inaccurate prediction of adsorption by IAST.[26, 57]

Table 4.1. The Lennard-Jones parameters for gas compounds considered in this work

adsorbate	ε/k_B (K)	σ (Å)
CO_2	236.1	3.75
CH_4	148	3.73
H_2	34.2	2.96
Ar	119.5	3.41
Kr	165.2	3.66

4.3.1 Gas Mixtures Adsorption at Low to Moderate Pressure

We first calibrate the cDFT predictions for binary mixtures because they provide a good benchmark for the theoretical description of mixture adsorptions. Besides, binary mixtures are important to understand the physics underlying adsorption selectivity. Figure 4.2A shows the adsorption isotherms for an equimolar mixture of Kr and Ar in MOF-5 at 297 K over a range of pressure up to 50 bar. The symbols are from GCMC carried out in this work, and the lines are predicted from IAST and the two versions of cDFT. Because Kr and Ar molecules are similar in terms of both size and interaction energy (viz. the LJ parameters), unsurprisingly, IAST shows near quantitative performance for the adsorption amounts of both species. Previous comparisons also indicate excellent agreement between IAST and GCMC at low to moderate pressure.[20] The good agreement may be attributed not only to the similarity between different species but also to the dominant effects of adsorbate-adsorbent interactions. For cDFT, both the mean-field approximation (cDFT-MFA) and weighted density approximation (cDFT-WDA) predict the adsorption isotherms in fair agreement with GCMC. While cDFT-WDA achieves a numerical performance slightly better than IAST, cDFT-MFA underestimates the adsorption amount for both Kr and Ar due to the neglect of correlation effects. In comparison with IAST, one major advantage of cDFT is that it does not

require single-component adsorption isotherms as the input. Besides, it contains atomistic details that are helpful for the computational design of adsorbent materials.

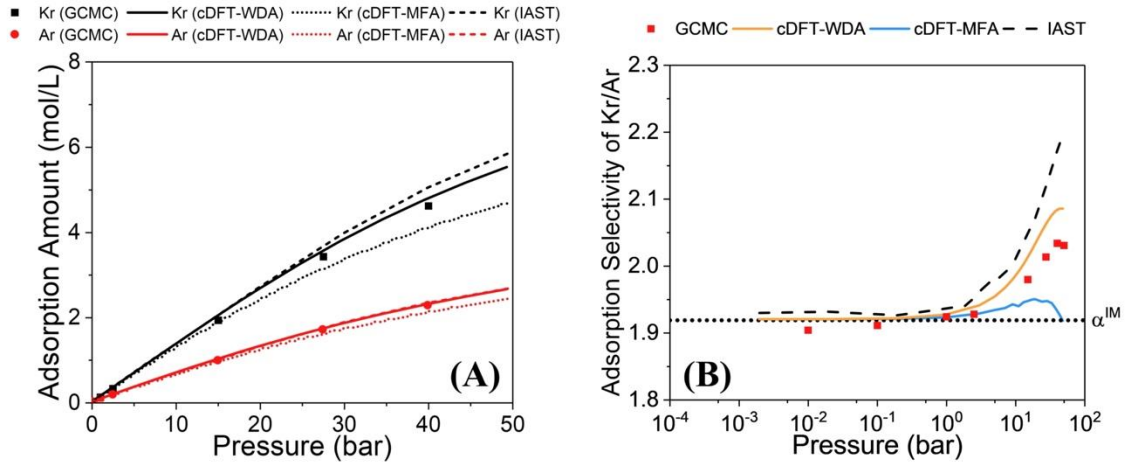

Figure 4.2. Adsorption amounts (A) and selectivity (B) for MOF-5 in contact with an equimolar mixture of Kr and Ar in the bulk at 297 K. In panel B, the dotted line represents the adsorption selectivity at infinite dilution.

Figure 4.2B shows the adsorption selectivity corresponding the adsorption isotherms shown in Figure 4.2A. In the limit of infinite dilution (viz, at extremely low bulk pressure), the selectivity of Kr/Ar predicted by either cDFT or by IAST converges to the ideal adsorption selectivity predicted by Henry's law. cDFT is able to reproduce the exact adsorption selectivity at infinite dilution because the excess Helmholtz energy vanishes in the ideal limit. In principle, IAST is also able to reproduce the ideal limit if highly accurate data are available for pure-component adsorption isotherms at infinite dilution.[26] However, the procedure is numerically problematic and requires much more iterations in GCMC simulations because small adsorption amount at low pressure leads to large errors in the adsorption selectivity. With the increase of pressure, cDFT-WDA predicts the adsorption selectivity still in quantitative agreement with that from GCMC

simulation while IAST shows noticeable deviations. In contrast, cDFT-MFA underestimates the adsorption amount for Kr and predicts unreasonable adsorption selectivity at moderate pressure. For other binary mixtures with more distinct physiochemical properties (e.g., CH_4/CO_2), we observe similar trends in adsorption isotherm and selectivity at low to moderate pressures (shown in Figure 4.3).

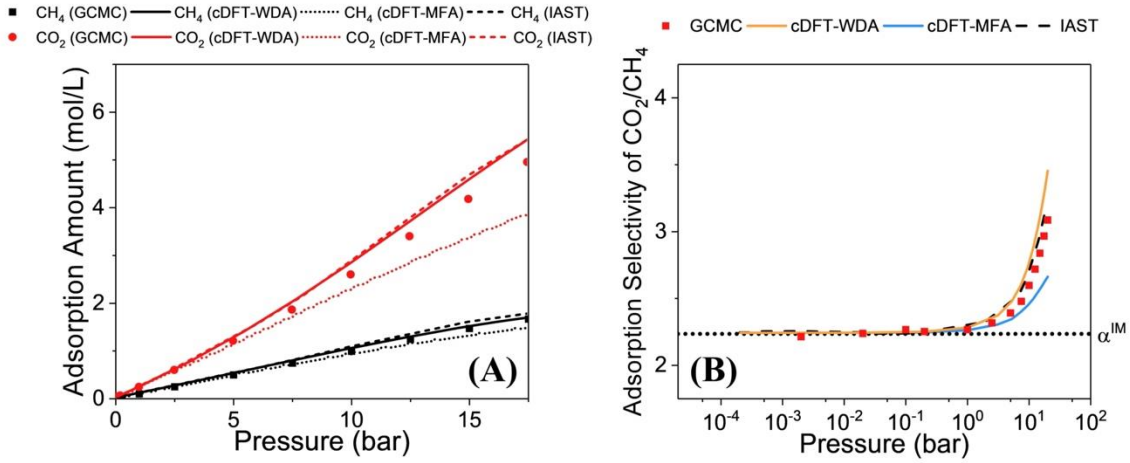

Figure 4.3. Adsorption isotherms (A) and selectivity (B) for an equimolar mixture of CH_4 and CO_2 in MOF-5 at 313.15 K up to 15 bar. The dotted line represents the adsorption selectivity at infinite dilution.

Practical applications of adsorption-based separation processes are often concerned with gas mixtures beyond binary systems (e.g., $H_2/N_2/CO/CH_4/CO_2$ in hydrogen purification). The adsorbate-adsorbate interactions become more complicated when there are more components in the mixture.[38] Therefore, in addition to binary gas mixtures, we also calibrate the cDFT methods with a ternary system, $H_2/CH_4/CO_2$, which has distinct size and interaction energy differences among different species. Figure 4.4 shows the adsorption isotherms at 313.15 K. Similar to that observed for the binary gas mixture, both IAST and cDFT-WDA slightly overestimate the simulation data for the

highly adsorbed component (viz., CO_2) while giving excellent predictions of the adsorption amount for the other two components (H_2/CH_4). However, cDFT-MFA underestimates the adsorption amount for both CO_2 and CH_4. In terms of the adsorption selectivity, IAST and both versions of cDFT are able to reproduce the results from GCMC simulation. It is also worth mentioning that, compared with experimental measurement in literature[15, 58], the adsorption behavior (viz. adsorption amount) of gas mixture in MOF-5 are well captured by the force field parameters adapted in this work (shown in Figure 4.5), which indicates its direct relevance to the practical applications.

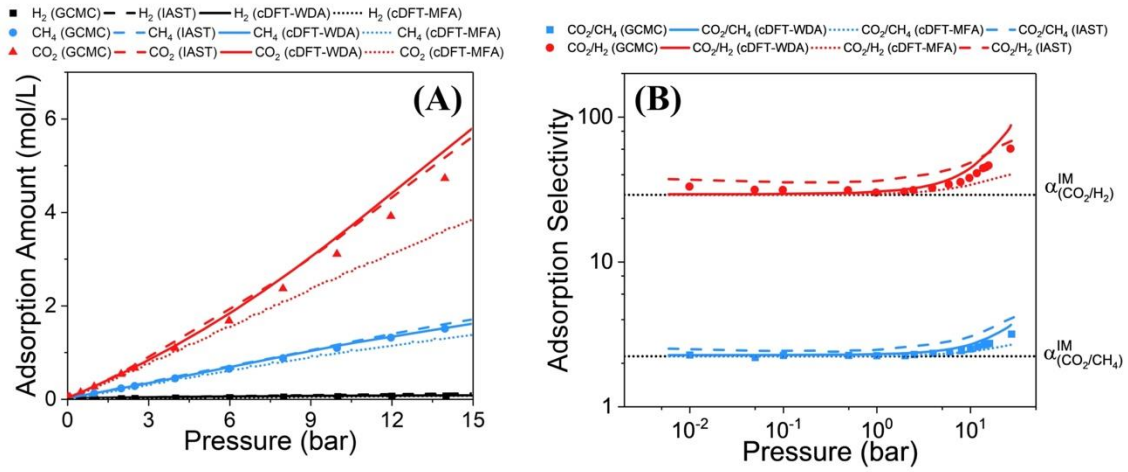

Figure 4.4. Adsorption isotherms (A) and selectivity (B) for a ternary mixture of H_2, CH_4 and CO_2 in MOF-5 at 313.15 K with the molar ratios in the bulk given by H_2:CH_4:CO_2=15:42.5:42.5. In panel B, the dotted line represents the adsorption selectivity at infinite dilution.

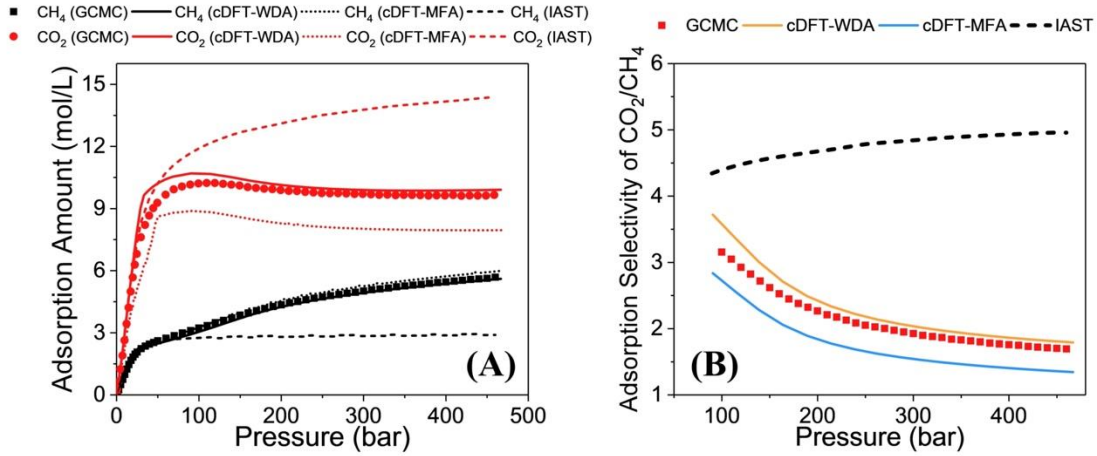

Figure 4.5. Adsorption isotherms for (A) equimolar mixture of CO_2 and CH_4 in MOF-5 at 297 K and ternary mixture of H_2, CH_4 and CO_2 in MOF-5 at 297 K with different bulk concentration: (B) H_2:CH_4:CO_2=15:42.5:42.5 (C) H_2:CH_4:CO_2=42.5:15:42.5. The dashed lines are calculated from the classical density functional theory with mean field approximation (cDFT-MFA), the solid lines are from the classical density functional theory with weighted density approximation (cDFT-WDA) and the symbol are experimental measurement from literature[15, 58].

4.3.2 Gas Mixtures Adsorption at High Pressure

Industrial applications of adsorption-based separation are mostly operated between low and moderate pressures (under 100 bar). Under those conditions, the IAST performance has been proven to be reasonable.[17, 20] However, separation of gas mixtures from power plants, especially when the process is built on the integrated gasification combined cycle (IGCC), is often carried out under higher pressure (up to 200 bar).[16] Owing to their excellent mechanic stability and thermal stability, many MOFs are promising to serve as effective adsorbents for those processes. To facilitate selection and computational design of MOFs, computational methods are needed for fast and accurate evaluation of adsorption isotherms and selectivity at high pressure.

Figure 4.6 shows the adsorption isotherms and selectivity for an equimolar gas mixture of Kr and Ar in MOF-5 at 297 K up to 450 bar of the bulk pressure. Although IAST shows good performance for the bulk pressure up to 50 bar, with the further increase of the bulk pressure, it over- and under-estimates the adsorption amounts of Kr and Ar, respectively. It is somewhat surprising that IAST fails to predict the adsorption isotherm at high pressure even for gas mixtures of similar molecular size and interaction energy. At high pressure, not only does the assumption of ideal solution break down, but the two-dimensional model is also problematic due to the significant inhomogeneity of gas density inside the pores. Compared with the adsorption isotherms calculated from GCMC simulation, both versions of cDFT perform better than IAST at high pressure. While the cDFT predictions of the adsorption amount are near perfect for Ar, noticeable differences are seen between cDFT-WDA and cDFT-MFA predictions for Kr, which slightly over- and under-estimates the adsorption amount, respectively, in comparison with the simulation data. Figure 4.6B shows that cDFT-WDA yields a much better selectivity than cDFT-MFA while IAST yields only semi-quantitative predictions.

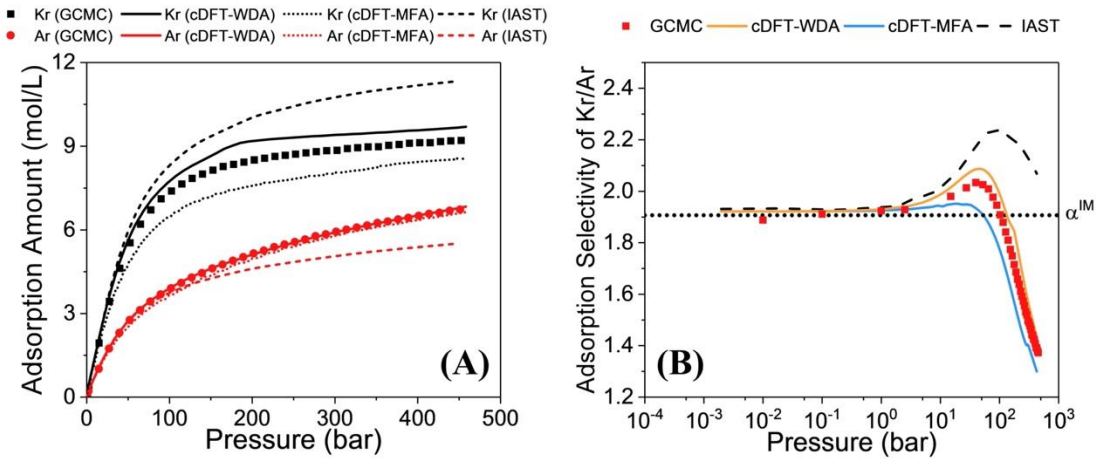

Figure 4.6. Adsorption isotherms (A) and selectivity (B) for an equimolar mixture of Kr and Ar in MOF-5 at 297 K. In Panel B, the dotted line represents the adsorption selectivity at infinite dilution.

For gas mixtures containing molecules with asymmetry in both size and interaction energy, the difference of adsorbate-adsorbate interactions is magnified at high pressure. As a result, the ideal-solution assumption becomes more problematic as the number of components increases. Figure 4.7 compares the adsorption isotherms and selectivity calculated from GCMC with those predicted by cDFT and IAST for a ternary mixture of H_2, CH_4 and CO_2 (with molar ratios in the bulk H_2:CH_4:CO2=15:42.5:42.5) in MOF-5 at 313.15 K. While cDFT-WDA is able to predict both the adsorption amounts and selectivity in quantitative agreement with the simulation results for the entire range of pressure, the predictions by cDFT-MFA are mostly semi-quantitative. Similar to the adsorption of an equimolar binary gas mixture of CH_4 and CO_2 (shown in Figure 4.8), the increase of pressure reduces the adsorption amount for CO_2 beyond a certain value while the adsorptions of CH_4 and H_2 keep on increasing as pressure rises. As CH_4 and H_2 molecules are smaller compared with CO_2, the favorable adsorption sites (near the

corners of the MOF pores) are more likely to be further occupied by CH_4 or H_2 than CO_2 at high gas pressure (shown in Figure 4.9).[59] The maximum CO2 adsorption amount at 115 bar is captured by both versions of cDFT but not by IAST. More specifically, cDFT-WDA quantitatively captures the competitive adsorption behavior while cDFT-MFA underestimates the adsorption amount for both CO_2 and H_2.

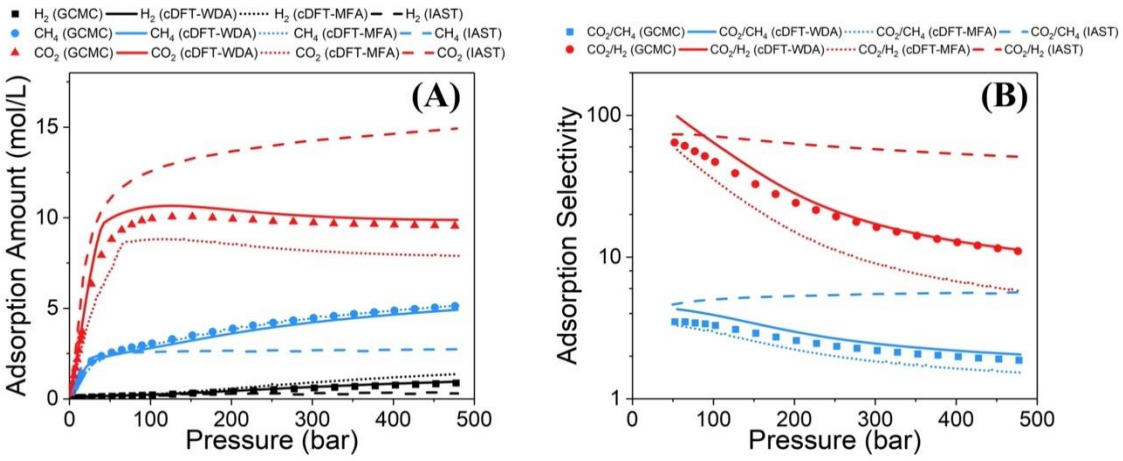

Figure 4.7. Adsorption isotherms (A) and selectivity (B) for a ternary mixture of H_2, CH_4 and CO_2 in MOF-5 at 313.15 K with bulk molar ratios of H_2:CH_4:CO_2=15:42.5:42.5. In Panel B, the dotted line represents the adsorption selectivity at infinite dilution.

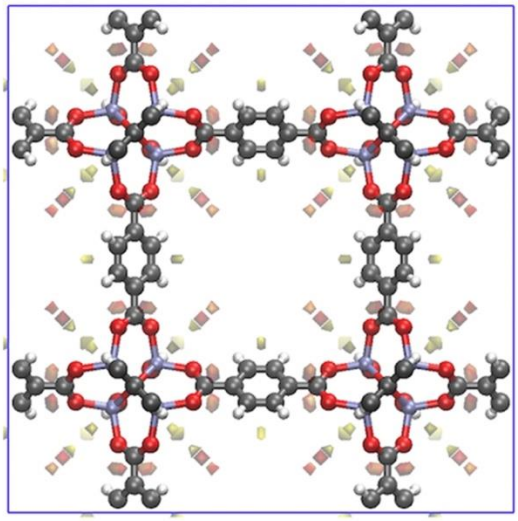

Figure 4.8. Adsorption isotherms (A) and selectivity (B) for an equimolar mixture of CH_4 and CO_2 in MOF-5 at 313.15 K up to 450 bar.

For the ternary mixture considered in this work, CO_2 and CH_4 have the same composition in the bulk phase. Interestingly, the presence of H_2 has different effects on the adsorptions of CH_4 and CO_2. Compared with the equimolar binary gas mixture CH_4/CO_2 (shown in Figure 4.8), H_2 has a stronger effect on CH_4 adsorption than that on CO_2 because these gas molecules have different favorable adsorption sites. The favorable adsorption site for H_2 is closer to that of CH_4 than that of CO_2 (shown in Figure 4.9). As a result, the competition between CH_4 and H_2 adsorptions is more significant than that between CO_2 and H_2.

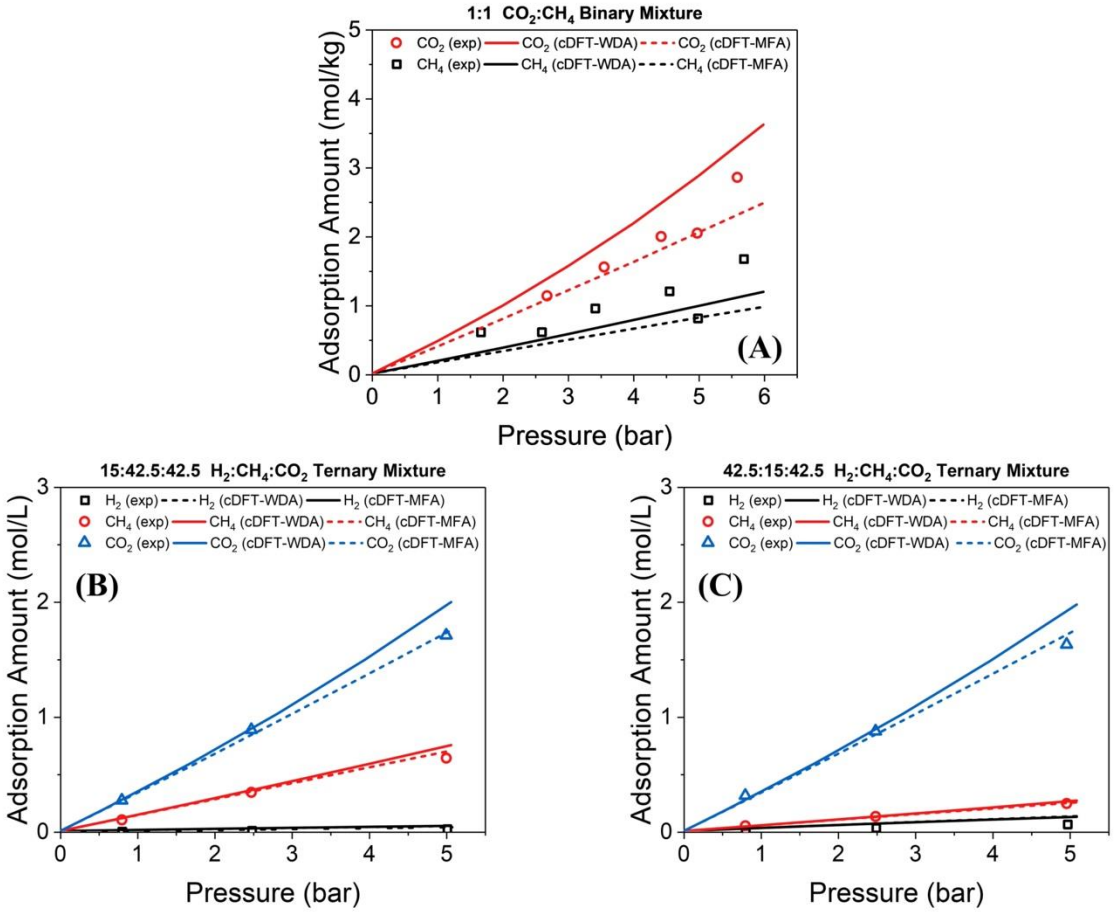

Figure 4.9. Density iso-surfaces of CH_4, CO_2 and H_2 for a ternary mixture with bulk composition (H_2:CH_4:CO_2=15:42.5:42.5) in MOF-5 at 313.15 K and gas pressure 100 bar. The red, orange and yellow colors represent the density iso-surfaces of CH_4, CO_2 and H_2, respectively. The local density of each iso-surface is taken as half of the maximum local density value. The grey, purple, red and white spheres represent carbon, zinc, oxygen and hydrogen atoms, respectively.

Compared with the results from GCMC simulation, both cDFT-WDA and cDFT-MFA make near quantitative predictions of the selectivity at high pressure. As that for the binary system, cDFT-WDA gives a better prediction than cDFT-MFA. In the bulk limit, cDFT-WDA reduces to the MBWR equation, which is a well-tested equation of state, while cDFT-MFA yields only qualitative results as one may expect from the van der

Waals equation. In contrast to cDFT predictions, the results from IAST are not even qualitative for the selectivity of CO_2/CH_4 at high pressure due to the neglect of adsorbate-adsorbate interactions.

4.3.3 Adsorption Sites and Density Isosurfaces

As mentioned above, cDFT gives not only adsorption isotherms and selectivity but also full atomistic details useful for adsorbent design. The adsorption sites for different adsorbates can be identified from the density isosurfaces. For example, Figure 4.10 shows the density profiles of CH_4, CO_2 and H_2 predicted by cDFT-WDA for the ternary gas mixture with molar composition (H_2:CH_4:CO_2=15:42.5:42.5) in MOF-5 at 313.15 K and the bulk pressure of 100 bar and 300 bar. At 100 bar, the gas molecules are mainly localized on the favorable adsorption sites. The local density of CO_2 is much higher than those of H_2 and CH_4, which explains the stronger adsorption of CO_2 than that of H_2 or CH_4. When the bulk pressure increases to 300 bar, the densities of CH_4 and H_2 extend to the pore centers and that of CO_2 declines, indicating that the favorable adsorption sites for CO_2 are more likely to be occupied by smaller gas molecules (i.e., CH_4 and H_2). In other words, the favorable adsorption sites taken by H_2 and CH_4 molecules repel CO_2 molecules, resulting in the reduction of the CO_2 adsorption in the isotherm.

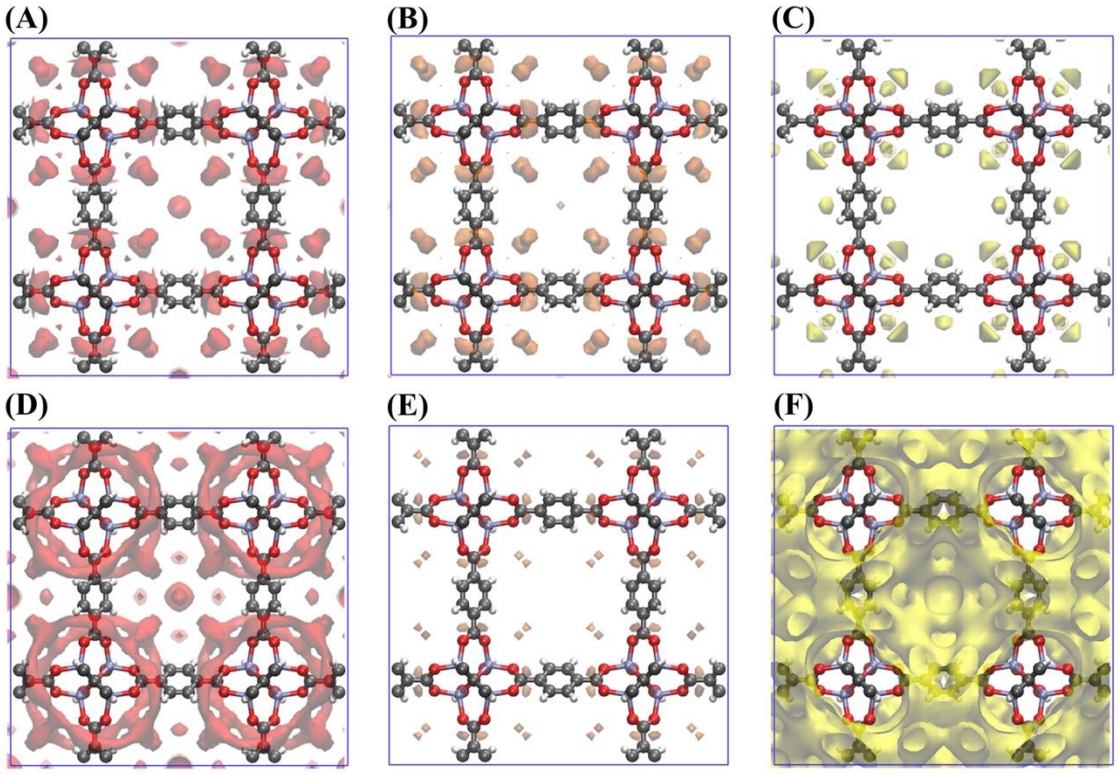

Figure 4.10. Density isosurfaces of CH_4, CO_2 and H_2 in MOF-5 for the adsorption of a ternary mixture with bulk molar composition (H_2:CH_4:CO_2=15:42.5:42.5) predicted by cDFT-WDA at 313.15 K and the gas pressure of 100 bar (top) and 300 bar (bottom). The red isosurfaces in (A) and (D) are for CH_4 with the local density of 0.017 molecules/Å^3. The orange isosurfaces in (B) and (E) are for CO_2 with the local density of 0.09 molecules/Å^3. The yellow isosurfaces in (C) and (F) are for H_2 with the local density of 0.0005 molecules/Å^3. The grey, purple, red and white spheres represent carbon, zinc, oxygen and hydrogen atoms, respectively.

With atomistic information for mixture adsorption available from cDFT, we can identify favorable adsorption sites in the complex structure not only for existing nanoporous materials, but also for in-silico designed nanoporous materials before synthesis. Clearly, the density isosurface is beyond the simple geometry analysis (e.g., pore size distribution) yet offers direct insights when designing new nanoporous materials for specific applications. For example, to remove even more CO_2 from ternary

mixture of H_2, CH_4 and CO_2 at high pressure, one may substitute the metal node and organic linker in MOF-5 with smaller secondary building blocks that allow for more adsorption of CO_2 according to favorable adsorption sites identified from Figure 4.10.

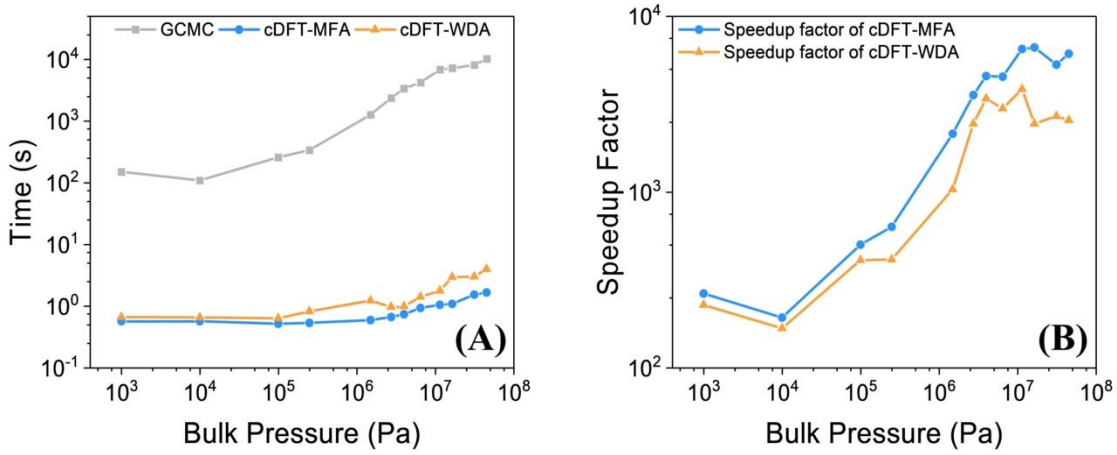

Figure 4.11. (A) Computation time of GCMC, cDFT-MFA and cDFT-WDA versus system pressure for equimolar binary mixture of Kr and Ar. (B) The speedup factors of cDFT-MFA and cDFT-WDA are based on the computation time of GCMC.

Finally, it is worth emphasizing that the main advantage of cDFT over GCMC is the computational efficiency. For example, Figure 4.11 shows the computation time of GCMC, cDFT-MFA and cDFT-WDA for equimolar binary mixture of Kr and Ar. Also shown in Figure 4.11 are the speedup factors of cDFT-MFA and cDFT-WDA compared with GCMC simulation. The computational time and speedup factor for ternary mixture of H_2, CH_4 and CO_2 are shown in Figure 4.12. All the cDFT calculations are executed with massively paralleled GPU-accelerated implementation on Nvidia Tesla P100 graphic card with one CPU core on Intel Xeon E5-2620 v4 while all the GCMC simulations are carried via RASPA with one CPU core on Intel Xeon E5-2640 v4. With massively paralleled GPU-accelerated implementation, both cDFT-MFA and cDFT-

WDA are 2~3 orders of magnitude faster than serial GCMC simulation for typical binary and ternary mixtures at moderate pressure, and the speedup becomes even more noticeable at higher pressure. The detailed discussion of this implementation can be found in our previous work.[51] Even compared with parallel GCMC code with GPU acceleration such as GPU Optimized Monte Carlo (GOMC)[60], massively paralleled GPU-accelerated cDFT is still much faster. For GOMC, the speedup factor is up to 30 while massively paralleled GPU-accelerated cDFT can achieve speedup factor up to several hundreds. Compared with cDFT-MFA, cDFT-WDA takes correlation effect into account and the extra correlation term needs to be re-evaluated in each iteration, which results in a slight increase of the computation time. Although the computational cost of cDFT increases proportional to the number of components, all the calculations can be finished within few seconds even for ternary mixture. The rapid and accurate evaluation of multicomponent adsorption isotherms is essential for the construction of large database for data-driven materials design.

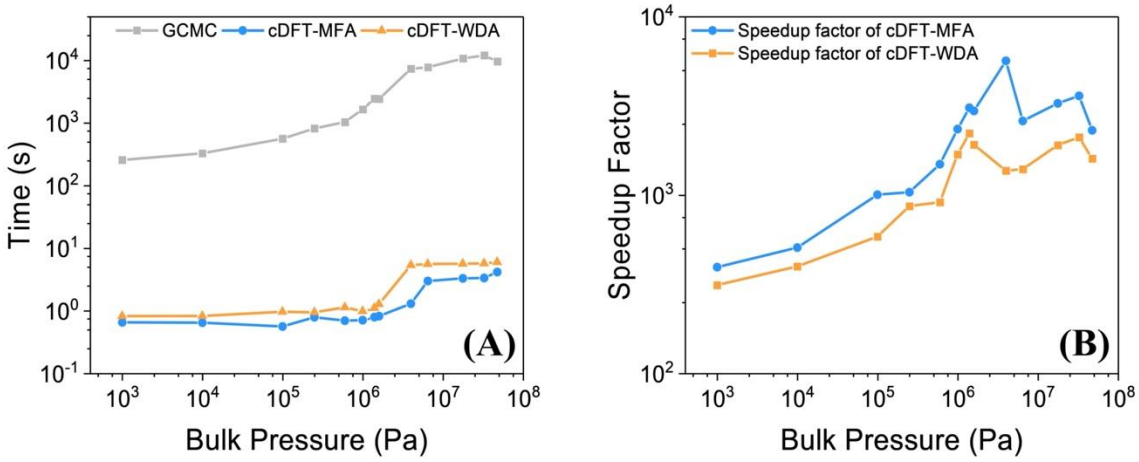

Figure 4.12. (A) Computation time of GCMC, cDFT-MFA and cDFT-WDA versus system pressure for ternary mixture of H_2, CH_4 and CO_2. (B) The speedup factors of cDFT-MFA and cDFT-WDA are based on the computation time of GCMC.

4.4 Conclusions

In this work, we propose two versions of classical density functional theory (cDFT) for describing adsorption of multicomponent gas mixtures by nanoporous materials. Their main difference lies in the formulation of the excess Helmholtz energy due to van der Waals attraction. One is based on the mean-field approximation (cDFT-MFA), which is commonly used in cDFT calculations and has been adopted in porous materials characterization. The other accounts for the correlation effects by using the weighted density approximation (cDFT-WDA). The two formulations of the excess Helmholtz energy have been tested for both binary and ternary systems. Compared with the results from grand canonical Monte Carlo (GCMC) simulations, cDFT-WDA is able to predict both the adsorption isotherms and selectivity near quantitatively. However, cDFT-MFA significantly underestimates the adsorption amount due to the absence of the correlation contribution to the excess Helmholtz energy. Compared with the ideal

adsorbed solution theory (IAST), both versions of cDFT show substantial improvements, especially at high pressure where adsorbate-adsorbate interactions and correlation effects become more significant. At low pressure, both two versions of cDFT and IAST can quantitatively predict the adsorption isotherm and selectivity of gas mixture in MOF-5. Moreover, cDFT offers atomistic details revealing the underlying mechanism of competitive adsorption in gas mixtures, which well explains the peak adsorption in CO2. The microscopic insights are helpful to design nanoporous materials for more efficient separation of multicomponent gas mixtures by adsorption. In addition, with the massively parallel GPU-accelerated implementation, both cDFT calculations can be accomplished at the scale of few seconds for each thermodynamic condition, which is a few orders of magnitude faster than GCMC simulation. Therefore, cDFT may be used as an alternative to IAST or GCMC for constructing a large and accurate adsorption database for multicomponent gas mixtures that are required for the data-driven inverse design of nanoporous materials.

Bibliography

1. Aaron, D. and C. Tsouris, *Separation of CO2 from flue gas: A review.* Sep Sci Technol, 2005. **40**(1-3): p. 321-348.

2. Ritter, J.A. and A.D. Ebner, *State-of-the-art adsorption and membrane separation processes for hydrogen production in the chemical and petrochemical industries.* Sep Sci Technol, 2007. **42**(6): p. 1123-1193.

3. Sircar, S. and T.C. Golden, *Purification of hydrogen by pressure swing adsorption.* Sep Sci Technol, 2000. **35**(5): p. 667-687.

4. Saha, D., et al., *Postextraction Separation, On-Board Storage, and Catalytic Conversion of Methane in Natural Gas: A Review.* Chem Rev, 2016. **116**(19): p. 11436-11499.

5. Barelli, L., et al., *Hydrogen production through sorption-enhanced steam methane reforming and membrane technology: A review.* Energy, 2008. **33**(4): p. 554-570.

6. Sholl, D.S. and R.P. Lively, *Seven chemical separations to change the world.* Nature, 2016. **532**(7600): p. 435-7.

7. Li, J.R., R.J. Kuppler, and H.C. Zhou, *Selective gas adsorption and separation in metal-organic frameworks.* Chem Soc Rev, 2009. **38**(5): p. 1477-504.

8. Feng, X., X. Ding, and D. Jiang, *Covalent organic frameworks.* Chem Soc Rev, 2012. **41**(18): p. 6010-6022.

9. Ozekmekci, M., G. Salkic, and M.F. Fellah, *Use of zeolites for the removal of H2S: A mini-review.* Fuel Process Technol, 2015. **139**: p. 49-60.

10. Chong, S., et al., *Applications of machine learning in metal-organic frameworks.* Coordin Chem Rev, 2020. **423**: p. 213487.

11. Yao, Z.P., et al., *Inverse design of nanoporous crystalline reticular materials with deep generative models.* Nature Machine Intelligence, 2021. **3**(1): p. 76-86.

12. Chung, Y.G., et al., *In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm.* Sci Adv, 2016. **2**(10): p. e1600909.

13. Simon, C.M., et al., *What are the best materials to separate a xenon/krypton mixture?* Chem Mater, 2015. **27**(12): p. 4459-4475.

14. Simon, C.M., et al., *The materials genome in action: identifying the performance limits for methane storage.* Energy Environ Sci, 2015. **8**(4): p. 1190-1199.

15. Kloutse, F., et al., *Experimental benchmark data of CH4, CO2 and N2 binary and ternary mixtures adsorption on MOF-5.* Sep Purif Technol, 2018. **197**: p. 228-236.

16. Ullah, R., et al., *Adsorption equilibrium studies of CO2, CH4 and N2 on various modified zeolites at high pressures up to 200 bars.* Microporous Mesoporous Mater, 2018. **262**: p. 49-58.

17. Walton, K.S. and D.S. Sholl, *Predicting multicomponent adsorption: 50 years of the ideal adsorbed solution theory.* AIChE J, 2015. **61**(9): p. 2757-2762.

18. Myers, A.L. and J.M. Prausnitz, *Thermodynamics of mixed-gas adsorption.* AIChE J, 1965. **11**(1): p. 121-127.

19. Simon, C.M., B. Smit, and M. Haranczyk, *pyIAST: Ideal adsorbed solution theory (IAST) Python package.* Comput Phys Commun, 2016. **200**: p. 364-380.

20. Cessford, N.F., N.A. Seaton, and T. Duren, *Evaluation of Ideal Adsorbed Solution Theory as a Tool for the Design of Metal-Organic Framework Materials.* Ind Eng Chem Res, 2012. **51**(13): p. 4911-4921.

21. Wang, K., S.Z. Qiao, and X.J. Hu, *On the performance of HIAST and IAST in the prediction of multicomponent adsorption equilibria.* Sep Purif Technol, 2000. **20**(2-3): p. 243-249.

22. Sakuth, M., J. Meyer, and J. Gmehling, *Measurement and prediction of binary adsorption equilibria of vapors on dealuminated Y-zeolites (DAY).* Chem Eng Process, 1998. **37**(4): p. 267-277.

23. Costa, E., et al., *Adsorption of binary and ternary hydrocarbon gas mixtures on activated carbon: experimental determination and theoretical prediction of the ternary equilibrium data.* AIChE J, 1981. **27**(1): p. 5-12.

24. Ryan, P., et al., *Computational Screening of Metal-Organic Frameworks for Xenon/Krypton Separation.* AIChE J, 2011. **57**(7): p. 1759-1766.

25. Gurdal, Y. and S. Keskin, *A new approach for predicting gas separation performances of MOF membranes.* J Membr Sci, 2016. **519**: p. 45-54.

26. Van Heest, T., et al., *Identification of Metal–Organic Framework Materials for Adsorption Separation of Rare Gases: Applicability of Ideal Adsorbed Solution Theory (IAST) and Effects of Inaccessible Framework Regions.* J Phys Chem C, 2012. **116**(24): p. 13183-13195.

27. Fu, J., Y. Tian, and J.Z. Wu, *Classical density functional theory for methane adsorption in metal-organic framework materials.* AIChE J, 2015. **61**(9): p. 3012-3021.

28. Bristow, J.K., D. Tiana, and A. Walsh, *Transferable Force Field for Metal-Organic Frameworks from First-Principles: BTW-FF*. J Chem Theory Comput, 2014. **10**(10): p. 4644-4652.

29. Bureekaew, S., et al., *MOF-FF - A flexible first-principles derived force field for metal-organic frameworks.* Phys Status Solidi B, 2013. **250**(6): p. 1128-1141.

30. Witman, M., B. Wright, and B. Smit, *Simulating Enhanced Methane Deliverable Capacity of Guest Responsive Pores in Intrinsically Flexible MOFs.* J Phys Chem Lett, 2019. **10**(19): p. 5929-5934.

31. Bousquet, D., F.X. Coudert, and A. Boutin, *Free energy landscapes for the thermodynamic understanding of adsorption-induced deformations and structural transitions in porous materials.* J Chem Phys, 2012. **137**(4): p. 044118.

32. Yu, Y.X. and J.Z. Wu, *Structures of hard-sphere fluids from a modified fundamental-measure theory.* J Chem Phys, 2002. **117**(22): p. 10156-10164.

33. Yu, Y.X., *A novel weighted density functional theory for adsorption, fluid-solid interfacial tension, and disjoining properties of simple liquid films on planar solid surfaces.* J Chem Phys, 2009. **131**(2): p. 024704.

34. Liu, Y., et al., *Development of a density functional theory in three-dimensional nanoconfined space: H2 storage in metal-organic frameworks.* J Phys Chem B, 2009. **113**(36): p. 12326-31.

35. Wu, J., *Density functional theory for liquid structure and thermodynamics*, in *Molecular thermodynamics of complex systems*, X. Lu, Y. Hu, and H. Chen, Editors. 2009, Springer: Berlin. p. 1-74.

36. Motevaselian, M. and N. Aluru, *An EQT-based cDFT approach for thermodynamic properties of confined fluid mixtures.* J Chem Phys, 2017. **146**(15): p. 154102.

37. Camacho Vergara, E.L., G.M. Kontogeorgis, and X. Liang, *Gas adsorption and interfacial tension with classical density functional theory.* Ind Eng Chem Res, 2019. **58**(14): p. 5650-5664.

38. Krishna, R., *Metrics for Evaluation and Screening of Metal-Organic Frameworks for Applications in Mixture Separations.* ACS Omega, 2020. **5**(28): p. 16987-17004.

39. Yu, Y.X. and J.Z. Wu, *A modified fundamental measure theory for spherical particles in microchannels.* J Chem Phys, 2003. **119**(4): p. 2288-2295.

40. Wu, J., *Variational Methods in Molecular Modeling*. 2016: Springer.

41. Wu, J. and Z. Li, *Density-functional theory for complex fluids.* Annu Rev Phys Chem, 2007. **58**: p. 85-112.

42. Wu, J., *Density functional theory for chemical engineering: From capillarity to soft materials.* AIChE J, 2006. **52**(3): p. 1169-1193.

43. Rappe, A.K., et al., *Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations.* J Am Chem Soc, 1992. **114**(25): p. 10024-10035.

44. Johnson, J.K., J.A. Zollweg, and K.E. Gubbins, *The Lennard-Jones Equation of State Revisited.* Mol Phys, 1993. **78**(3): p. 591-618.

45. Roth, R., et al., *Fundamental measure theory for hard-sphere mixtures revisited: the White Bear version.* J Phys: Condens Matter, 2002. **14**(46): p. 12063.

46. Henderson, D., *Monte Carlo and perturbation theory studies of the equation of state of the two-dimensional Lennard-Jones fluid.* Mol Phys, 1977. **34**(2): p. 301-315.

47. Cotterman, R.L., B.J. Schwarz, and J.M. Prausnitz, *Molecular thermodynamics for fluids at low and high densities. Part I: Pure fluids containing small or large molecules.* AIChE J, 1986. **32**(11): p. 1787-1798.

48. Fu, J., et al., *Density Functional Methods for Fast Screening of Metal Organic Frameworks for Hydrogen Storage.* J Phys Chem C, 2015. **119**(10): p. 5374-5385.

49. Carnahan, N.F. and K.E. Starling, *Equation of state for nonattracting rigid spheres.* J Chem Phys, 1969. **51**(2): p. 635-636.

50. Hager, W.W. and H.C. Zhang, *Algorithm 851: CG DESCENT, a conjugate gradient method with guaranteed descent.* ACM Transactions on Mathematical Software, 2006. **32**(1): p. 113-137.

51. Zhou, M. and J. Wu, *A GPU implementation of classical density functional theory for rapid prediction of gas adsorption in nanoporous materials.* J Chem Phys, 2020. **153**(7): p. 074101.

52. Steele, W.A., *The physical interaction of gases with crystalline solids: I. Gas-solid energies and properties of isolated adsorbed atoms.* Surf Sci, 1973. **36**(1): p. 317-352.

53. Martin, M.G. and J.I. Siepmann, *Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes.* J Phys Chem B, 1998. **102**(14): p. 2569-2577.

54. Buch, V., *Path integral simulations of mixedpara ‐ D2andortho ‐ D2clusters: The orientational effects.* J Chem Phys, 1994. **100**(10): p. 7610-7629.

55. Lastoskie, C., K.E. Gubbins, and N. Quirke, *Pore size heterogeneity and the carbon slit pore: a density functional theory model.* Langmuir, 1993. **9**(10): p. 2693-2702.

56. Pellenq, R.J. and D. Nicholson, *Intermolecular potential function for the physical adsorption of rare gases in silicalite.* J Phys Chem, 1994. **98**(50): p. 13339-13349.

57. Babarao, R., S. Dai, and D.E. Jiang, *Effect of Pore Topology and Accessibility on Gas Adsorption Capacity in Zeolitic-Imidazolate Frameworks: Bringing Molecular Simulation Close to Experiment.* J Phys Chem C, 2011. **115**(16): p. 8126-8135.

58. Kloutse, F., et al., *Hydrogen separation by adsorption: Experiments and modelling of H2-N2-CO2 and H2-CH4-CO2 mixtures adsorption on CuBTC and MOF-5.* Microporous Mesoporous Mater, 2018. **271**: p. 175-185.

59. Zhou, W., et al., *Molecular insights into competitive adsorption of CO 2/CH 4 mixture in shale nanopores.* RSC Adv, 2018. **8**(59): p. 33939-33946.

60. Nejahi, Y., et al., *GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids.* SoftwareX, 2019. **9**: p. 20-27.

Chapter 5. Massively Parallel GPU-Accelerated String Method for Fast and Accurate Prediction of Molecular Diffusivity in Nanoporous Materials

The diffusivity of guest molecules in nanoporous materials is instrumental for practical applications ranging from gas separation to catalysis and energy storage. Conventional methods to predict diffusion coefficients are computationally demanding, in particular for polyatomic molecules with small diffusivity in nanoporous materials. In this chapter, we have implemented a massively parallel graphic processing unit (GPU)-accelerated string method to calculate the minimum energy path for the diffusion of polyatomic molecules in nanoporous materials. The GPU parallelization enables fast prediction of molecular diffusivity in nanoporous materials, speeding up the computation by a factor of over 500 in comparison with serial CPU calculations. The massively parallel GPU-accelerated string method yields diffusion coefficients in excellent agreement with results from molecular dynamics while reducing the computational cost by several orders of magnitude. It will thus open up opportunities for high-throughput screening and inverse design of nanoporous materials.

5.1 Introduction

Recent years have seen the rapid development of nanoporous materials with a vast variety of building blocks.[1-6] Nanoporous materials (e.g., metal-organic frameworks) can now be designed and synthesized by assembling organic ligands and metal cluster with appropriate topology. As a result, large materials databases become commonplace promising data-driven applications via high-throughput screening and computational design.[7-9] Transport properties such as diffusion coefficient are closely related to many

important applications. For example, diffusivity dictates the performance of nanoporous materials, including zeolites, metal-organic and covalent organic frameworks, for gas separation and ion sieving.[10-13] Efficient computational methods for fast yet accurate prediction of transport properties are always in great demand for searching the best nanoporous materials in a structural database and/or for the inverse design.[10, 14]

For guest molecules in a nanoporous material, the diffusion coefficients can be measured using experimental techniques such as quasi-elastic neutron scattering (QENS) and pulsed-field gradients-nuclear magnetic resonance (PFG-NMR).[15, 16] Such experiments are laborious and not suitable for high-throughput operations due to the time-consuming nature of sample preparation and measurement.[17, 18] As a result, experimental data are rarely available for molecular diffusion coefficients of chemical species in large libraries of nanoporous materials. Alternatively, diffusion coefficients can be predicted from a number of theoretical methods.[14, 19-21] Among them, molecular dynamics (MD) simulation has been most widely used to investigate the diffusion of gas molecules in nanoporous materials. Despite its popularity, construction of a diffusion-coefficient database by 'brute force' MD simulation is computationally prohibitive. The task is challenging in particular when one is concerned with the separation of organic molecules (e.g., paraxylene/orthoxylene and benzene/cyclohexane) using nanoporous materials due to the slow diffusivity (less than 10^{-12} m^2/s).

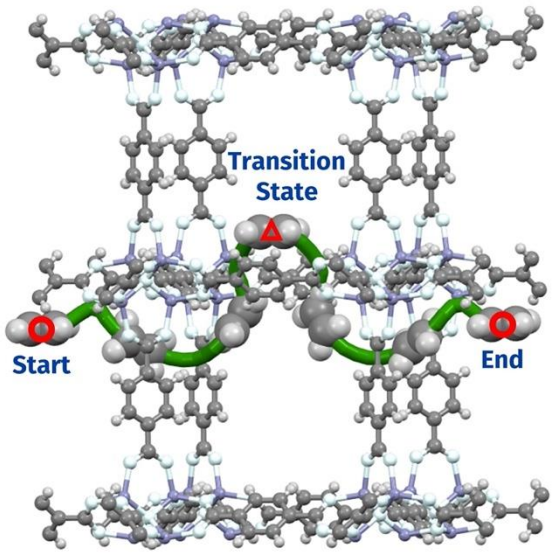

Figure 5.1. Minimum energy path for the center of mass (green line) of ethene in MOF-5. Molecular configuration of ethene at different position along the minimum energy path is also shown. The size of atoms is rescaled and for the purpose of illustration only. Grey, white, ice blue and purple represent carbon, hydrogen, oxygen and zinc, respectively.

Many theoretical attempts have been made to circumvent the computational limit of MD simulation in predicting diffusion coefficients.[10, 11] A well-established alternative is by using the transition-state theory (TST).[21] While diffusivity is typically calculated from the Einstein equation via mean-square displacement (MSD) over long equilibrium steps in MD simulation, TST predicts diffusion coefficients based on a minimum energy path (MEP) that is solely determined by the energy landscape of guest-host interactions (shown in Figure 5.1). Mathematical tools such as nudged elastic band (NEB) and string methods have been commonly used to calculate the MEP.[22-26] While NEB is mostly used in quantum-mechanical calculations of transport properties such as ion diffusivity, the string method is more suitable to obtain the highly curved MEP dictating gas diffusion in nanoporous materials.[27-31] More specifically, the string

method is able to identify the diffusion pathways based on the energy gradients such that each path follows an exactly minimum energy route. Besides NEB and string methods, other mathematical tools, such as tunnel and transition-state search, cluster analysis and grid searching, are also promising.[11, 32] Computationally, TST is able to predict diffusion coefficients much more efficient than molecular simulation because it entails no thermal fluctuations or atomic motions. Regrettably, existing applications of TST methods are mostly limited to the diffusion of simple gas molecules as represented by the single-site Lennard-Jones (LJ) potential. Not only is the extension of the MEP calculation to polyatomic molecules mathematically challenging, but the computational efficiency is severely compromised due to the rapid increase of dimensionality in representing MEP for polyatomic molecules.

In a previous work,[33] we demonstrated that, given a fine-enough three-dimensional potential grid, the string method can be used to accurately assess the minimum energy path (MEP) for the diffusion of simple gas molecules in nanoporous materials. However, the same procedure is not directly applicable to polyatomic molecules because the memory of a typical desktop computer is infeasible to handle the external potential using a multi-dimensional grid with a sufficiently fine resolution essential in MEP calculations. If the external potential is calculated on-the-fly as the string evolves, it would be an enormous computational burden for serial implementation with conventional central processing unit (CPU). Different from CPU, a graphic processing unit (GPU) has many more arithmetic logic units (ALUs, a.k.a. threads) thereby it is capable of high-throughput data processing. Inspired by recent progress of

massively parallel GPU-acceleration of simulation methods with excellent performance,[34-36] we have implemented in this work a massively parallel GPU-accelerated string method for predicting the diffusivity of polyatomic molecule in a large library of nanoporous materials. The algorithm speeds up the theoretical predictions of diffusion coefficients with the string method by a factor of ~500 in comparison with serial CPU implementation. Importantly, the theoretical results are in excellent agreement with MD simulation data for a number of materials. We also benchmark the computational efficiency for high-throughput screening of metal-organic frameworks (MOFs) for ethane/ethylene separation. More than 90% of calculations of the diffusion coefficient in the nanoporous materials can be completed within 30 seconds. By analyzing over 3080 structures from the CoRE MOF 2019 library, we are able to identify promising materials and desirable structural features leading to the highest membrane selectivity. We expect that the expanded computational capability will likely open up avenues for the construction of a large computational database for molecular diffusivity thus empowering data-driven approaches to the inverse design of nanoporous materials.

5.2 Methods and Models

5.2.1 Transition-State Theory

According to the transition-state theory[21], the self-diffusion coefficient for a guest molecule inside a nanoporous material can be calculated from

$$D_0 = \frac{1}{2} k a^2 \tag{5.1}$$

where D_0 is the self-diffusion coefficient, k is the hopping rate (i.e., transmission rate), and a is the hopping distance between two neighboring cages. At infinite dilution, the

hopping rate can be obtained from the minimum energy path (MEP) for the molecular diffusion following the Bennett-Chandler approach

$$k = \sqrt{\frac{k_B T}{2\pi m}} \frac{\exp\left[-\beta V^{ext}\left(s^*\right)\right]}{\int_0^1 \exp\left[-\beta V^{ext}\left(s\right)\right] ds} \tag{5.2}$$

where k_B is the Boltzmann constant, T stands for the absolute temperature, $\beta = 1/\left(k_B T\right)$, m represents the molecular mass, V^{ext} is the potential energy due to the interaction of the guest molecule with the porous material. In the transition state theory, the minimum energy path is described in terms of a dimensionless variable s, which represents the normalized reaction coordinate for the molecular transition between neighboring cages. In general, s depends on the molecular configuration and the center of mass of the guest molecule.

In this work, we assume that both the nanoporous material and the guest molecule have fixed structures. As a result, the reaction coordinate can be uniquely defined by six collective variables $s\left(\theta_1,...,\theta_6\right)$. The first three variables are related to the molecular position and the other three variables represent the Euler angles of the guest molecule, i.e., $s\left(\theta_1,...,\theta_6\right) = s\left(\mathbf{r},\omega\right)$, where $\mathbf{r} = \left(x,y,z\right)$ represents the position for the molecular center of mass (COM), and $\omega = \left(\alpha,\beta,\gamma\right)$ describes how a polyatomic molecule is oriented relative to its original input structure (as shown in Figure 5.2).

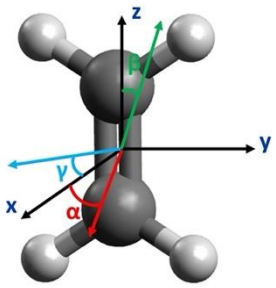

Figure 5.2. Definition of orientation variables (viz., Euler angles) for an ethene molecule relative to the material frame.

To implement the string method numerically, we describe the minimum energy path by using a series of discrete points referred to as images. At each point/image, the dimensionless variables s can be expressed in terms of $\mathbf{r}$ and ω

$$s_i = \sqrt{\left(\frac{l_{\mathbf{r},i}}{L_{\mathbf{r}}}\right)^2 + \left(\frac{l_{\omega,i}}{L_{\omega}}\right)^2} \tag{5.3}$$

where $l_{\mathbf{r},i} = \sum_{j=2}^{i} \left|\mathbf{r}_j - \mathbf{r}_{j-1}\right|$ and $l_{\omega,i} = \sum_{j=2}^{i} \left|\omega_j^* - \omega_{j-1}^*\right|$ are the string arc lengths for the spatial and rotational variables at image i, while $L_{\mathbf{r}}$ and L_{ω} are the arc lengths of the entire string for $\mathbf{r}$ and ω, respectively.

In this work, we use atomistic models for both nanoporous materials and guest molecules. As in a standard molecular force field, the non-bonded interactions are described by the Lennard-Jones (LJ) potential. In addition to the short-range repulsion and van der Waals (vdW) attraction, the electrostatic interactions due to atomic partial charges are accounted for with the Coulomb potential. For atoms in the framework materials, the universal force field (UFF) is adopted for the LJ parameters, while the charge equilibration method from the RASPA software package is used to assign the

point charges of individual atoms.[37-39] The unit cell of each framework material is duplicated along the axis so that the edge length is more than two times the cutoff distance. Whereas the diffusion path may vary with the loadings due to gas-gas interactions, we expect that the effect is relatively insignificant because the free-energy landscape is dominated by the external energy.

For direct comparison with simulation data, we use the force field parameters for polyatomic molecules and nanoporous materials the same as those used in molecular dynamics simulations. For most polyatomic molecules considered in this work, the simulation results were based on the united atom force field (TraPPE-UA) without electrostatic interactions.[40, 41] For nitrogen and carbon dioxide, each atom is modeled as a single Lennard-Jones (LJ) site with the point charge the same as that used in the TraPPE force field.[42] To balance the charge neutrality, a positive charge is placed in the center of nitrogen molecule. The DFT-derived force field is used for benzene.[43] Table 5.1 shows the force-field parameters for all molecules considered in this work. Structural properties of nanoporous materials, such as pore diameters and void fraction, are calculated with Zeo++.[44] Images and videos of atomistic molecular structures presented in this work are rendered from visual molecular dynamics (VMD) and Mercury.[45, 46]

Table 5.1. LJ parameters and partial charge for polyatomic molecules considered in this work

	σ (Å)	ε/k_B (K)	charge (*e*)
CH_4 (methane)	3.730	148	
CH_3- (ethane)	3.760	108	
CH_2- (ethene)	3.680	92.8	
C- (xylene)	3.850	20.0	
CH- (xylene)	3.695	50.5	
CH_3- (xylene)	3.750	98.0	
C- (benzene)	3.470	47.81	-0.15
H- (benzene)	2.850	7.55	0.15
C- (carbon dioxide)	2.80	27	0.4
O- (carbon dioxide)	3.05	79	-0.2
N- (nitrogen)	3.31	36	-0.482
COM (nitrogen)	0	0	0.964

Given the position and configuration of a guest molecule, the external potential accounts for its interaction with the nanoporous material and is given by

$$V^{ext} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_f} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \tag{5.4}$$

where ε and σ stand for the LJ energy and size parameters, respectively, ε_0 stands for the vacuum permeability, N_g and N_f are the number of atoms in each guest molecule and that from the nanoporous material. In calculation of V^{ext}, we use the Ewald summation method for electrostatic interactions, and the Lorentz-Berthelot mixing rule is used for the energy and size parameters between different atoms. The periodic boundary conditions are applied to all directions with the vdW interactions truncated and shifted to zero at 12.9 Å.

5.2.2 Simplified String Method

Within the framework of the transition-state theory (TST), both the computational cost and accuracy are critically dependent on the construction of the minimum energy path (MEP). In our previous work,[33] we demonstrated that the string method provides an efficient way to identify MEPs, leading to an accurate prediction of self-diffusion coefficients for simple gas molecules in nanoporous materials. In principle, the string method is equally applicable to more complicated polyatomic molecules with the minimum energy path obtained by evolving discrete points (a.k.a. images) along "a string" towards the direction of decreasing the external potential. As the number of atoms in the guest molecule increases, the minimum energy path becomes much more difficult to calculate due to the drastic increase of pairwise interactions and the images in the reaction coordinate. In this work, we employ a simplified yet more accurate version of the string method to obtain the minimum energy path.[23] Compared with the original string method, the simplified string method is numerically more stable and accurate, yet it is computationally more efficient.[23]

According to the simplified string method,[23] the evolution of the normalized reaction coordinate is driven by the full gradient of the external potential

$$\frac{ds_i}{dt} = -\nabla V^{ext}\left(s_i\right) = -\left[\frac{\partial V^{ext}\left(s_i\right)}{\partial x} + \frac{\partial V^{ext}\left(s_i\right)}{\partial y} + \frac{\partial V^{ext}\left(s_i\right)}{\partial z} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \alpha} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \beta} + \frac{\partial V^{ext}\left(s_i\right)}{\partial \gamma}\right] \tag{5.5}$$

where t is a fictitious time used in the iteration to search for the minimum energy path,

and s_i represents image *i* on the string. During each iteration, the guest molecule is first updated according to

$$\theta_{i,k}^{\text{II}}(t) = \theta_{i,k}(t) - \Delta t \frac{\partial V^{ext}(s_i)}{\partial \theta_k}|_t, \qquad k = 1,...,6 \tag{5.6}$$

where θ_i represents a molecular coordinate (position or angle) corresponding to image *i*, the partial derivative is evaluated at fictious time *t*, and the superscript $^{\text{II}}$ represents the updated string. Throughout this work, the forward Euler method is used to calculate the derivative of the external potential with respect to θ_i, and Δt is set as 1×10^{-4}.

To identify the diffusion path, we first calculate the energy landscape for the guest molecule at the entrance plane via a discrete grid. The position and orientation of the guest molecule that minimize the external potential are used as the starting image of the string. Due to the periodic boundary conditions, the guest molecule has the configuration at the starting and ending images. Their difference lies only in the reaction coordinate, i.e., parameter *s* along the direction of the minimum energy path. The initial string is generated by positioning the images evenly between the starting and ending points. After that, each iteration updates the position and orientation of the guest molecule according to the normalized reaction coordinate *s* (eq [5.6]). After string evolution in each step, interpolation/reparameterization (eq [5.7]) is needed to retain the continuous shape of string through the nanoporous materials.

When a guest molecule diffuses through a nanoporous material, the preferred molecular orientation depends on the position at the molecular center of mass (COM). As

a result, different images have orientations independent from each other. We only need to interpolate the molecular COM position of the evolved images according to the arc length

$$\mathbf{r}_i(t+\Delta t)=\mathbf{r}_{k-1}^{\|}(t)+\left[\frac{i-1}{N-1}L_{\mathbf{r}}^{\|}(t)-l_{\mathbf{r},k-1}^{\|}(t)\right]\frac{\mathbf{r}_k^{\|}(t)-\mathbf{r}_{k-1}^{\|}(t)}{\left|\mathbf{r}_k^{\|}(t)-\mathbf{r}_{k-1}^{\|}(t)\right|} \tag{5.7}$$

where $l_{\mathbf{r},i}^{\|}=\sum_{j=2}^{i}\left|\mathbf{r}_j^{\|}-\mathbf{r}_{j-1}^{\|}\right|$ is the string arc length for spatial coordinates (*x*, *y*, *z*) at image *i* after evolution, $L_{\mathbf{r}}^{\|}$ is the entire string arc length for molecular COM position after evolution, and *N* is the number of images used in the string. To prevent the abrupt change of the molecular orientation, a smooth function is used for interpolating between neighboring images[22]

$$\omega_i=(1-\delta)\omega_i+\frac{\delta}{2}(\omega_{i-1}+\omega_{i+1}) \tag{5.8}$$

where δ is the parameter to control the degree of smoothness. A small number, $\delta=1\times10^{-4}$, is used in this work to ensure the accuracy of molecular orientation that minimizes the external potential.

5.2.3 GPU Implementation

In implementing the string method to predict the diffusivity of polyatomic molecules in nanoporous materials, the computational cost is mostly affiliated with the calculation of the external potential and its derivates. In our previous work where the string method was applied to a single LJ particle, the external potential was pre-calculated by placing the guest molecule in a three-dimensional grid. The external potential at any point can be interpolated with a linear scheme. The procedure is not

directly applicable to polyatomic molecules because the grid size grows exponentially with the dimensionality. The calculation of the external potential as a function of position and orientation in a fine grid is not only computationally prohibitive, but it is also logistically challenging to store the potential energy data over such a large grid. Without the support of a supercomputer with an enormous memory, the external potential has to be calculated on-the-fly along with the string evolution.

In this work, we implement the massively parallel simplified string method with the graphic processing unit (GPU). Figure 5.3A shows a schematic procedure. Compared with central processing unit (CPU), GPU is designed for high-throughput data processing and paralleled tasks. GPU enables the parallel calculation of the external potential on-the-fly and is much less memory demanding because it involves only the external potential related to the images along the string instead of entire energy landscape for the guest molecule inside framework material. As the external potential and its derivates are calculated for each image on the string independent of each other, a paralleled implementation can significantly improve the speed and reduce the computational cost.

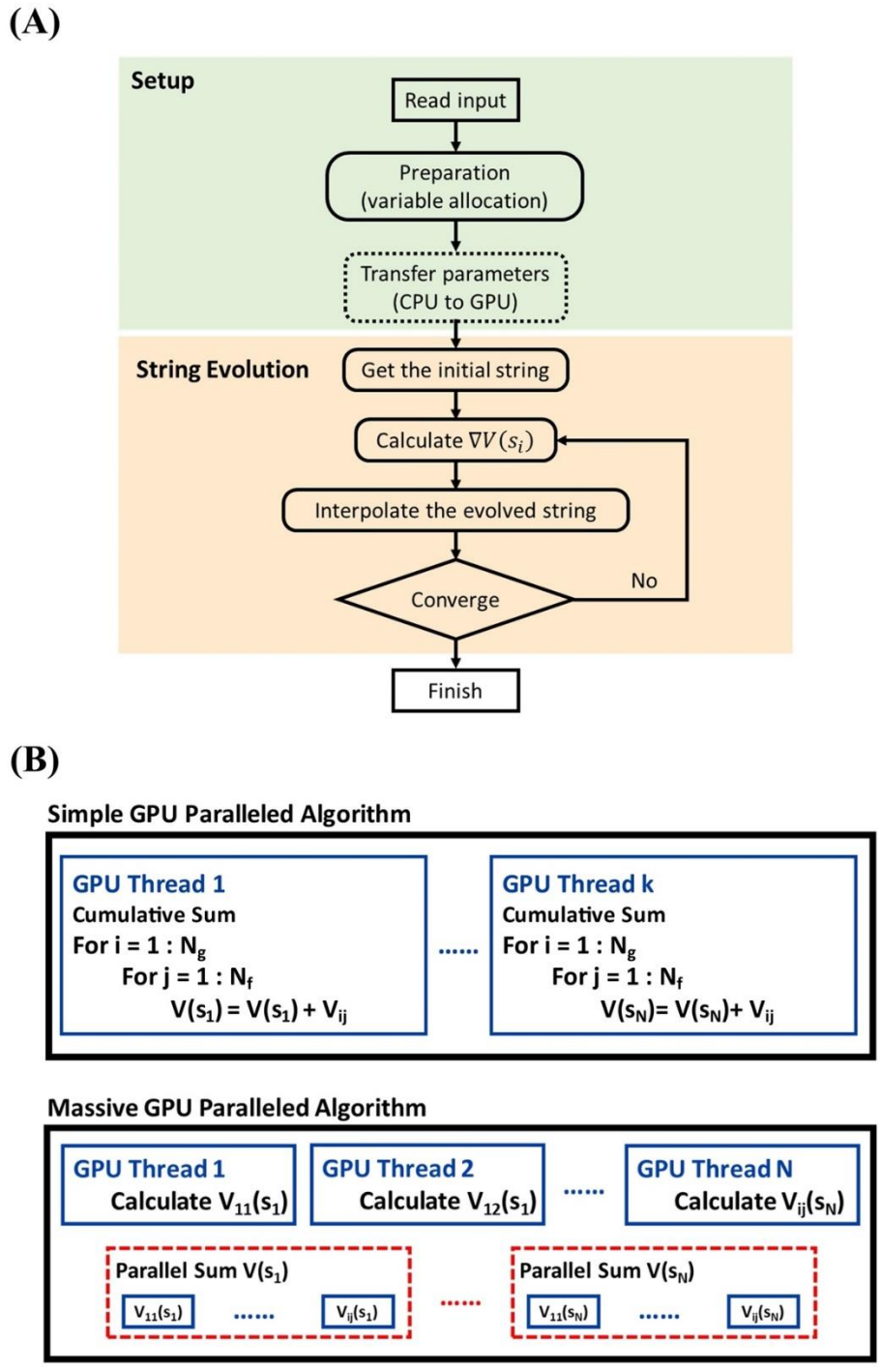

Figure 5.3. (A) Computational flowchart for the GPU implementation of the simplified string method. (B) Schematic illustration of simple and massive GPU paralleled algorithms for calculating the external potential for images on the string. Blue box represents the GPU thread.

We minimize the memory transfer between the host (CPU) and the device (GPU) in implementation of the string method. As shown in Figure 5.3, the memory transfer takes place only for reading the input information (such as the force-field parameters and the atomic structures of the nanoporous material and guest molecule), and for checking the convergence and outputting the final string configuration. The calculation is carried out only at the GPU device throughout the string evolution. The most expensive step in string evolution lies in the calculation of the derivatives of the external potential. Again, parallel implementation can significantly reduce the computational cost. As shown in Figure 5.3B, two different GPU paralleled algorithms have been implemented and tested in this work. To calculate the derivatives of the external potential, we may assign a GPU thread for a given set parameters (x, y, z, α, β, γ). For a given set model parameters, the external potential is calculated on a single GPU thread by the cumulative summation of all pair potentials (V_{ij}) between atom i from the polyatomic molecule (guest) and atom j from the nanoporous material (host). While this parallel scheme is intuitive and simple to implement, it does not utilize all available GPU threads especially when the number of images on the string is relatively small. Alternatively, we can assign one GPU thread for each pair of the interatomic potential (V_{ij}) and calculate the external potential by the summation of V_{ij} for all atomic pairs evaluated via multiple GPU threads. In this work, we use the CUDA UnBound (CUB) library, a configurable C++ template library developed by Nvidia for Compute Unified Device Architecture (CUDA), to carry out the summation of V_{ij} on GPU.[47] With this massively parallel implementation, all GPU threads can be fully utilized even when the number of string images is relatively small.

One caveat of the massively paralleled implementation is that it leads to a higher demand of the memory usage. Because it stores all the pairwise interaction before summation, the massive GPU paralleled algorithm limits its capability handling a large set of string images. It is also worth mentioning that similar GPU-accelerated algorithms can also be developed to calculate the potential energy surface of polyatomic molecule in nanoporous materials and we have demonstrated such implementation for molecules modeled by single LJ site in our previous work.[34] The thermodynamic quantities would enable a rapid evaluation of properties, such as zero-coverage adsorption amount, in nanoporous materials for gas storage and separation.

5.3 Results and Discussion

5.3.1 GPU Speedup

To benchmark different GPU-accelerated parallel methods for implementing the string method, we take the diffusion of an ethene molecule in MOF-5 as a model system. Figure 5.4 compares the computational costs for two implementations of GPU-accelerated parallel algorithms as a function of the system size as measured in terms of the number of V^{ext} calculations (# of V^{ext}) and the number of V_{ij} calculations (# of V_{ij}). Here, the speedup factor is obtained by the comparison of the performance for Nvidia Tesla P100, which is used for all our GPU calculations, with that for the serial CPU implementation on Intel Xeon E5-2640. Both simple and massive paralleled GPU implementations outperform the serial CPU implementation regardless of the system size. The number of V^{ext} calculations is solely determined by the number of images on the string, while the number of V_{ij} calculations for each image depends on the number of

atoms at the guest molecule and the number of atoms from the nanoporous material. When the number of images on the string (equivalently, the number of V^{ext} calculations) increases, the speedup factor rises exponentially for the simple GPU paralleled implementation until it reaches a plateau after all GPU threads are utilized. For the GPU device tested in this work, the speedup factor approaches an asymptotic limit when it processes more than 100,000 images along the diffusion pathway (viz., the string). The maximum speedup by the simple GPU paralleled algorithm is about 2500 folds of the CPU serial implementation.

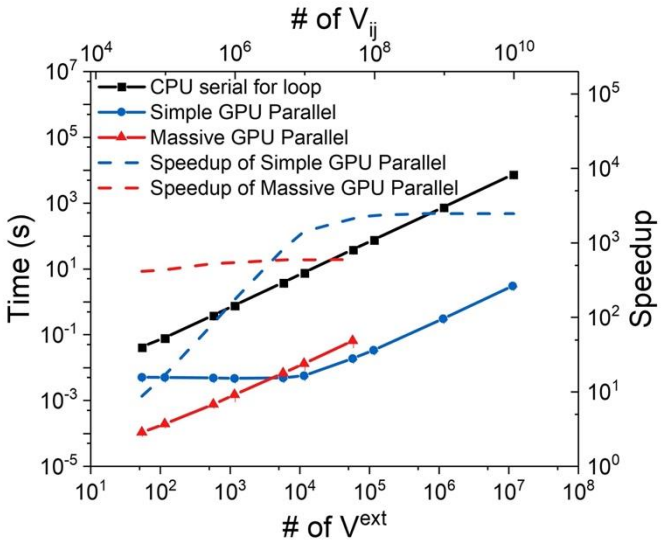

Figure 5.4. Comparison of the computational time versus the number of the total potential (V^{ext}) and pair potential (V_{ij}) evaluations for predicting ethene diffusion in MOF-5. The speedup factor is benchmarked with CPU calculations conducted on Intel Xeon E5-2640. All GPU calculations are carried out on Nvidia Tesla P100.

For the massive GPU paralleled implementation, the computational cost increases exponentially with the number of string images irrespective of the system size. Because the fully paralleled algorithm maximizes the usage of all active GPU threads, the speedup

factor, which is around 500, is almost independent of the number of string images. As mentioned above, the massive GPU paralleled implementation consumes more memory space than the simple GPU paralleled implementation (as shown in Figure 5.5). Thus, the upper limit for the number images that can be processed by the massive GPU paralleled implementation is much lower than that for simple GPU paralleled implementation. For most nanoporous materials considered in this work, the unit cell size varies from 10 Å to 30 Å, such that a string with hundreds of images would be sufficient to preserve all atomistic details along the minimum energy path (MEP). As a result, the massive GPU paralleled implementation has a better performance and thus used in all the following calculations.

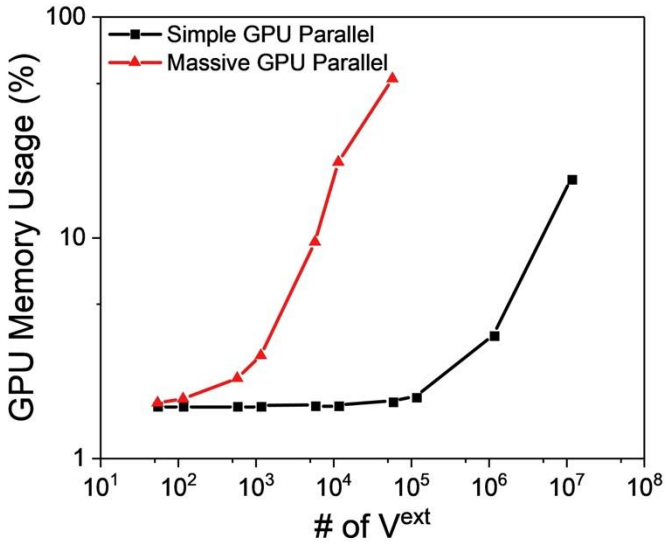

Figure 5.5. Percentage of GPU memory usage versus the number of V^{ext} calculations for ethene diffusion in MOF-5.

5.3.2 Calibration with MD simulation

We see above that excellent computational performance can be achieved by massively parallel GPU-accelerated implementation of the string method. But how accurate is the theoretical procedure for predicting diffusivity coefficients in comparison with conventional methods? In this section, we compare our theoretical predictions for the diffusion coefficients of 8 polyatomic molecules in MOF-5, a well-studied metal organic framework (MOF), with those from molecular dynamics (MD) simulations. As mentioned before, our theoretical predictions are based on the transition-state theory (TST) with the minimum energy path (MEP) calculated from the massive GPU paralleled implementation of the string method.

As shown in Figure 5.6A, the diffusion coefficients calculated in this work agree well with those from MD simulation over a broad range of values. The good agreement affirms the accuracy of the minimum energy path obtained from the GPU-accelerated calculations. Overall, TST predicts the self-diffusivity of various polyatomic molecules in MOF-5 slightly higher than that from MD simulation. The systematic error is introduced probably because TST neglects the barrier recrossing in molecular hopping.[48] The barrier recrossing of gas molecule becomes more significant especially at finite loadings. In this case, the dynamically corrected TST (dcTST) can better estimate the hopping rate ($k_{dcTST}=\kappa k_{TST}$) by correcting the recrossing event with the transmission coefficient (κ).[49, 50] Another possible reason is that most MD simulations are not carried in the single molecule limit thus the simulation results are affected by intermolecular interactions between the guest molecules. In principle, TST can be used to predict diffusivity

coefficients at finite gas pressure if MEP is replaced by the free-energy landscape. Alternatively, the self-diffusivity at finite loading can be calculated from the diffusivity coefficient at infinite dilution in combination with the excess-entropy scaling method.[33, 51] It is worth mentioning that, in comparison with MD, one of the most significant advantages of TST is computational efficiency. While it takes up to several thousands of CPU hours to simulate the diffusion coefficient of CO_2 (at the scale of 10^{-9} m^2/s), a relatively small polyatomic molecule, the same calculation can be finished in this work within 30 seconds for each material by using a single GPU card. Although both gas molecules (guest) and framework materials (host) are assumed to be rigid in this work, the string method can also be used to obtain the minimum energy path when flexibility of gas molecules and framework is significant such as large molecule squeezing through tight aperture. In the latter case, the computation will be more demanding because we need to consider both the guest-host interactions and the intramolecular potential.

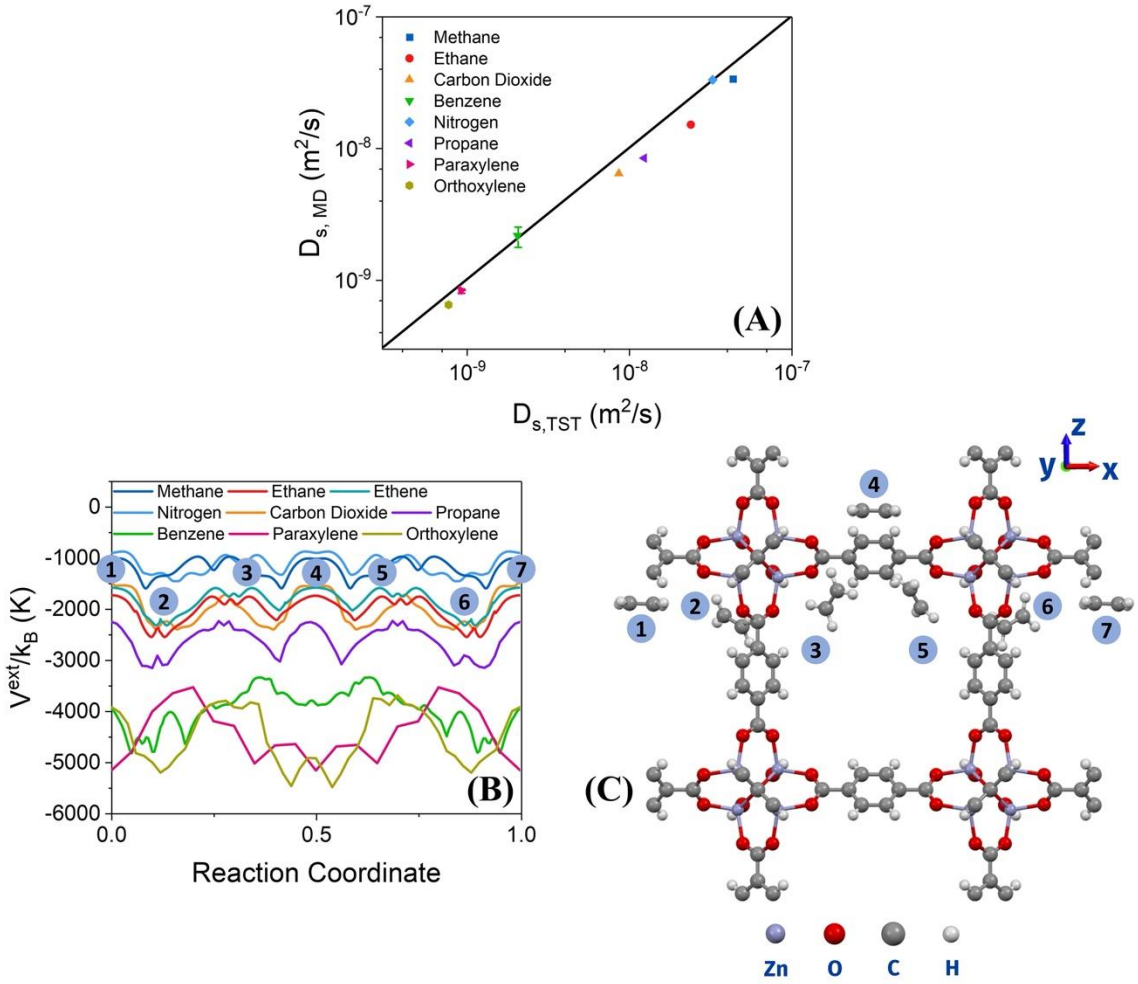

Figure 5.6. (A) Comparison of diffusion coefficients for 8 polyatomic molecules in MOF-5 predicted by the transition-state theory (TST) and by molecular dynamics (MD) simulations. The MD results are from the literature.[16, 52-55] (B) The minimum energy paths calculated from the GPU paralleled implementation of the simplified string method versus the reaction coordinate. Here the numbers 1-7 stand for images along the minimum energy path in MOF-5 for the ethene molecule. (C) The positions and orientations of the ethene molecule corresponding to the 7 images labeled in (B). (Ethene is modeled as a diatomic molecule according to TraPPE-UA force field. The molecular structures of ethene in (C) are only for illustration purpose).

Figure 5.6B presents the minimum energy paths (MEPs) for 8 polyatomic molecules tested in this work. For small molecules such as ethane, carbon dioxide and propane, their energy landscapes along the diffusion path have a similar shape and exhibit

only slightly different potential energy barriers, suggesting a similar hopping pattern for the diffusion of small molecules in MOF-5. According to the molecular orientation along the MEP shown in Figure 5.6C, the center of mass (COM) for the ethene molecule stays close to the metal cluster and organic linker in MOF-5 instead of going through MOF-5 in the center of pore in order to maintain the minimized hots-guest interactions. When the ethene molecule enters across the pore, its orientation also changes so that the smaller edge of the molecular plane would be directed toward the angle minimizing the external potential.

Figure 5.6B shows the minimum energy paths (MEPs) for benzene, paraxylene and orthoxylene. These paths are significantly different from those for smaller molecules such as ethene due to the molecular size. Even when comparing the potential energies and molecular orientation along MEPs of large molecules, they are quite different from each other because, for large molecules, even slight modifications (add/relocate) on the functional group would lead to significant difference in their preferred orientation at the energy barrier and rotational and translational activation energy along the reaction coordinate. In Figure 5.7, we compare the position and molecular orientation of p-type xylene with those corresponding to o-type xylene along the MEP in MOF-5. As discovered in NMR studies and by MD simulation,[56] the p-type xylene molecule, especially its methyl group, is located around the pore center when the hopping takes place between neighboring cages (shown in Figure 5.8). For the o-type xylene, the COM position is close to the metal cluster (viz., at the corner of pore) before hoping to the neighboring cages. In addition, the molecular orientation of p-type xylene changes much

less than o-type xylene along the MEP because p-type xylene has higher activation energy due to the rotational move as shown in Figure 5.7.

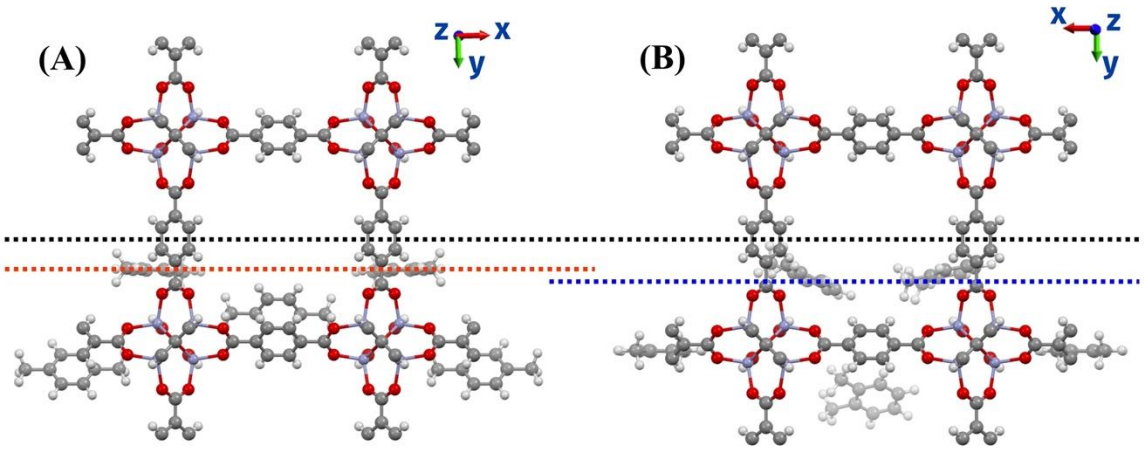

Figure 5.7. Molecular positions and orientations of p-type xylene (A) and o-type xylene (B) in MOF-5 along the minimum energy path. (P-type and o-type xylene are modeled as eight united group sites according to TraPPE-UA force field. The detailed guest molecular structures are for illustration purpose only). Black, red and blue dashed line represents the center of pore in MOF-5, center of mass for p-type xylene when crossing the pore and center of mass for o-type xylene when crossing the pore, respectively.

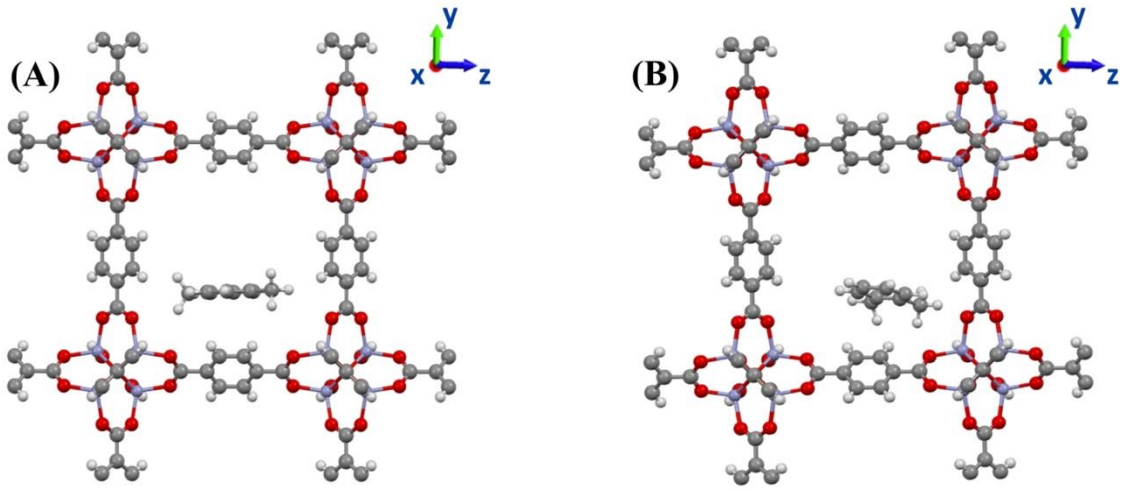

Figure 5.8. Molecular position and orientation of p-type xylene (A) and o-type xylene (B) in MOF-5 when hopping across pore (along x direction) along the minimum energy path. (P-type and o-type xylene are modeled as eight united group sites according to TraPPE-UA force field. The detailed guest molecular structures are for illustration purpose only).

5.3.3 High-Throughput Screening

Efficient prediction of diffusivity will likely open up opportunities for high-throughput screening and, eventually, for the inverse design of nanoporous materials for practical applications such as gas separation. Here, we demonstrate the capability of massively parallel GPU-accelerated string method for high-throughput screening of MOFs useful for ethane/ethene separation, a challenging yet important task in the chemical industry. Since ethane and ethene molecules have similar physical characteristics such as the size and shape, previous studies indicate that MOFs with pore limit diameter (PLD) between 3 Å and 4 Å are most efficient in terms of selectivity.[57] We use PLD as an initial criterion to select 3080 candidates from the computational-ready, experimental (CoRE) MOF 2019 database which covers over 14 000 porous structures.[58]

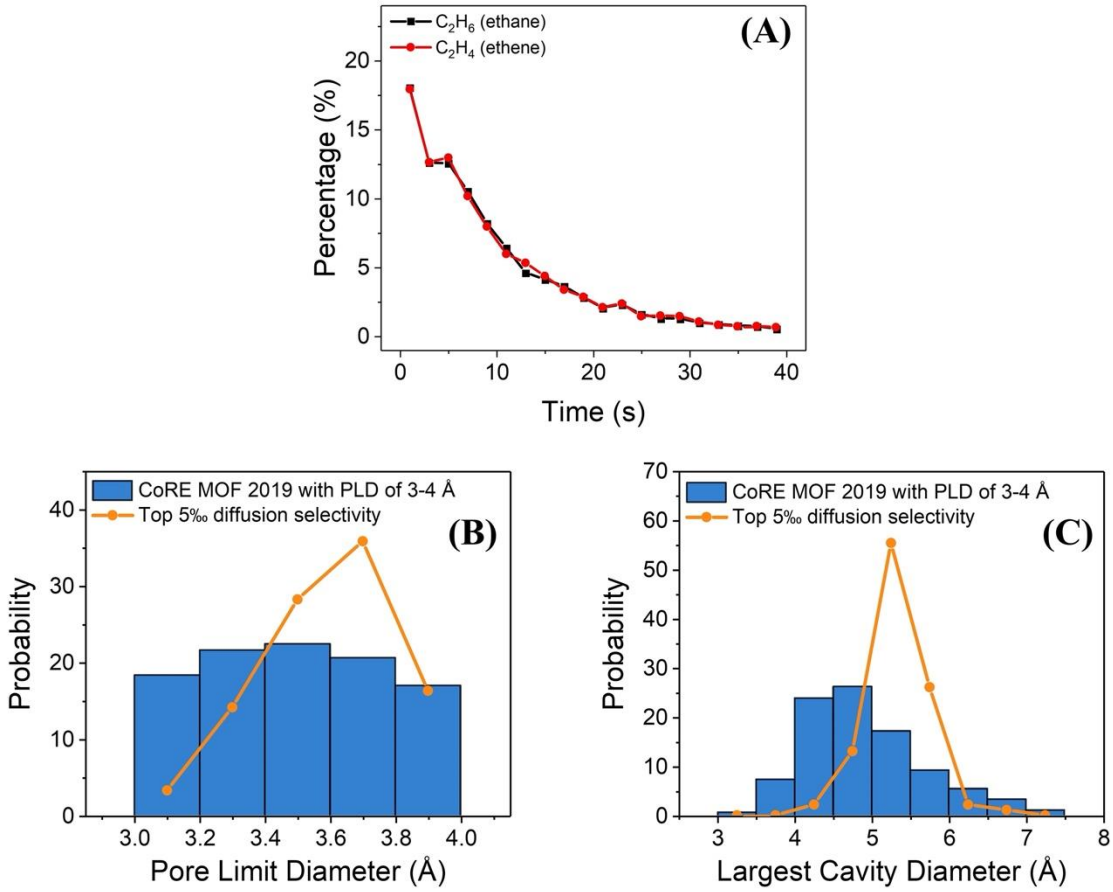

Figure 5.9. (A) Distribution of the computational time in high-throughput screening of 3080 MOF candidates for the separation of ethane/ethene gases at room temperature. Distributions of (B) the pore limit diameter (PLD) and (C) largest cavity diameter (LCD) of those MOFs with the highest diffusion selectivity.

As shown in Figure 5.9A, the diffusivity calculation for most MOFs (more than 90%) can be accomplished within less than 30 seconds, which is significantly faster than conventional methods such as molecular dynamics simulation (up to hundreds of CPU hours per material).[33] The massively parallel GPU-accelerated string method is also much faster than emerging methods that search only the tunnel space or the transition state with the polyatomic molecules represented by a single-site LJ model, which typically cost ~0.5 CPU hour per material.[10] The computational cost of the string

method depends not only on how many iterations it would take to identify the minimum energy path but also on the number of atoms in the system including those from both the guest molecule and the framework material (host). Ethene (3.23×4.18×4.84 Å) and ethane (3.81×4.08×4.82 Å) have similar molecular shape and size. Because both are modeled as diatomic molecules in the TraPPE force field, the computational costs of finding the minimum energy paths for ethane and ethene are almost identical.[40]

In Figure 6B and C, we present the distributions of the pore limit diameter (PLD) and the largest cavity diameter (LCD) for MOFs in the database with top 0.5% diffusion selectivity for the separation of ethane/ethene gases at room temperature. Here, the diffusion selectivity is calculated from the ratio of diffusivity, $S_{diff,1/2} = D_{0,1} / D_{0.2}$ and 1 and 2 refers to ethene and ethane, respectively. For materials in the CoRE MOF 2019 database, the PLDs are evenly distributed between 3 Å and 4 Å, whereas the distribution of MOFs with the highest diffusion selectivity has a notable peak between 3.4 Å and 3.8 Å. Although the range of the PLD from 3.4 Å to 3.8 Å is significantly smaller than the kinetic diameter of ethane (4.443 Å) and ethene (4.163 Å) molecules derived from the second virial coefficients, it falls into the perfect range for separating ethene from ethane according to the molecular size and shape. The peak value (3.4-3.8 Å) is larger than the smallest edge of an ethene molecule (3.28 Å) but smaller than that of ethane (3.81 Å).[57] Compared with the LCD distribution of MOFs in the background, which follows approximately a normal distribution with the mean between 4.5 Å and 5 Å, MOFs with top 0.5% diffusion selectivity has a slightly higher mean, between 5 Å and 5.5 Å, in the LCD distribution. According to our previous work,[29, 59, 60] a nanoporous material

with the LCD larger than the molecular size would impose more attraction along the minimum energy path, which is beneficial to achieve the diffusivity coefficient at the scale of practical interest.

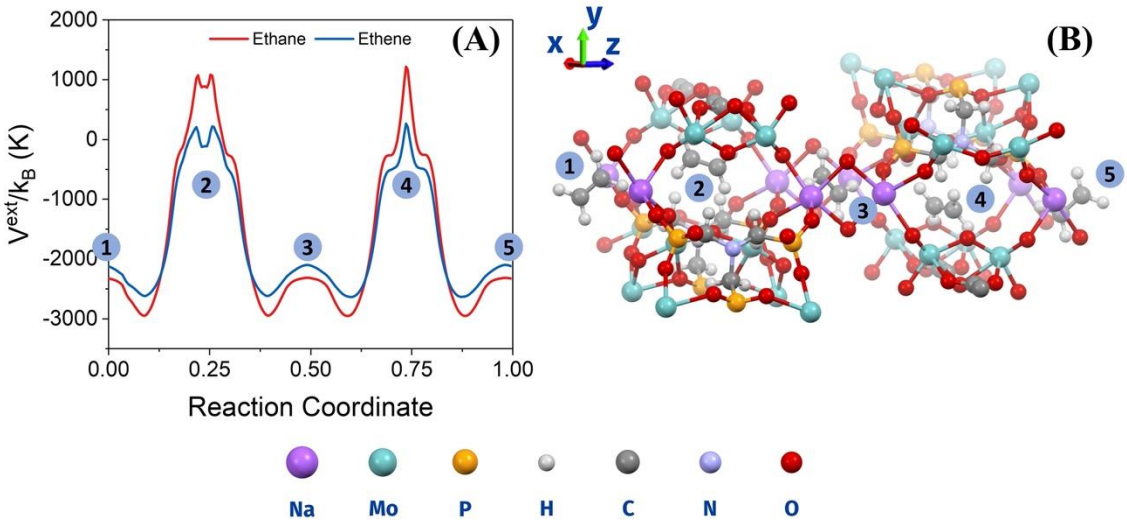

Figure 5.10. (A) Energy landscape along the minimum energy path for ethane and ethene in YIGFIF, a nanoporous material with the highest diffusion selectivity for the separation of ethane and ethene gases at 300 K. (B) The molecular position and orientation of an ethene molecule along the minimum energy path in YIGFIF.

In the Table 5.2, we present the diffusion coefficients and structural properties of top 10 MOFs with the highest diffusivity. Among all MOFs investigated in this work, YIGFIF has a diffusion selectivity of 57.69, which is the highest for the separation of ethane/ethene at 300 K. Figure 5.10 shows the energy landscape of ethane and ethene molecules along the minimum energy path in YIGFIF. For ethane diffusing along the MEP in YIGFIF, the COM position is almost identical to that for ethene. As shown in Figure 5.10B, the rotation of an ethene molecule inside YIGFIF is restricted due to strong confinement along the MEP in YIGFIF (PLD: 3.38 Å and LCD 5.02 Å). However, the nanoporous material exerts a repulsive energy on ethane at the transition state much

stronger than that on ethene. The larger diffusion barrier may be attributed to the minimum cross-section area of YIGFIF (~3.38 Å ×4.94 Å), which can be utilized to sieve ethane and ethene with an excellent diffusion selectivity. Although YIGFIF is the most promising MOF candidate for the separation of ethane/ethene as the membrane materials according to our screening, it has not yet been experimentally tested for any practical applications.[61] Compared with conventional ethene-selective adsorbent materials (selectivity up to 48.7), membrane separation with YIGFIF would be much less energy-intensive for industrial application with around 20% higher separation selectivity.[57, 62] For the state-of-art ethane-selective adsorbent materials (selectivity up to 4.4), YIGFIF can achieve a much higher separation selectivity of ethane/ethene.[63]

Table 5.2. Diffusion coefficients and structural properties of MOFs with the highest diffusion selectivity for ethane/ethene separation

Ref Code	$S_{diff,C_2H_4/C_2H_6}$	PLD (Å)	LCD (Å)	void fraction	$D_{C_2H_4}$ (m^2/s)	$D_{C_2H_6}$ (m^2/s)
YIGFIF	57.69	3.376	5.022	0.521	2.12×10^{-11}	3.68×10^{-13}
SUPSIG02	53.67	3.883	5.892	0.522	2.41×10^{-12}	4.50×10^{-14}
OGIBOV	50.00	3.663	4.920	0.428	1.45×10^{-12}	2.90×10^{-14}
WIXRUQ	37.87	3.644	5.220	0.400	4.09×10^{-12}	1.08×10^{-13}
YAQZEX	34.22	3.356	5.113	0.436	9.17×10^{-11}	2.68×10^{-12}
WUNTUV	28.85	3.469	5.530	0.476	4.27×10^{-13}	1.48×10^{-14}
WEDXAG	24.81	3.751	5.270	0.374	2.58×10^{-12}	1.04×10^{-13}
GAWCUF	24.17	3.203	5.312	0.461	3.19×10^{-13}	1.32×10^{-14}
YOFXEY	23.86	3.801	5.346	0.463	9.21×10^{-12}	3.86×10^{-13}
TEWFOS	23.64	3.413	5.270	0.449	1.17×10^{-11}	4.95×10^{-13}

5.4 Conclusion

In this work, we have implemented a GPU-accelerated string method to calculate the minimum energy path (MEP) for polyatomic molecules in nanoporous materials. The MEP calculation is essential for predicting diffusivity using the transition-state theory.

Both simple GPU parallel algorithm and massive GPU parallel algorithm are tested and benchmarked with serial CPU calculations. Compared with the serial CPU implementation on Intel Xeon E5-2640, GPU implementations on Nvidia Tesla P100 may speedup the diffusivity calculation up to ~2500 folds via the simple GPU parallel algorithm. The outstanding performance is attributed to massive threads available on GPU and the minimized memory transfer between CPU (host) and GPU (device). Although the simple GPU paralleled implementation can achieve up to three orders of magnitude speedup compared to the serial CPU implementation, the speedup factor depends on the number of images on the diffusion pathway (viz. the reaction coordinate represented by a string) and is much lower than that could be achieved by massive GPU paralleled implementation, especially when the number of images on the string is relatively small due to the insufficient usage of GPU threads. For massive GPU parallel implementation, a constant speedup factor around 500 is achieved regardless of the number of images on the string, an indication of excellent parallelization for MEP calculations. Because most nanoporous materials have the largest edge in the unit cell less than 30 Å, the massive GPU paralleled implementation of the string method is more advantageous for calculation of MEP for polyatomic molecules in nanoporous materials.

The diffusion coefficients of guest molecules in nanoporous materials can be calculated from MEP via the transition-state theory (TST). The results are compared with the diffusion coefficients from molecular dynamics (MD) simulation for 8 polyatomic molecules in MOF-5. Excellent agreement between theory and simulation is achieved, further indicating the accuracy of MEP obtained by the GPU-accelerated string method.

Because TST underestimates the diffusion barrier and recrossing of molecular hopping, and because most MD simulations are not run in the single-molecule limit, the diffusivity from TST is slightly larger than that from simulation. While the diffusion of small molecules (e.g., ethene, nitrogen and carbon dioxide) in MOF-5 shares a similar hopping pattern, large molecules such as p-type xylene and o-type xylene have significantly different trajectories for their positions and orientations along the minimum energy path. Different from MD simulation whereby diffusivity is calculated from the statistical average of molecular movements in random, the minimum energy path calculated from GPU-accelerated massively parallel string method offers the microscopic details of molecular hopping that can be utilized to guide the rational design of nanoporous materials for the separation of polyatomic molecules.

Finally, we have demonstrated the capability of massively parallel GPU-accelerated string method for high-throughput screening MOFs for the separation of ethane and ethene, two polyatomic molecules of similar size and shape that are of tremendous importance for the chemical industry. In the high-throughput screening calculations, 3080 MOFs are selected from computational-ready, experimental MOF database CoRE MOF 2019 according to their pore limit diameters (PLDs). The massively parallel GPU-accelerated string method is used to calculate the diffusion selectivity for ethane/ethene separation. The diffusivity calculation can be completed within 30 seconds per material for more than 90% of MOFs, which is significantly faster than MD simulation (cost ~ hundreds of CPU hours per material). The GPU calculation outperforms even emerging methods such as TuTraSt (cost ~ 0.5 CPU hour for a single

LJ site).[10, 33] While the calculation of slow diffusion behavior (less than 1×10^{-12} m^2/s) is computationally prohibitive for molecular dynamics, the computational cost for the string method does not change with the scale of the diffusion coefficient. According to our high-throughput screening calculations, MOFs with the pore limit diameter (PLD) from 3.4 to 3.8 Å and the largest cavity diameter (LCD) between 5 and 5.5 Å can efficiently separate ethene (with the molecular dimensions of 3.23×4.18×4.84 Å) from ethane (with the molecular dimensions of 3.81×4.08×4.82 Å) by their molecular sizes and shapes. YIGFIF (PLD= 3.38 Å and LCD=5.02 Å) from the CoRE MOF 2019 database has been identified with the highest diffusion selectivity for ethene/ethane separation, with a theoretical selectivity that can reach up to 57.69 at 300 K.

The GPU-accelerated massively parallel implementation of string method enables efficient and accurate calculation of diffusion coefficients for polyatomic molecules in nanoporous materials. We expect that the computational platform will be generally useful for high-throughput screening of nanoporous materials, for example, as the membrane for the separation of polyatomic molecules. It can also be used to construct high-fidelity properties database for the inverse design of nanoporous materials.

Bibliography

1. Rangnekar, N., et al., *Zeolite membranes - a review and comparison with MOFs.* Chem Soc Rev, 2015. **44**(20): p. 7128-54.

2. Al-Rowaili, F.N., et al., *A Review on Recent Advances for Electrochemical Reduction of Carbon Dioxide to Methanol Using Metal-Organic Framework (MOF) and Non-MOF Catalysts: Challenges and Future Prospects.* Acs Sustainable Chemistry & Engineering, 2018. **6**(12): p. 15895-15914.

3. Ali, M., et al., *Recent advancements in MOF‐based catalysts for applications in electrochemical and photoelectrochemical water splitting: A review.* International Journal of Energy Research, 2021. **45**(2): p. 1190-1226.

4. Li, Y., et al., *Laminated self-standing covalent organic framework membrane with uniformly distributed subnanopores for ionic and molecular sieving.* Nat Commun, 2020. **11**(1): p. 599.

5. Fan, H., et al., *MOF-in-COF molecular sieving membrane for selective hydrogen separation.* Nat Commun, 2021. **12**(1): p. 38.

6. Li, J., et al., *Two-Dimensional Covalent Organic Frameworks (COFs) for Membrane Separation: a Mini Review.* Industrial & Engineering Chemistry Research, 2019. **58**(34): p. 15394-15406.

7. Chung, Y.G., et al., *Computation-Ready Experimental Metal-Organic Framework (CoRE MOF) 2019 Dataset*. 2019: Zenodo.

8. Moghadam, P.Z., et al., *Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future.* Chemistry of Materials, 2017. **29**(7): p. 2618-2625.

9. Tranchemontagne, D.J., et al., *Secondary building units, nets and bonding in the chemistry of metal-organic frameworks.* Chem Soc Rev, 2009. **38**(5): p. 1257-83.

10. Mace, A., S. Barthel, and B. Smit, *Automated Multiscale Approach To Predict Self-Diffusion from a Potential Energy Field.* J Chem Theory Comput, 2019. **15**(4): p. 2127-2141.

11. Haldoupis, E., S. Nair, and D.S. Sholl, *Efficient calculation of diffusion limitations in metal organic framework materials: a tool for identifying materials for kinetic separations.* J Am Chem Soc, 2010. **132**(21): p. 7528-39.

12. Abraham, J., et al., *Tunable sieving of ions using graphene oxide membranes.* Nature Nanotechnology, 2017. **12**(6): p. 546-550.

13. Wijmans, J.G. and R.W. Baker, *The Solution-Diffusion Model - a Review.* Journal of Membrane Science, 1995. **107**(1-2): p. 1-21.

14. Dubbeldam, D. and R.Q. Snurr, *Recent developments in the molecular modeling of diffusion in nanoporous materials.* Molecular Simulation, 2007. **33**(4-5): p. 305-325.

15. Jobic, H., et al., *Influence of isotherm inflection on the loading dependence of the diffusivities of n-hexane and n-heptane in MFI zeolite. Quasi-elastic neutron scattering experiments supplemented by molecular simulations.* J Phys Chem B, 2006. **110**(5): p. 2195-201.

16. Ford, D.C., et al., *Self-Diffusion of Chain Molecules in the Metal-Organic Framework IRMOF-1: Simulation and Experiment.* J Phys Chem Lett, 2012. **3**(7): p. 930-3.

17. Urbanczyk, M., W. Kozminski, and K. Kazimierczuk, *Accelerating diffusion-ordered NMR spectroscopy by joint sparse sampling of diffusion and time dimensions.* Angew Chem Int Ed Engl, 2014. **53**(25): p. 6464-7.

18. Price, K.E., L.H. Lucas, and C.K. Larive, *Analytical applications of NMR diffusion measurements.* Anal Bioanal Chem, 2004. **378**(6): p. 1405-7.

19. Altintas, C. and S. Keskin, *Molecular simulations of MOF membranes for separation of ethane/ethene and ethane/methane mixtures.* RSC Adv, 2017. **7**(82): p. 52283-52295.

20. Avci, G., S. Velioglu, and S. Keskin, *High-Throughput Screening of MOF Adsorbents and Membranes for H2 Purification and CO2 Capture.* ACS Appl Mater Interfaces, 2018. **10**(39): p. 33693-33706.

21. Kärger, J., D.M. Ruthven, and D.N. Theodorou, *Diffusion in Nanoporous Materials, 2 Volume Set*. 2012: John Wiley & Sons.

22. Maragliano, L., et al., *String method in collective variables: minimum free energy paths and isocommittor surfaces.* J Chem Phys, 2006. **125**(2): p. 24106.

23. E, W., W. Ren, and E. Vanden-Eijnden, *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events.* J Chem Phys, 2007. **126**(16): p. 164103.

24. Weinan, E., W.Q. Ren, and E. Vanden-Eijnden, *String method for the study of rare events.* Physical Review B, 2002. **66**(5): p. 052301.

25. Jónsson, H., G. Mills, and K.W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions.* 1998.

26. Sheppard, D., et al., *A generalized solid-state nudged elastic band method.* J Chem Phys, 2012. **136**(7): p. 074103.

27. Cvitas, M.T. and S.C. Althorpe, *Locating Instantons in Calculations of Tunneling Splittings: The Test Case of Malonaldehyde.* J Chem Theory Comput, 2016. **12**(2): p. 787-803.

28. Zhou, M., A. Vassallo, and J. Wu, *Toward the inverse design of MOF membranes for efficient D2/H2 separation by combination of physics-based and data-driven modeling.* Journal of Membrane Science, 2020. **598**: p. 117675.

29. Zhou, M.S., et al., *Fractionation of Isotopic Methanes with Metal-Organic Frameworks.* Journal of Physical Chemistry C, 2019. **123**(12): p. 7397-7407.

30. Mahdizadeh, S.J. and E.K. Goharshadi, *Multicomponent gas separation and purification using advanced 2D carbonaceous nanomaterials.* Rsc Advances, 2020. **10**(41): p. 24255-24264.

31. Kocer, C.P., et al., *Lithium Diffusion in Niobium Tungsten Oxide Shear Structures.* Chem Mater, 2020. **32**(9): p. 3980-3989.

32. Coifman, R.R., et al., *Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems.* Multiscale Modeling & Simulation, 2008. **7**(2): p. 842-864.

33. Tian, Y., X. Xu, and J. Wu, *Thermodynamic Route to Efficient Prediction of Gas Diffusivity in Nanoporous Materials.* Langmuir, 2017. **33**(42): p. 11797-11803.

34. Zhou, M. and J. Wu, *A GPU implementation of classical density functional theory for rapid prediction of gas adsorption in nanoporous materials.* J Chem Phys, 2020. **153**(7): p. 074101.

35. Acun, B., et al., *Scalable Molecular Dynamics with NAMD on the Summit System.* IBM J Res Dev, 2018. **62**(6): p. 1-9.

36. Nitsche, M.A., et al., *GPU Accelerated Implementation of Density Functional Theory for Hybrid QM/MM Simulations.* J Chem Theory Comput, 2014. **10**(3): p. 959-67.

37. Rappe, A.K. and W.A. Goddard, *Charge Equilibration for Molecular-Dynamics Simulations.* Journal of Physical Chemistry, 1991. **95**(8): p. 3358-3363.

38. Rappe, A.K., et al., *Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations.* J Am Chem Soc, 1992. **114**(25): p. 10024-10035.

39. Dubbeldam, D., et al., *RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials.* Mol Simul, 2016. **42**(2): p. 81-101.

40. Wu, Y., et al., *Adsorption and separation of ethane/ethylene on ZIFs with various topologies: Combining GCMC simulation with the ideal adsorbed solution theory (IAST).* Chemical Engineering Science, 2015. **124**: p. 144-153.

41. Wick, C.D., M.G. Martin, and J.I. Siepmann, *Transferable potentials for phase equilibria. 4. United-atom description of linear and branched alkenes and alkylbenzenes.* The Journal of Physical Chemistry B, 2000. **104**(33): p. 8008-8016.

42. 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.

43. Amirjalayer, S. and R. Schmid, *Adsorption of Hydrocarbons in Metal–Organic Frameworks: A Force Field Benchmark on the Example of Benzene in Metal–Organic Framework 5.* The Journal of Physical Chemistry C, 2012. **116**(29): p. 15369-15377.

44. Willems, T.F., et al., *Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials.* Microporous and Mesoporous Materials, 2012. **149**(1): p. 134-141.

45. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

46. Macrae, C.F., et al., *Mercury 4.0: From visualization to analysis, design and prediction.* Journal of applied crystallography, 2020. **53**(1): p. 226-235.

47. Merrill, D., *Cuda unbound (cub) library.* NVIDIA-Labs, 2015.

48. Vanden-Eijnden, E. and F.A. Tal, *Transition state theory: variational formulation, dynamical corrections, and error estimates.* J Chem Phys, 2005. **123**(18): p. 184103.

49. Beerdsen, E., D. Dubbeldam, and B. Smit, *Understanding diffusion in nanoporous materials.* Phys Rev Lett, 2006. **96**(4): p. 044501.

50. Dubbeldam, D., et al., *Molecular simulation of loading-dependent diffusion in nanoporous materials using extended dynamically corrected transition state theory.* Journal of Chemical Physics, 2005. **122**(22): p. 224712.

51. Dyre, J.C., *Perspective: Excess-entropy scaling.* J Chem Phys, 2018. **149**(21): p. 210901.

52. Skoulidas, A.I. and D.S. Sholl, *Self-diffusion and transport diffusion of light gases in metal-organic framework materials assessed using molecular dynamics simulations.* Journal of Physical Chemistry B, 2005. **109**(33): p. 15760-15768.

53. Cabrales-Navarro, F.A., J.L. Gomez-Ballesteros, and P.B. Balbuena, *Molecular dynamics simulations of metal-organic frameworks as membranes for gas mixtures separation.* Journal of Membrane Science, 2013. **428**: p. 241-250.

54. Amirjalayer, S. and R. Schmid, *Mechanism of benzene diffusion in MOF-5: A molecular dynamics investigation.* Microporous and Mesoporous Materials, 2009. **125**(1-2): p. 90-96.

55. Witherspoon, V.J., et al., *Translational and rotational motion of C8 aromatics adsorbed in isotropic porous media (MOF-5): NMR studies and MD simulations.* The Journal of Physical Chemistry C, 2017. **121**(28): p. 15456-15462.

56. Witherspoon, V.J., et al., *Translational and Rotational Motion of C8 Aromatics Adsorbed in Isotropic Porous Media (MOF-5): NMR Studies and MD Simulations.* Journal of Physical Chemistry C, 2017. **121**(28): p. 15456-15462.

57. Bao, Z., et al., *Molecular Sieving of Ethane from Ethylene through the Molecular Cross - Section Size Differentiation in Gallate - based Metal – Organic Frameworks.* Angewandte Chemie International Edition, 2018. **57**(49): p. 16020-16025.

58. Chung, Y.G., et al., *Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019.* Journal of Chemical and Engineering Data, 2019. **64**(12): p. 5985-5998.

59. Zhou, M., A. Vassallo, and J. Wu, *Toward the inverse design of MOF membranes for efficient D2/H2 separation by combination of physics-based and data-driven modeling.* Journal of Membrane Science, 2020. **598**.

60. Wang, J.Q., et al., *Computational screening and design of nanoporous membranes for efficient carbon isotope separation.* Green Energy & Environment, 2020. **5**(3): p. 364-373.

61. Yang, L., et al., *Self-assembly of two ring-shaped hexanuclear Mo(VI) clusters.* Crystengcomm, 2013. **15**(27): p. 5452-5457.

62. Yang, S.H., et al., *Supramolecular binding and separation of hydrocarbons within a functionalized porous metal-organic framework.* Nature Chemistry, 2015. **7**(2): p. 121-129.

63. Li, L., et al., *Ethane/ethylene separation in a metal-organic framework with iron-peroxo sites.* Science, 2018. **362**(6413): p. 443-446.

Chapter 6. Towards the Inverse Design of MOF Membranes for Efficient D_2/H_2 Separation by Combination of Physics-Based and Data-Driven Modeling

Hydrogen isotopes are useful for scientific research, energy generation and medical treatment. However, their industrial production is expensive because conventional processes for separation of hydrogen isotopologues are mostly based on energy-intensive macroscopic procedures with extremely low separation efficiency. Metal-organic frameworks (MOFs) provide a promising route to D_2/H_2 separation by leveraging their well-defined chemical and structural features. In this chapter, we report high-throughput screening of 12,723 experimentally synthesizable MOF membranes for D_2/H_2 separation by predicting gas adsorption and transport properties underpinning the separation efficiency. A membrane performance score is introduced to identify top ranked MOFs with the best selectivity and capacity. The extensive data generated from the physics-based modeling enables application of machine learning methods to predict desirable features of novel nanoporous materials for more efficient separation of hydrogen isotopes.

6.1 Introduction

Isotopologues are chemical species with the same molecular structure but a different number of neutrons in certain atom types. Because of their unique properties, hydrogen isotopologues have been extensively used in scientific research, energy production and medical treatment, ranging from neutron scattering[1, 2], isotopic tracing[3-5] and nuclear fusion reaction[6], to medical imaging and cancer therapy[7]. As the only difference is the number of neutrons, hydrogen isotopologues have virtually

identical chemical properties, making their separation at industrial scale exceedingly difficult. Conventional processes like cryogenic distillation and thermal diffusion suffer from low separation efficiency (e.g., the selectivity of D_2/H_2 is only about 1.5 at 24 K) and intensive energy consumption[8]. Whereas newly proposed technologies (e.g., atomic vapor laser isotope separation[9] and magnetically activated and guided isotope separation[10]) can achieve higher separation selectivity and consume less energy, their capacity for D_2/H_2 separation is severely limited thus hampering industrial applications.

Recently, metal-organic frameworks (MOFs) have been proposed for more efficient separation of hydrogen isotopes by utilizing quantum sieving effects. The mechanism was initially discovered by Beenakker and coworkers in 1995 and can be attributed to the disparity in the zero-point energies of isotopologues [11]. For gas separation with porous materials, the quantum sieving effect is most pronounced when the difference between the pore size and the diameter of gas molecules becomes comparable with the molecular de Broglie length. Because a hydrogen isotopologue shows a smaller effective size and a higher binding energy as the number of neutrons increases, a porous material adsorbs a heavier isotopologue more favorably, and makes it diffuse faster, than a lighter isotopologue. By tuning the difference in the zero-point energies of adsorbed hydrogen isotopologues, FitzGerald et al. demonstrated that quantum sieving could lead to D_2/H_2 selectivity of 1.5 at a temperature as high as 150 K[12]. Because the good performance is mainly attributed to the difference between the binding energies of hydrogen isotopes, the procedure is also known as chemical-affinity

quantum sieving (CAQS). A further improvement of the separation performance could be achieved by combing different quantum sieving mechanisms[13].

MOFs have been recognized as one of the most promising nanoporous structures for gas storage and separation. Compared with alternative nanoporous materials like activated carbons or zeolites, MOFs have the advantages of tunable pore size, geometry, and local chemical composition[14-17]. Besides, MOFs are particularly promising for D_2/H_2 separation not only because of ultrahigh porosity and large specific surface area but also for the wide varieties of metal clusters and organic linkers that can be finely tuned to amplify the quantum sieving effects[18]. Previous experimental research and molecular dynamics (MD) simulation have shown that MOF structures can be utilized to separate D_2/H_2 with a selectivity up to 41.4 at 20 K[13, 19-21]. In addition, a flexible MOF (MIL-53) has been investigated for separation of D_2/H_2 at 40 K; a selectivity of 13.6 was achieved by controlling the "breathing effect"[22]. Impressive selectivity for D_2/H_2 separation was also predicted for other sub-nanometer structures such as carbon nanotubes (CNTs)[23, 24], albeit they are less promising for industrial applications due to difficulties in materials synthesis[18].

Whereas quantum sieving effects are sensitive to the microscopic details of gas-pore interactions, such effects have not been systematically investigated and their influence on D_2/H_2 separation remains largely unknown. Previously[25, 26], we studied separation of isotopic methanes by high-throughput screening of both hypothetical and experimentally attainable MOF databases based on classical density functional theory (cDFT) calculations. In this work, we demonstrate that similar procedures can be

generalized to account for the quantum sieving effects by using the Feynman-Hibbs (FH) method[27]. Through theoretical prediction of Henry's constants and self-diffusivity coefficients, the physics-based models allow us to evaluate 12,723 experimentally synthesizable MOF structures for D_2/H_2 separation by using the membrane performance score. The extensive properties data generated from physic-based modeling provide a sound basis for application of machine-leaning methods to identify important features of nanoporous materials that may achieve both high selectivity and separation capacity.

6.2 Methods and Models

6.2.1 Molecular Model

We consider D_2/H_2 separation with various MOFs at 77 K, a cryogenic temperature commonly used for characterization of porous materials by nitrogen adsorption. The latest computation-ready, experimentally synthesizable MOF database (CoRE 2019) is used for high-throughput screening and data generation[28]. Different from those in hypothetical MOF databases, all MOF structures in the CoRE database have been experimentally synthesized, thus paving the way for experimental verification of theoretical predictions.

At the cryogenic condition, hydrogen molecules exhibit non-negligible quantum effects that cannot be captured with classical methods. In this work, we use the Lennard-Jones (LJ) model to describe the classical component of intermolecular interactions. The quantum effects are accounted for by modifying the pair potential with the 4^{th}-order approximation to the Feynman-Hibbs (FH) equation[27]

$$U(r)=U^{LJ}(r)+\frac{\beta\hbar^2}{24\mu}\left[U^{LJ}\,'(r)+\frac{2U^{LJ}\,''(r)}{r}\right]+\frac{\beta^2\hbar^4}{1152\mu^2}\left[U^{LJ}\,''''(r)+\frac{4U^{LJ}\,'''(r)}{r}+\frac{15U^{LJ}\,'(r)}{r^3}\right]$$

(6.1)

where r denotes the center-to-center distance, U^{LJ} represents the 12-6 LJ potential, $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the absolute temperature, $\hbar$ is the reduced Planck constant, $\mu = m_i m_j / (m_i + m_j)$ is the reduced mass for the interacting particles i and j with mass m_i and m_j, respectively. In eq (6.1), the number of primes represents the order of derivatives of the LJ potential with respect to the distance. The LJ parameters, $\sigma = 0.296$ nm and $\epsilon / k_B = 34.2$ K, are the same for H_2 and D_2[29]; the isotopologues are distinguished only by the quantum corrections related to their molecular weights. The universal force field (UFF) is used to represent all MOF atoms and a cutoff distance of 12.9 Å is applied to the LJ potential[30]. The Lorentz-Berthelot mixing rule is employed for describing interaction between different atoms.

Figure 6.1 illustrates the pair potential between hydrogen isotopologues according to the quantum-corrected LJ models (Supporting Information). While the 4th order approximation of the FH equation is computationally more expensive than the quadratic approximation commonly used in molecular simulations, addition of the higher order terms is important to fully capture the quantum effect in particular for hydrogen gases in a confined space[27]. As temperature falls, we may see more discrepancy among the intermolecular potentials for different hydrogen isotopologues because the quantum effect becomes more pronounced. Consistent with a previous report by Beenakker et

al[11], a heavier hydrogen isotopologue has a smaller size but a higher attractive energy than a lighter isotopologue. The size difference can be attributed to the fact that a heavier isotope has a narrower translational wave function thus a smaller de Broglie wavelength. On the other hand, the energy difference arises from the disparity in the zero-point energies of H_2 and D_2 molecules.

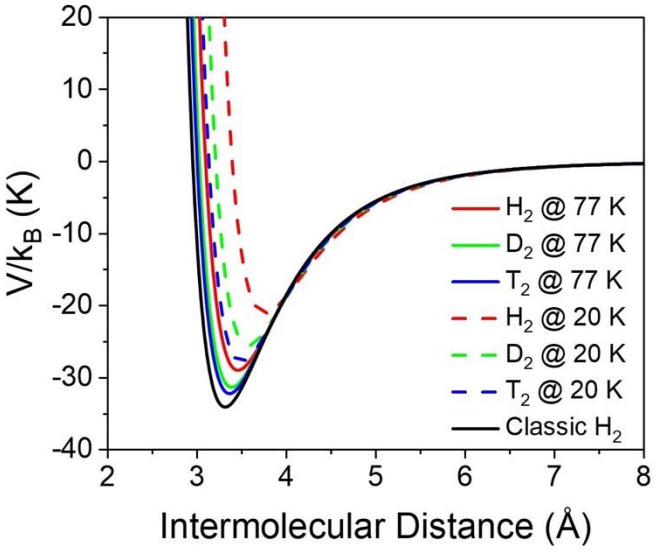

Figure 6.1. Classic and quantum-corrected intermolecular potential for H_2, D_2 and T_2 at 77 K and 20 K.

6.2.2 Ideal Adsorption Selectivity

We evaluate the separation efficiency for D_2/H_2 adsorption in MOF materials based on the ideal adsorption solution theory (IAST)[31]. The adsorption selectivity is thus defined as the ratio of Henry's constant for the heavier isotopologue relative to that for the lighter isotopologue:

$$\alpha^{IM} = K_{h,2} / K_{h,1} \tag{6.2}$$

where superscript "IM" stands for ideal systems, i.e., gas adsorption at extremely low

pressure such that interactions between gas molecules are negligible. For gas adsorption, Henry's constant is calculated from

$$K_h = \frac{1}{k_B TV} \int \exp\left[-\beta \varphi^{ext}(\mathbf{r})\right] d\mathbf{r} \tag{6.3}$$

where **r** stands for the position of a gas molecule, V is the system volume, and $\varphi^{ext}(\mathbf{r})$ represents the external potential for a gas molecule at **r** due to its interaction with the MOF atoms. With the quantum effects accounted for by the FH equation, we have different adsorbate-adsorbent interaction energies for H_2 and D_2 thereby different Henry's constants.

6.2.3 Ideal Membrane Selectivity

MOFs can be used for D_2/H_2 separation either as an adsorbent or as a porous membrane. In Henry's law region (*viz*., at low pressure), the membrane selectivity is defined in terms of Henry's constant and the self-diffusivity coefficient of gas molecules at infinite dilution[17, 32, 33]:

$$k^{IM} = \frac{K_{h,2}}{K_{h,1}} \frac{D_{0,2}}{D_{0,1}} = \frac{P_2}{P_1} \tag{6.4}$$

where permeability P is defined as the product of Henry's constant, K_h, and the self-diffusivity coefficient at infinite dilution, D_0[33]. Unlike adsorption selectivity, the membrane selectivity depends on both the thermodynamic and the transport properties of individual gas compounds.

As reported in an earlier work[34], the diffusion coefficient at infinite dilution can be predicted by using the transition-state theory (TST):

$$D_0 = \frac{1}{2}\varpi a^2 \tag{6.5}$$

where *a* stands for the distance between the equilibrium positions of the gas molecule in two neighboring cages (the initial and final states of transmission), and ν is the gas hopping rate. The latter can be calculated from the potential energy along the diffusion path[35]

$$\varpi = \sqrt{\frac{k_B T}{2\pi m}} \frac{\exp\left[-\beta\varphi^{ext}\left(s^*\right)\right]}{\int_0^1 \exp\left[-\beta\varphi^{ext}\left(s\right)\right] ds} \tag{6.6}$$

where *m* denotes the molecular mass, the integral is performed along the reaction coordinate of gas hopping, and superscript * represents the transition state of a gas molecule hoping between neighboring cages. Similar to our previous work[26], the minimum energy path can be calculated by using a simplified string method. Although the above procedure is based on the classical theory, it has been shown that the quantum-corrected potential is able to capture not only the equilibrium quantum sieving effect but also the kinetic quantum sieving consistent with experimental observations[36, 37].

It should be noted that, in addition to quantum sieving, quantum tunneling may also contribute to gas transport at the cryogenic temperature. In principle, we can take into account the quantum tunneling effect by adding a tunneling correction factor *Q* to the right side of eq (6.5),

$$Q = \frac{e^{\alpha}}{\gamma - \alpha}\left(\gamma e^{-\alpha} - \alpha e^{-\gamma}\right) \tag{6.7}$$

where $\alpha = E / k_B T$, $\gamma = 2a\pi^2 \left(2mE\right)^{1/2} / h$, E is the energy barrier for gas diffusion, *2a*

stands for the hopping distance, and *h* represents the Planck constant[38]. It has been shown that quantum tunneling is significant for H_2/D_2 separation only when the energy barrier is much larger than the difference between the zero-point energies[38]. Because such condition is rarely met for D_2 and H_2 in MOF materials, we neglect the correction factor affiliated with the quantum tunneling effect for all calculations reported in this work.

6.2.4 Robeson Boundary

The Robeson boundary provides an empirical correlation between permeability and the separation factor (the ratio of permeability). Although there are no empirical parameters available for D_2/H_2, empirical parameters of the Robeson boundary for H_2/N_2 are used in this work since they have the closest molecular parameters as hydrogen isotopes. These parameters are all obtained from Robeson's original work[39].

The Robeson upper bound can be expressed as following:

$$P_i = k\alpha_{ij}^n \tag{6.8}$$

where P_i is the permeability of spice i, k is an empirical parameter related to the solubility and membrane properties, n is an empirical parameter related with the molecular size, and α_{ij} is the permeability ratio of species i and j (a.k.a. separation factor). In eq (6.8), n can be calculated in the following way:

$$n = \frac{-1.127916}{d_j - d_i} \tag{6.9}$$

where d_i is the molecular size of spice i. The numerator on the right-hand side of eq (6.9) is fitted from Robeson's original work.

In this work, we assume that the molecular size is the same as the LJ diameter. The value of parameter k is predicted from the Freeman equation:

$$k^{-\frac{1}{n}} = \frac{S_i}{S_j} S_i^{-\frac{1}{n}} \exp\left\{\frac{1}{n}\left[b - f\left(\frac{1-a}{RT}\right)\right]\right\} \tag{6.10}$$

where S is the solubility constant, a has a universal value of 0.64, b is 11.5 for glassy polymers membrane used in this work and f has a value of 12,600 cal/mol. The solubility constant can be expressed:

$$\ln S_i = b + 0.023(\varepsilon_i / k_B) \tag{6.11}$$

where ε is the LJ energy parameter, and b=14.4557 is obtained from Robeson's original work.

6.2.5 Machine Learning Models

In addition to physics-based models described above, we use data-based models (a.k.a. machine learning methods) to identify the structural features of MOF membranes with good separation efficiency. Specifically, we have tested the performance of four types of machine learning models that may be able to reveal the desirable features of nanoporous materials for H_2/D_2 separation: *support-vector machine* (SVM), *random forest* (RF), *gradient-boosted trees* (GBT) and *deep neural network* (DNN). SVM is one of the most prevailing machine learning models for classification; it projects the original data into a higher-dimensional space with the help of hyperplane construction[34, 35]. In comparison with alternative models, SVM is expected to achieve a better distinction of different MOFs within the original data because the kernel method is able to well recognize data patterns. RF is a machine-learning model consisting of a large amount of

decision trees (“if-then” logic consequence) to extract useful information from the input data[40, 41]. The final result given by RF is the average vote of all decision trees in the model. In general, RF is able to avoid overfitting better than a single decision tree because the vote of biased decision trees would be averaged in the end. Similar to RF, GBT employs an ensemble of decision trees (weak learners) to achieve a strong learning power. Different from RF, decision trees in GBT are shallow trees with high bias and low variance instead of fully grown decision trees. Finally, DNN is well known for its excellent performance in predicting new features with a large number of input parameters. It is based on the idea of *artificial neural network* (ANN) that mimics the function of a human brain[42-44]. While ANN uses only a layer of neurons (transfer functions), DNN incorporates multiple layers of neurons thereby having a better interpretation power.

For all machine learning models considered in this work, we optimize the parameters with the *k*-fold cross-validation method. The procedure has a single parameter (*k*) that is referred to the number of groups that a given data sample is to be split into. In this work, we take a *k*-fold value of 17, the first prime number that is divisible to the entire dataset.

6.3 Results and Discussion

6.3.1 Adsorption Versus Permeation

Nanoporous materials can be used as an effective medium for gas separation based on either adsorption or permeation. The latter often achieves a higher selectivity because permeation in a membrane is related to both thermodynamic and transport properties. However, computational studies of MOFs for gas separations are mostly

focused on the adsorption behavior because calculation of transport properties renders additional challenges. In a previous work[34], we proposed an efficient theoretical procedure to calculate the diffusivity by using a simplified string method to calculate the minimum-energy path. The theoretical procedure allows us to calculate adsorption isotherms and diffusion coefficients with negligible computational cost in comparison with conventional methods.

Figure 6.2 shows the selectivity for a large library of MOF materials (CoRE 2019) versus their capacity for D_2/H_2 separation by adsorption and by membrane permeation at 77 K. For separation of the hydrogen isotopologues by gas adsorption, the selectivity quickly declines as Henry's constant increases, implying that MOFs with high D_2/H_2 selectivity are compromised by a low separation capacity. Approximately, the ideal adsorption selectivity falls exponentially with Henry's constant, and the latter provides a direct measure of adsorption capacity. Conversely, membrane separation exhibits no negative correlation between the selectivity and capacity. As shown in Figure 6.2B, most MOFs are able to achieve an ideal membrane selectivity significantly higher than that suggested by the Robeson boundary (details are provided in Supporting Information). The high selectivity values indicate that a MOF membrane would have a much better performance than a conventional polymer membrane for separating D_2/H_2. Figure 6.2B also shows that most MOFs are able to achieve k^{IM} much higher than the classical limit of the Knudsen diffusion (~ 0.7), albeit a few MOFs have k^{IM} around 0.7 because of the lack of sufficiently small pores for the kinetic quantum sieving effect to take place.

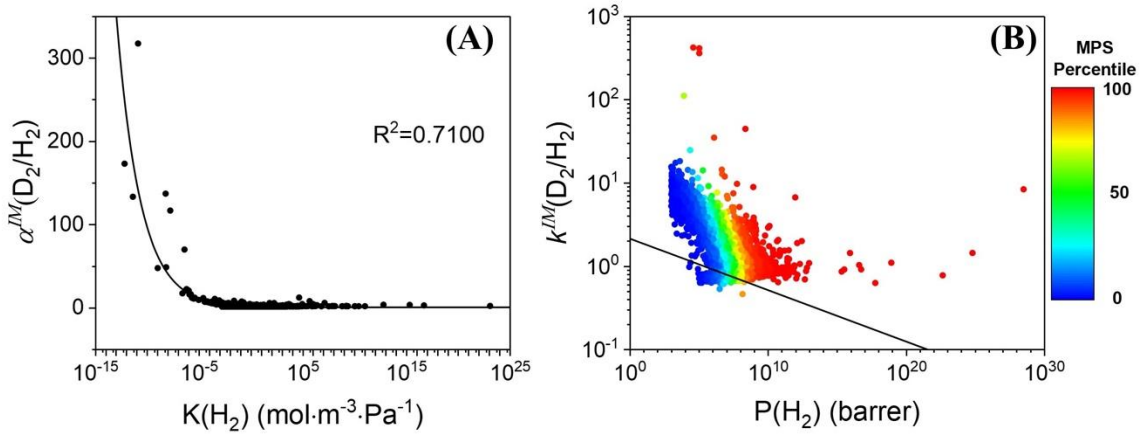

Figure 6.2. (A) Ideal selectivity for D_2/H_2 separation at 77 K by MOF adsorption versus Henry's constant for H_2. The points are calculated from eq (6.2) for 12,723 MOFs from the CoRE 2019 library, and the solid line is empirically fitted with $\alpha^{IM} = 139.7e^{3694000K} + 1.33$. (B) Ideal membrane selectivity verse permeability of H_2 at 77 K. The points are predicted from eq (6.4) for the same CoRE MOF materials, and the solid line represents the Robeson boundary. The color in (B) denotes the percentile of Membrane Performance Score (MPS): the red and blue colors represent the highest and lowest MPS, respectively, and the green color represents the intermediate MPS.

Table 6.1 lists the properties of top 5 MOFs from the CoRE 2019 library with the highest membrane selectivity k^{IM}. In Supporting Information (Table 6.2), the selectivity and the structural features of top 3 MOFs identified according to k^{IM} are compared with other MOFs that have been reported in the literature for D_2/H_2 separation. The selectivity of MOFs identified in this work is higher than those previously reported nanoporous materials by almost one order of magnitude. As discussed later, the top performing MOFs significantly enhance the kinetic quantum sieving (KQS) effect owing to their small pore limit diameters.

Table 6.1. Properties of top MOFs for D_2/H_2 separation at 77 K identified according to their ideal membrane selectivity k^{IM} ranking.

MOF	P(D_2) (barrer)	D_0(D_2) ($m^2 \cdot s^{-1}$)	K_h(D_2) ($mol \cdot m^{-3} \cdot Pa$)	k^{IM} (D_2/H_2)
ROQFUA07	1.689×10^7	5.47×10^{-18}	1.03×10^9	421.5
ROQNES05	4.506×10^7	6.75×10^{-18}	2.23×10^9	413.6
ROQFUA08	3.943×10^7	1.35×10^{-17}	9.77×10^8	360.6
ZOJWAY	9.232×10^5	2.54×10^{-15}	1.22×10^5	110.7
ECIVUH	1.057×10^{10}	8.93×10^{-12}	3.96×10^5	44.4

Table 6.2. Structural features of MOFs with the highest selectivity identified in this work and those reported by others.

MOF	PLD (Å)	LCD (Å)	Void Fraction	Selectivity	
ROQFUA07	2.587	4.936	0.4694	421.5 (77 K)	This Work
ROQNES05	2.607	4.937	0.4758	413.6 (77 K)	This Work
ROQFUA08	2.600	4.954	0.4712	360.6 (77 K)	This Work
SIFSIX-3-Zn	3.787	4.208	0.4230	53.8 (20 K)	Literature[45]
MIL-53	2.6-8.5			13.6 (77 K)	Literature[22]

6.3.2 Structural Features of Highly Selective MOF Membranes

To understand molecular mechanisms underpinning the efficiency of D_2/H_2 selectivity, we have calculated the geometric features of the MOFs with top 5% ideal membrane selectivity, i.e., the largest cavity diameter (LCD), the pore limit diameter (PLD), and the void fraction. All geometry calculations are based on Zeo++ software with UFF[46]. As MOF membranes are able to achieve better performance than MOF adsorbents for D_2/H_2 separation in terms of both selectivity and capacity, the top ranked MOF adsorbents are not considered in our structural analysis.

Figure 6.3 shows the distributions of the pore limit diameter (A), the largest cavity diameter along the diffusion path (B), and the void fraction (C) of the top ranked MOF membranes. Distributions of the void volume and surface area of these materials

are presented in Figure 6.4. In Figure 6.3A, we see a significant enhancement on the PLD distribution between 2 to 3 Å, suggesting that small pores are a desirable feature for D_2/H_2 separation. Intuitively, MOFs with PLD about 2-3 Å would better sieve D_2/H_2 at 77 K because the effective LJ diameter for hydrogen molecules is around 3 Å (see Table 6.3). As over 90% MOFs with top 5% ideal membrane selectivity have a PLD between 2 to 4 Å, it is important to replace the background distribution from the CoRE database for MOFs with PLD of 2-4 Å to avoid misinterpretation of how the structural features correlate with the membrane performance. Figure 6.5 shows the distribution of unit cell volumes of the MOF membranes with top 5% ideal membrane selectivity compared with the same distribution but for all MOFs from the CoRE 2019 library and for MOFs with PLD in the range of 2-4 Å. While modifying the unit cell volume could not improve the performance for D_2/H_2 separation, one might erroneously conclude that the top MOF membranes would correlate with the unit cell volume if its distribution for the top 5% MOFs is directly compared with that for all MOFs in the CoRE 2019 library. Figure 6.5B shows an enhanced distribution of the unit cell volume at 0-2000 $Å^3$ simply because MOFs with PLD of 2-4 Å have a relatively small unit cell volume (0-2000 $Å^3$).

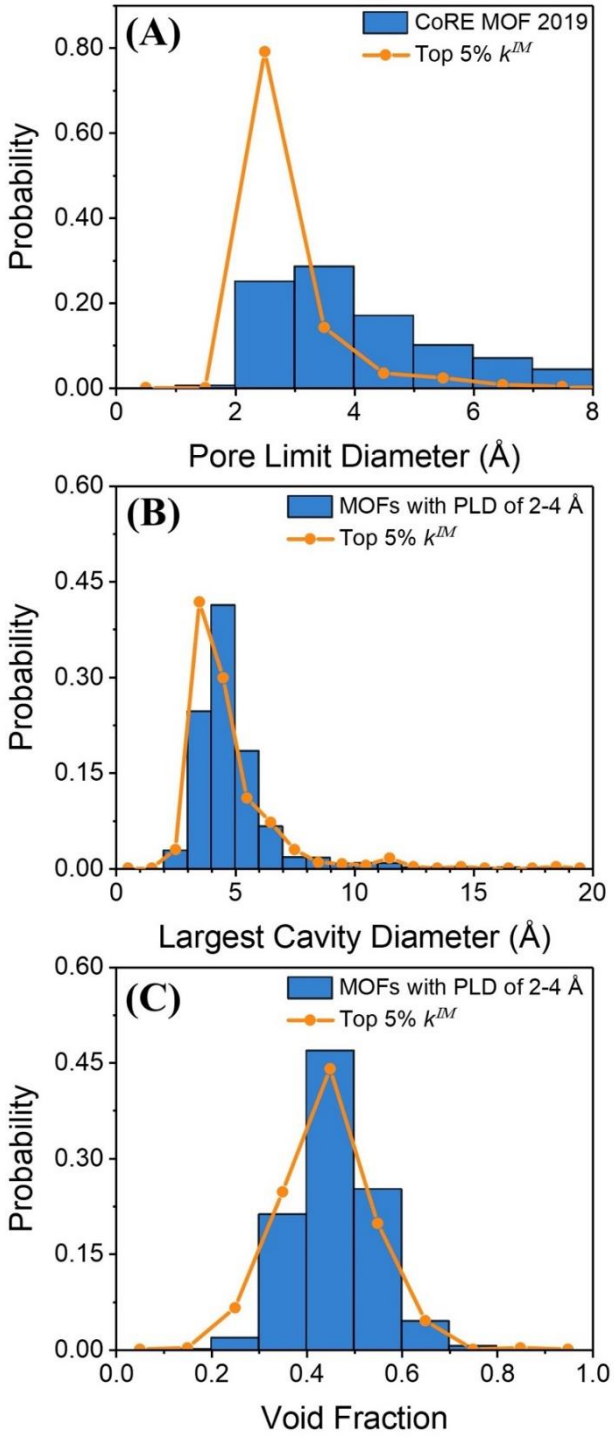

Figure 6.3. Distributions of the pore limit diameter (A), the largest cavity diameter along the diffusion path (B), and the void fraction (C) for MOFs in the CoRE 2019 library with the top 5% ideal membrane selectivity for D_2/H_2 separation at 77 K.

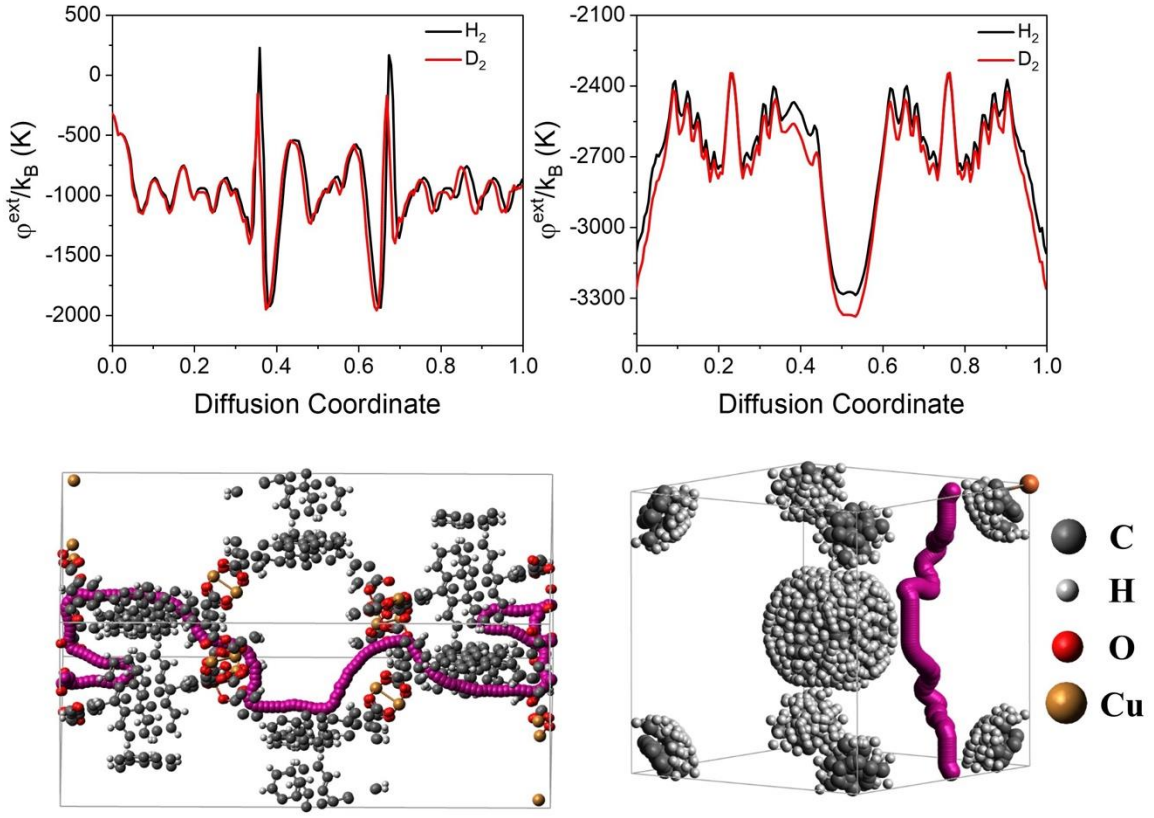

Figure 6.4. Distribution of void volume (A) and surface area (B) for MOFs with top 5% ideal membrane selectivity for D_2/H_2 separation.

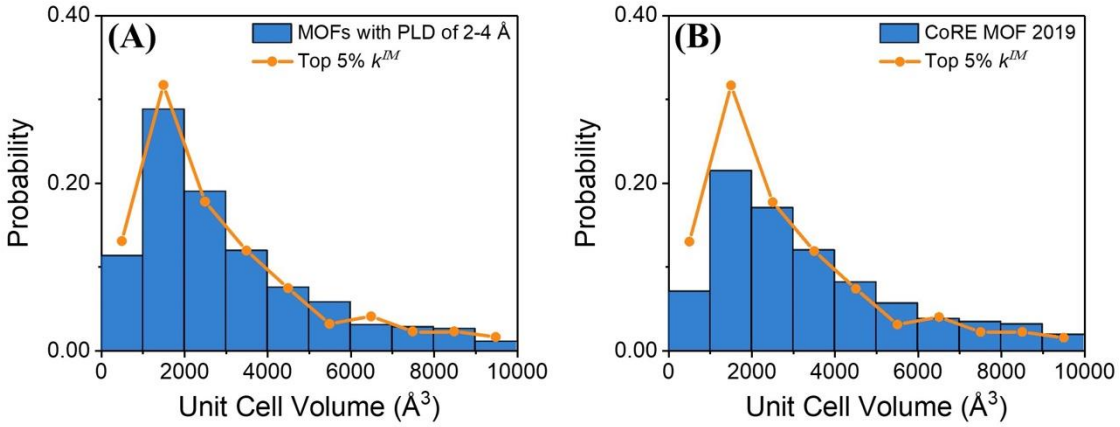

Figure 6.5. Distribution of unit cell volume for MOFs with top 5% ideal membrane selectivity and MOFs in the CoRE 2019 database (A) and MOFs with PLD between 2 and 4 Å (B).

Table 6.3. Effective LJ parameters of D_2 and H_2 at 77 K and 20 K

Molecule	77 K		20 K	
	σ (Å)	ε/k_B (K)	σ (Å)	ε/k_B (K)
H_2	3.1037	29.0823	3.4019	21.1377
D_2	3.0340	31.4524	3.2175	25.6396

Figure 6.3 and Figure 6.5 show that the LCD and void fraction distributions for the top ranked MOF membranes are noticeably different from the background distributions. While PLD plays a major role in determining the diffusion barrier, LCD affects gas permeation along the rest of diffusion coordinate. The smaller the LCD, the less favorable would be the minimum energy path for molecular hopping, leading to a smaller diffusion coefficient and gas permeability. Over 70% of MOFs with high ideal membrane selectivity have LCD in the range of 3-5 Å because the external potential for hydrogen molecules is strongly attractive within this range.

In addition to PLD and LCD, the void fraction is an important parameter to characterize the non-occupied space inside the MOF materials. It has been shown that a small change in the void fraction would affect the permeability of hydrogen molecules by orders of magnitude[47, 48]. However, permeability does not monotonically increase with the void fraction but depends also on the lattice type (e.g., fcc, bcc or simple cubic). Although a smaller void fraction would lead to a higher ratio of Henry's constants, it may also result in a smaller adsorption capacity. On the other hand, a larger void fraction may lead to larger pores but with diminishing quantum sieving effects for D_2 and H_2 separation. Figure 6.3C shows that nearly half of MOFs with top 5% ideal membrane selectivity have an intermediate void fraction around 0.45.

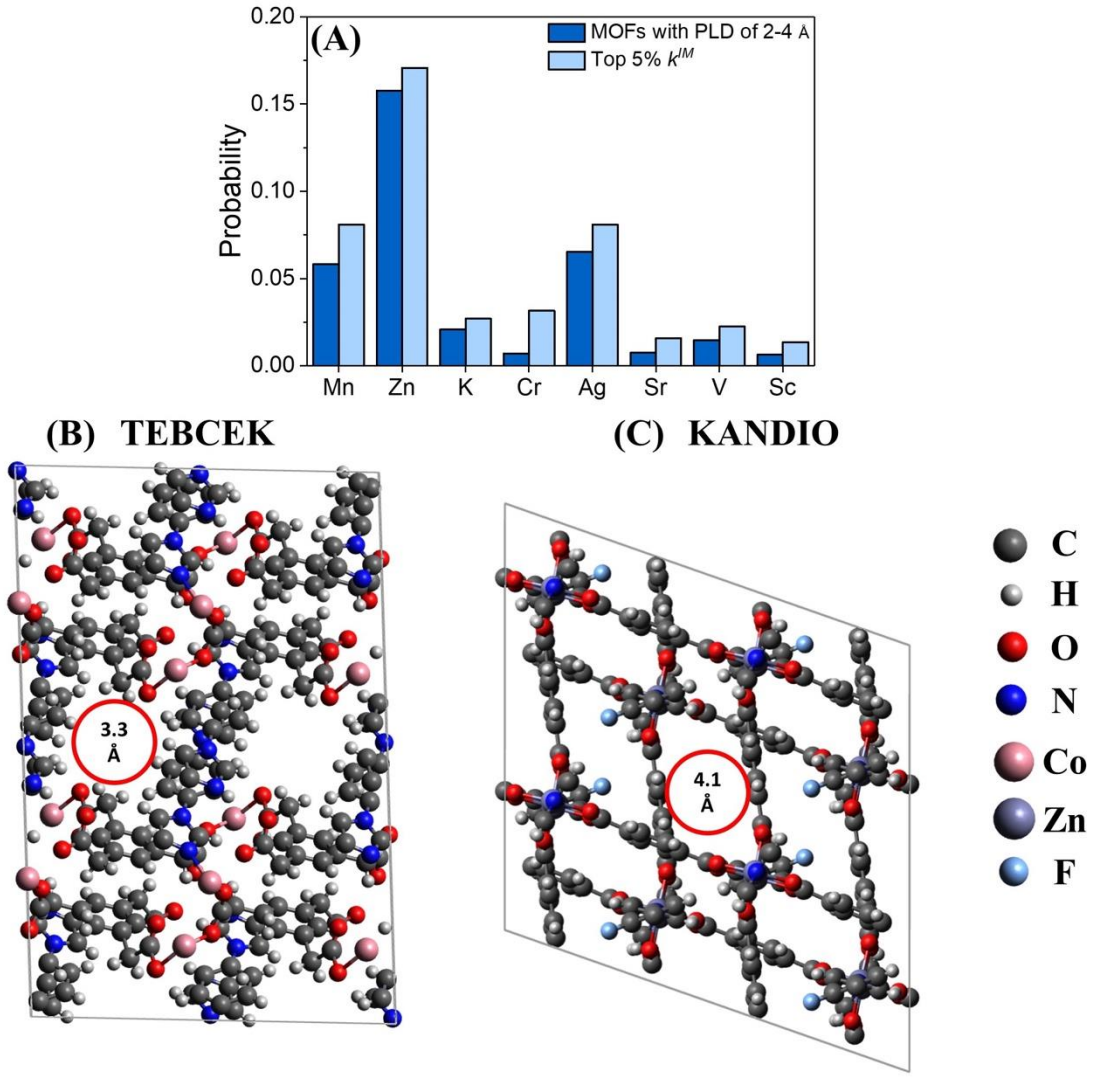

Figure 6.6. (A) Distributions of metal elements in the top 5% MOFs in terms of the ideal membrane selectivity and those with PLD in the range of 2-4 Å. All these elements exhibit a probability difference by at least 0.5% between the two groups of MOF structures. Example structures of MOFs containing metal elements with high (B) and low (C) energy parameters. The red circles denote the pore limit diameters.

In addition to the structural features, we have investigated the types of metal elements in the top ranked MOFs promising as a membrane for D_2/H_2 separation and their role in determining the separation performance. Figure 6.6 shows the percentages of

a few major metal elements in MOFs with top 5% ideal membrane selectivity and in those MOFs with PLD in the range of 2-4 Å. The results for all other metal elements are listed in Table 6.2. Most metal elements have only minor percentage difference between MOFs with top 5% ideal membrane selectivity and MOFs with PLD in the range of 2-4 Å.

Table 6.4. The LJ parameters for major metal elements in MOFs from the CoRE 2019 library with top 5% ideal membrane selectivity (elements with bolded fonts are considered as metal sites with high energy parameter)

Element	ε/k_B (K)	σ (Å)	Element	ε/k_B (K)	σ (Å)
Mn	6.54	2.64	Ag	18.11	2.80
Zn	**62.40**	**2.46**	**Sr**	**118.26**	**3.24**
K	17.61	3.40	V	8.05	2.80
Cr	7.55	2.69	Sc	9.56	2.94

The metal elements shown in Figure 6.6A have a probability in MOFs with top 5% ideal membrane selectivity at least 0.5% higher than that in MOFs with PLD in the range of 2-4 Å. Table 6.4 presents the LJ parameters for all these elements. Surprisingly, only a few metal elements have the LJ energy parameters higher than those for hydrogen molecules. As the Lorentz-Berthelot mixing rule is used to predict attraction between different species, metal elements with higher energy parameters would be most favored for D_2/H_2 separation. In other words, a larger energy parameter for the adsorbent is preferred because the small difference between the chemical species to be separated would be magnified. As each MOF is consisted of metal nodes and organic linkers, a.k.a. secondary building units (SBUs), quantum sieving can still be achieved for metal sites with lower energy parameter by pairing it with SBUs of appropriate choice. For metal sites with a lower energy parameter, a smaller SBU would then be needed to make confinement small enough to distinguish H_2/D_2 and enable effective quantum sieving.

Figures 3B and 3C present representative structures of MOF membranes containing metal sites with high and low energy parameters, respectively. TEBCEK contains cobalt atoms ($\epsilon / k_B = 7.05$ K) while KANDIO contains zinc atoms ($\epsilon / k_B = 62.40$ K). These two structures have similar ideal membrane selectivity but KANDIO has larger pore sizes (LCD = 6.0 Å and PLD = 4.1 Å) than TEBCEK (LCD = 3.3 Å and PLD = 2.6 Å). In order to have equivalent membrane selectivity (KANDIO vs. TEBCEK, 5.7 vs. 4.6), a smaller pore size is required for MOFs containing metal sites with a lower binding energy such that their impacts on gas diffusivity and adsorption amount can be compensated with each other.

6.3.3 Membrane Performance Score

The adsorbent performance score has been widely used to evaluate porous materials for gas separation because it takes into account both selectivity and separation capacity. However, a similar metric was missing for membrane separations. Here we define the *Membrane Performance Score* (MPS) according to the membrane selectivity and permeability:

$$MPS = S_{fast/slow} \times P_{fast} \tag{6.12}$$

where $S_{fast/slow}$ is the membrane selectivity of the fast diffusing species over the slow diffusing species, P_{fast} is the permeability of the fast specie. As shown in Figure 6.2B, a high MPS can be achieved by either a high selectivity or a high permeability. For separation of H_2 and D_2, most MOFs with high MPS can be attributed to high permeability but intermediate selectivity.

Table 6.5. Top 5 MOFs for D_2/H_2 separation at 77 K according to the membrane performance score

MOF	$P(D_2)$ (barrer)	$D^0(D_2)$ ($m^2 \cdot s^{-1}$)	$k^{IM}(D_2/H_2)$	MPS (barrer)
RUBLEH	2.94×10^{29}	3.97×10^{-10}	8.3	2.44×10^{30}
ERANAO	9.96×10^{24}	2.40×10^{-8}	1.42	1.42×10^{25}
YEGKIG	3.71×10^{22}	1.78×10^{-9}	0.77	6.30×10^{22}
XOPVOO	9.93×10^{18}	2.11×10^{-8}	1.09	1.08×10^{19}
FEKBED	3.94×10^{17}	5.88×10^{-12}	0.62	1.02×10^{18}

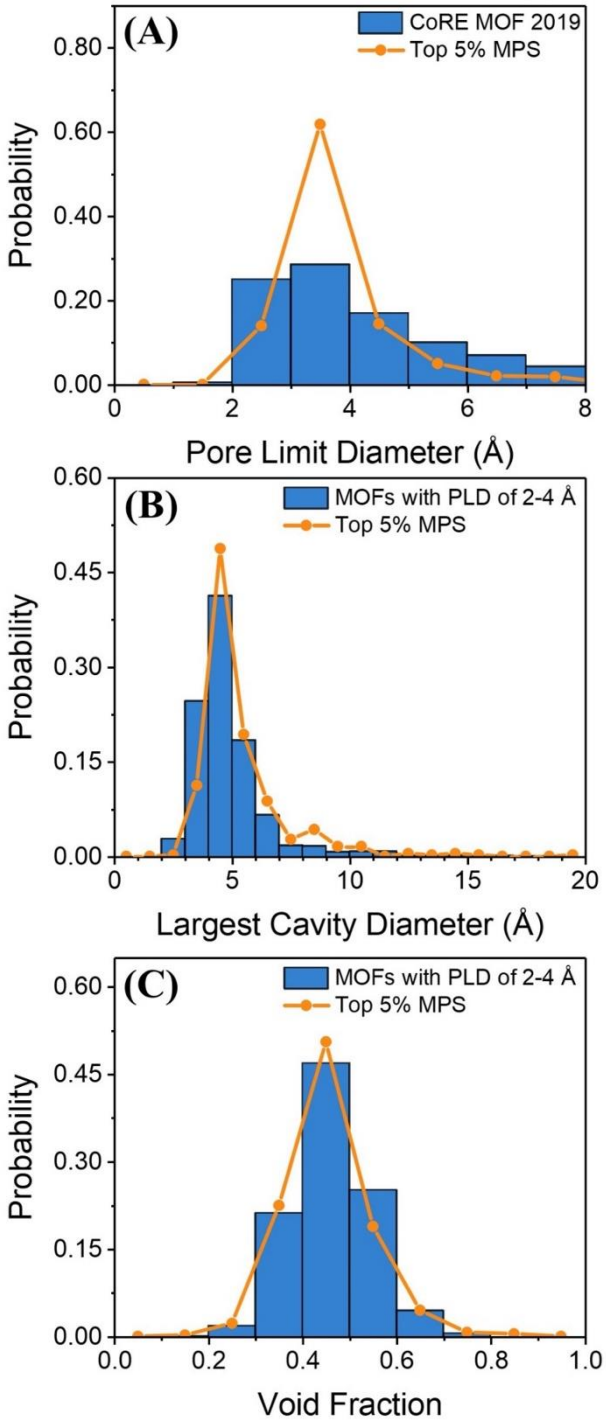

Figure 6.7. Distributions of the pore limit diameter (A), the largest cavity diameter along the diffusion path (B), and the void fraction (C) for MOFs with top 5% MPS for D_2/H_2 separation.

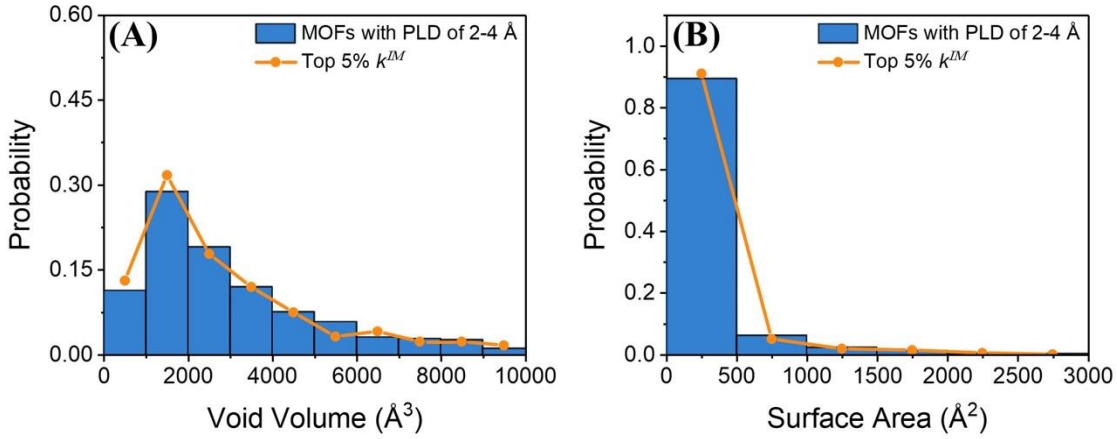

Figure 6.8. Distribution of void volume (A) and surface area (B) for MOFs with top 5% ideal membrane selectivity.

Table 6.5 lists the theoretical values of permeability, ideal diffusivity, and membrane selectivity for top 5 MOFs with the highest MPS values. Figure 6.7 and Figure 6.8 provide an analysis of the structural features of the top 5% MOFs according to the MPS ranking. Compared with MOFs with top 5% ideal membrane selectivity, MOFs with top 5% MPS have enhanced distributions for both PLD and LCD at a larger pore diameter, and the void fraction distribution shifts to the direction of less confinement. For D_2/H_2 separation, it is much easier to have a high permeability rather than a high membrane selectivity. Therefore, MOFs with top 5% MPS have less confined structures in order to achieve high permeability. Indeed, all MOFs listed in Table 6.5 have an extremely high permeability but with only moderate membrane selectivity.

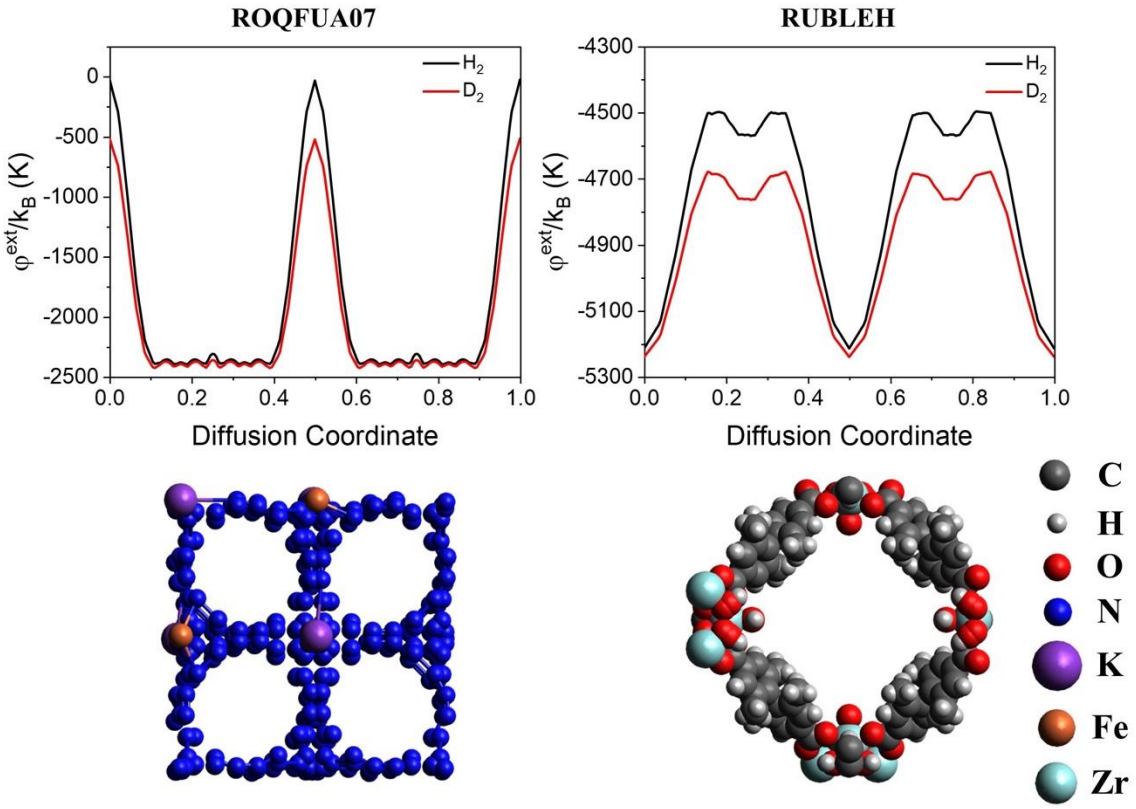

Figure 6.9. The energy landscape along the diffusion coordinate and the structures of MOFs with the highest ideal membrane selectivity (left panel - ROQFUA07) and those with the highest membrane performance score (right panel - RUBLEH).

Figure 6.9 shows the energy landscapes and the structures of MOFs with the highest ideal membrane selectivity (ROQFUA07) and with the highest membrane performance score (RUBLEH). We may identify significant differences in the energy landscapes for ROQFUA07 and RUBLEH. As discussed above, the highest ideal membrane selectivity is affiliated with an energy barrier that distinguishes hydrogen isotopologues through quantum sieving. However, the highest membrane performance score is attributed to high permeability and intermediate membrane selectivity. Not only is the difference between the energy barriers for H_2 and D_2 diffusion in ROQFUA07

much higher than those in RUBLEH, but the energy landscape along the minimum energy path differs greatly between these two MOF materials as well. RUBLEH is strongly attractive to both hydrogen isotopologues along the minimum energy path, leading to a high permeability and a high membrane performance score. By contrast, ROQFUA07 imposes little attraction at the diffusion barrier, implying that the process is dominated by repulsive interactions. The difference in the energy landscape results in the permeability of hydrogen isotopologues in RUBLEH much higher than that in ROQFUA07. The structural features of MOF with the highest ideal membrane selectivity (ROQFUA07) are similar to those recommended by Nguyen et al.[36] In order to achieve good D_2/H_2 separation, PLD (2.587 Å) should be significantly smaller than s_{H_2} while LCD (4.936 Å) should be slightly larger than s_{H_2}.

6.3.4 Promising Features of MOF Membrane Predicted by Machine Learning

While physics-based modelling allows us to analyze the materials behavior based on atomistic details, machine learning provides a complementary route to discovering new materials that satisfy predefined specific practical needs. A first step toward the so-called inverse design is to identify patterns or promising features from a large amount of data relating materials performance to the atomic constituents.

As discussed above, structural features, such as PLD, LCD and void fraction, are intrinsically related to the performance of MOF membranes for D_2/H_2 separation. In principle, these properties may be directly used as input parameters (*viz*., descriptors) for machine learning. Although the general trend can be captured by regression analysis, we found that these structural features are oversimplified and insufficient to establish

quantitative correlations between the MOF structures and the physical properties underlying their performance for D_2/H_2 separation (shown in Figure 6.10). Even with additional descriptors reflecting the metal elements, a quantitative correlation is still beyond the capacity of regression analysis. Therefore, in the following, we use classification instead of regression methods for data-based modelling.

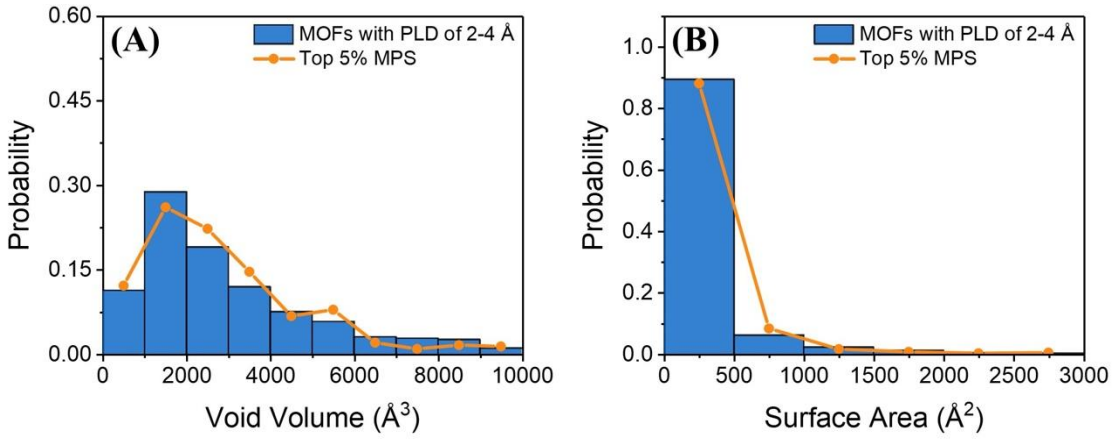

Figure 6.10. Ideal membrane selectivity predicted by (A) DNN, (B) RF, (C) GBT and (D) SVM versus calculated by physics-based model. Left panel is predicted by input of PLD, LCD and void fraction while right panel is predicted by additional percentage of metal elements in MOFs.

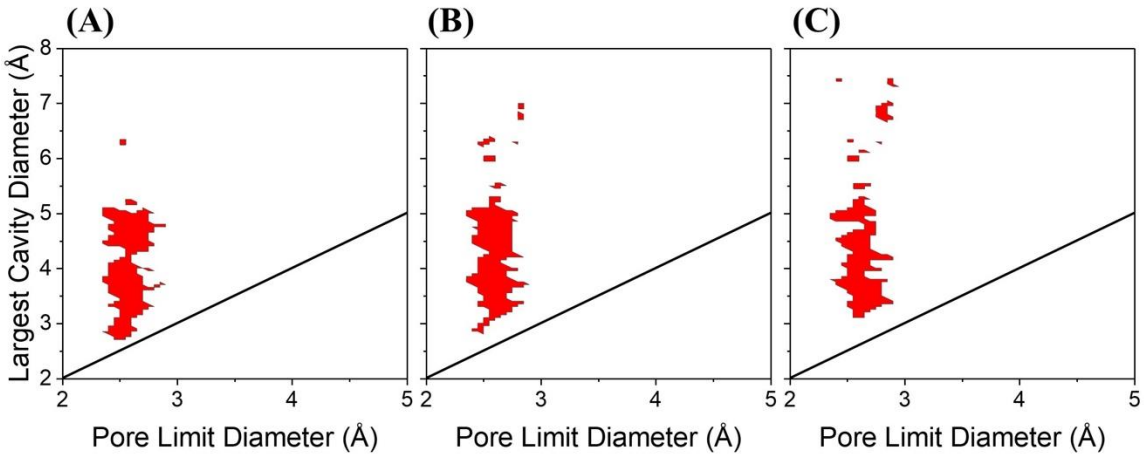

Figure 6.11. Structural features predicted by machine learning models for MOFs with top 10% ideal membrane selectivity (red color). Here the void fraction is fixed at (A) 0.4, (B) 0.45 and (C) 0.5. The black lines show conditions where PLD equals to LCD. Because PLD must be larger than LCD, the area below the black line is physically impossible.

We have analyzed the structural features of MOFs using *support-vector machine* (SVM), *random forest* (RF), *gradient-boosted trees* (GBT), and *deep neural network* (DNN). Table 6.6 compares the accuracy of different machine learning models. All four classification models are able to recognize MOFs with top 10% ideal membrane selectivity. With additional information on metal elements, DNN shows most significant improvement in accuracy. On the other hand, RF, GBT and SVM methods provide satisfactory classification with only structural features, e.g., in terms of the PLD, LCD and void fraction of each MOF structure. With the kernel method projecting data to a higher dimensional space, SVM is able to distinguish small differences between MOFs better than RF and GBT, which employ many decision trees of the same dimension. Surprisingly, RF, GBT and SVM methods show no improvement of the classification accuracy with the extra information for metal elements as the input. Because high-quality classification can be readily achieved with SVM, RF or GBT using the structural features

as the input, additional information on metal elements might be considered merely as a noise. Such information is too discrete and scarce and becomes detrimental to recognize the data patterns with SVM, RF or GBT[49].

Table 6.6. Percentage of metal elements in different categories of MOFs

Element	2Å< PLD <4Å (%)	Top 5% k^{IM} (%)	Element	2Å< PLD <4Å (%)	Top 5% k^{IM} (%)	Element	2Å< PLD <4Å (%)	Top 5% k^{IM} (%)
Cu	12.51	8.54	K	2.07	2.70	Ho	0.68	0.67
Cd	11.35	10.79	Yb	0.81	0.90	Bi	0.31	0.00
W	1.35	1.12	I	0.91	0.45	Nb	0.25	0.00
Na	1.81	1.12	Cr	0.71	3.15	Y	0.66	0.90
Co	10.51	7.19	Ni	4.34	3.82	Ba	0.93	0.22
Mn	5.82	8.09	Cs	0.35	0.00	Pt	0.34	0.00
Tb	2.40	2.70	Si	1.03	0.45	Lu	0.21	0.22
U	1.32	0.45	Ag	6.53	8.09	Hf	0.01	0.00
La	2.20	2.02	Ca	1.37	1.12	Ir	0.04	0.00
In	1.34	1.80	Fe	4.51	4.04	Sb	0.04	0.00
Ce	1.32	1.12	Ga	1.28	1.35	Au	0.31	0.00
Pr	1.51	0.67	Dy	1.62	0.90	Sc	0.63	1.35
Nd	2.22	1.57	Pd	0.09	0.22	Rb	0.29	0.22
Sm	1.66	1.80	Sr	0.75	1.57	Sn	0.13	0.45
Eu	3.25	1.57	B	0.90	1.12	Re	0.26	0.00
Gb	2.15	1.80	Mg	1.41	0.90	Hg	0.21	0.00
Mo	1.44	0.90	V	1.47	2.25	Rh	0.07	0.00
Be	0.19	0.67	Al	1.90	2.02	As	0.07	0.22
Zn	15.77	17.08	Ru	0.60	0.45	Se	0.16	0.00
Br	0.88	0.00	Ti	0.35	0.22	Ge	0.10	0.00
Er	1.47	1.57	Th	0.16	0.45	Te	0.07	0.22
Pb	0.19	0.00	Zr	1.06	1.35	Np	0.09	0.00
Tm	0.32	0.67	Li	0.78	0.90	Pu	0.01	0.00

Among the four classification methods, SVM has the highest accuracy and it is thus used to predict structural features for MOFs with the best membrane selectivity. The classification accuracy, which is defined as the percentage of MOFs correctly predicted by the four machine learning methods for the entire database, is presented in Supporting Information (Table 6.7). Figure 6.11 shows the predicted results. Here the red color denotes PLD and LCD values predicted by SVM for MOFs with top 10% ideal membrane selectivity at three different void fractions. The optimum PLD is approximately the same as the LJ diameter of the hydrogen molecules. In other words, MOFs with PLD comparable to the molecular size would yield a large difference in the energy barriers in H_2 or D_2 diffusion thus promoting separation. According to SVM, LCD should be slightly larger than the molecular size of isotopic hydrogen in order to utilize the kinetic quantum sieving effect for D_2/H_2 separation, and the range of LCD shifts to larger diameters when the void fraction increases. While MOFs with PLD comparable to the molecular size of hydrogen would impose a nearly repulsive interaction, those with LCD slightly larger than the molecular size would be able to provide more attraction along the minimum energy path thereby increasing the molecular hopping rate. Because a larger LCD would accommodate a larger void fraction, the specific range of LCD depends on the void fraction of the MOFs to be designed to achieve optimal membrane performance.

Table 6.7. Accuracy of classification for different machine learning models

	DNN	RF	GBT	SVM
Accuracy without metal sites (%)	49.27	60.97	57.03	65.58
Accuracy with metal sites (%)	56.47	59.39	55.57	64.00

6.4 Conclusions

In this work, we used physics-based models to evaluate the performance of 12,723 MOFs from the CoRE 2019 database for D_2/H_2 separation. The selectivity and separation capacity were calculated for each MOF when it is used either as adsorbent or membrane material. We find that excellent D_2/H_2 selectivity can be achieved through gas adsorption but the MOF performance is compromised by low separation capacity. By contrast, MOFs can also be used as a membrane material for D_2/H_2 separation with a good balance of selectivity and capacity. Even at relatively high temperature (77 K), the D_2/H_2 selectivity for best MOF membranes identified in this work (Table 6.1) is almost one order of magnitude higher than those previously reported in the literature. Because all MOF structures in the CoRE database have been experimentally synthesized, our theoretical results are directly testable with experimental measurements.

An analysis of the structural features and the metal compositions of promising MOF membranes indicates that high D_2/H_2 selectivity can be achieved when the pore limit diameter (PLD) is comparable to the Lennard-Jones (LJ) diameter of hydrogen molecules. PLD plays an important role in determining the diffusion barrier, which is closely affiliated with the kinetic quantum sieving effects. For practical applications, we introduced the membrane performance score to evaluate the overall performance of the MOF membranes in terms of both selectivity and permeability. While MOFs with high membrane selectivity are characterized with extremely small pores, those with high membrane performance scores are less confined in order to achieve high permeability. In addition to PLD, gas permeation in MOFs depends on the largest cavity diameter (LCD)

and the metal composition. The smaller the LCD, the less favorable would be the minimum energy path for molecular hopping, leading to a smaller diffusion coefficient and gas permeability. For MOFs with different metal compositions, metal elements with a lower binding energy would require a smaller secondary building unit in order to achieve the same kinetic quantum sieving effect.

Combination of physics-based models for high-throughput screening and data-based modelling for identification of useful geometric features facilitates the inverse design of MOFs for better separation performance. With the extensive data generated from physics-based modelling, we have identified useful features of MOF materials by exploring four different machine learning models. Although none of these models gives a quantitative correlation of membrane selectivity or the membrane performance scores, satisfactory results can be obtained by using support-vector machine (SVM) to reproduce the structural features of promising MOF membranes. For effective D_2/H_2 separation, SVM predicts that the kinetic quantum sieving effects are most significant when the PLD of MOFs is comparable to, while LCD is slightly larger than, the molecular dimeter of isotopic hydrogens. The specific range of LCD depends on the void fraction of the MOF materials. Those structural features predicted by integrating physics-based modeling with machine learning provide useful insights into the rational design of new MOF structures for more efficient D_2/H_2 separation.

Bibliography

1. Zaccai, G., *How soft is a protein? A protein dynamics force constant measured by neutron scattering.* Science, 2000. **288**(5471): p. 1604-7.

2. Machida, A., et al., *Site occupancy of interstitial deuterium atoms in face-centred cubic iron.* Nat Commun, 2014. **5**: p. 5063.

3. Stiopkin, I.V., et al., *Hydrogen bonding at the water surface revealed by isotopic dilution spectroscopy.* Nature, 2011. **474**(7350): p. 192-5.

4. Keppler, F., et al., *Methoxyl groups of plant pectin as a precursor of atmospheric methane: evidence from deuterium labelling studies.* New Phytol, 2008. **178**(4): p. 808-14.

5. Povinec, P.P., et al., *Isotope tracing of submarine groundwater discharge offshore Ubatuba, Brazil: results of the IAEA-UNESCO SGD project.* J Environ Radioact, 2008. **99**(10): p. 1596-610.

6. Souers, P.C., *Hydrogen properties for fusion energy*. 1986, Berkeley: University of California Press.

7. Sanderson, K., *Big interest in heavy drugs.* Nature, 2009. **458**(7236): p. 269.

8. Rae, H.K., *Selecting Heavy Water Processes*, in *Separation of Hydrogen Isotopes*. 1978. p. 1-26.

9. Bokhan, P.A., et al., *Laser Isotope Separation in Atomic Vapor*. 2006.

10. Mazur, T.R., B. Klappauf, and M.G. Raizen, *Demonstration of magnetically activated and guided isotope separation.* Nature Physics, 2014. **10**(8): p. 601-605.

11. Beenakker, J.J.M., V.D. Borman, and S.Y. Krylov, *Molecular-Transport in Subnanometer Pores - Zero-Point Energy, Reduced Dimensionality and Quantum Sieving.* Chemical Physics Letters, 1995. **232**(4): p. 379-382.

12. FitzGerald, S.A., et al., *Highly selective quantum sieving of D2 from H2 by a metal-organic framework as determined by gas manometry and infrared spectroscopy.* J Am Chem Soc, 2013. **135**(25): p. 9458-64.

13. Kim, J.Y., et al., *Exploiting Diffusion Barrier and Chemical Affinity of Metal-Organic Frameworks for Efficient Hydrogen Isotope Separation.* J Am Chem Soc, 2017. **139**(42): p. 15135-15141.

14. Li, J.R., R.J. Kuppler, and H.C. Zhou, *Selective gas adsorption and separation in metal-organic frameworks.* Chem Soc Rev, 2009. **38**(5): p. 1477-504.

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

16. Wilmer, C.E., et al., *Large-scale screening of hypothetical metal–organic frameworks.* Nature Chemistry, 2011. **4**: p. 83.

17. Keskin, S., et al., *Atomically detailed models of gas mixture diffusion through CuBTC membranes.* Microporous and Mesoporous Materials, 2009. **125**(1-2): p. 101-106.

18. Kim, J.Y., H. Oh, and H.R. Moon, *Hydrogen Isotope Separation in Confined Nanospaces: Carbons, Zeolites, Metal-Organic Frameworks, and Covalent Organic Frameworks.* Adv Mater, 2019. **31**(20): p. e1805293.

19. Garberoglio, G., *Quantum sieving in organic frameworks.* Chemical Physics Letters, 2009. **467**(4-6): p. 270-275.

20. Oh, H., et al., *Quantum cryo-sieving for hydrogen isotope separation in microporous frameworks: an experimental study on the correlation between effective quantum sieving and pore size.* Journal of Materials Chemistry A, 2013. **1**(10): p. 3244-3248.

21. Cao, D., et al., *Ultrahigh effective H2/D2 separation in an ultramicroporous metal–organic framework material through quantum sieving.* Journal of Materials Chemistry A, 2018. **6**(41): p. 19954-19959.

22. Kim, J.Y., et al., *Selective Hydrogen Isotope Separation via Breathing Transition in MIL-53(Al).* J Am Chem Soc, 2017. **139**(49): p. 17743-17746.

23. Challa, S.R., D.S. Sholl, and J.K. Johnson, *Light isotope separation in carbon nanotubes through quantum molecular sieving.* Physical Review B, 2001. **63**(24).

24. Garberoglio, G. and J.K. Johnson, *Hydrogen isotope separation in carbon nanotubes: calculation of coupled rotational and translational States at high densities.* ACS Nano, 2010. **4**(3): p. 1703-15.

25. Tian, Y., W.Y. Fei, and J.Z. Wu, *Separation of Carbon Isotopes in Methane with Nanoporous Materials.* Industrial & Engineering Chemistry Research, 2018. **57**(14): p. 5151-5160.

26. Zhou, M.S., et al., *Fractionation of Isotopic Methanes with Metal-Organic Frameworks.* Journal of Physical Chemistry C, 2019. **123**(12): p. 7397-7407.

27. Kumar, A.V. and S.K. Bhatia, *Quantum effect induced reverse kinetic molecular sieving in microporous materials.* Phys Rev Lett, 2005. **95**(24): p. 245901.

28. Chung, Y.G., et al., *Computation-Ready Experimental Metal-Organic Framework (CoRE MOF) 2019 Dataset*. 2019: Zenodo.

29. Buch, V., *Path integral simulations of mixedpara-D2andortho-D2clusters: The orientational effects.* J Chem Phys, 1994. **100**(10): p. 7610-7629.

30. Rappe, A.K., et al., *Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations.* J Am Chem Soc, 1992. **114**(25): p. 10024-10035.

31. Myers, A.L. and J.M. Prausnitz, *Thermodynamics of mixed-gas adsorption.* AIChE J, 1965. **11**(1): p. 121-127.

32. Keskin, S. and D.S. Sholl, *Efficient methods for screening of metal organic framework membranes for gas separations using atomically detailed models.* Langmuir, 2009. **25**(19): p. 11786-11795.

33. Haldoupis, E., S. Nair, and D.S. Sholl, *Efficient calculation of diffusion limitations in metal organic framework materials: a tool for identifying materials for kinetic separations.* J Am Chem Soc, 2010. **132**(21): p. 7528-39.

34. Tian, Y., X. Xu, and J. Wu, *Thermodynamic Route to Efficient Prediction of Gas Diffusivity in Nanoporous Materials.* Langmuir, 2017. **33**(42): p. 11797-11803.

35. Frenkel, D. and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications*. 2nd ed. 2002, San Diego: Academic Press. 638.

36. Nguyen, T.X., H. Jobic, and S.K. Bhatia, *Microscopic observation of kinetic molecular sieving of hydrogen isotopes in a nanoporous material.* Phys Rev Lett, 2010. **105**(8): p. 085901.

37. Murphy, M.J., G.A. Voth, and A.L.R. Bug, *Classical and quantum transition state theory for the diffusion of helium in silica sodalite.* Journal of Physical Chemistry B, 1997. **101**(4): p. 491-503.

38. Anslyn, E.V., *Modern physical organic chemistry*. 2006, Sausalito CA: University Science/Sausalito CA.

39. Robeson, L.M., *The upper bound revisited.* Journal of Membrance Science, 2008. **320**(1-2): p. 390-400.

40. Tin Kam, H., *The random subspace method for constructing decision forests.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(8): p. 832-844.

41. Rokach, L. and O. Maimon, *Data Mining with Decision Trees*. Series in Machine Perception and Artificial Intelligence. 2014.

42. Goh, G.B., N.O. Hodas, and A. Vishnu, *Deep learning for computational chemistry.* J Comput Chem, 2017. **38**(16): p. 1291-1307.

43. Waszczyszyn, Z. *Fundamentals of Artificial Neural Networks*. 1999. Vienna: Springer Vienna.

44. Towell, G.G. and J.W. Shavlik, *Knowledge-based artificial neural networks.* Artificial Intelligence, 1994. **70**(1-2): p. 119-165.

45. Han, G., et al., *Screening of Metal-Organic Frameworks for Highly Effective Hydrogen Isotope Separation by Quantum Sieving.* ACS Appl Mater Interfaces, 2018. **10**(38): p. 32128-32132.

46. Willems, T.F., et al., *Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials.* Microporous and Mesoporous Materials, 2012. **149**(1): p. 134-141.

47. Perez-Mas, L., et al., *Maximizing the absorption of small cosolutes inside neutral hydrogels: steric exclusion versus hydrophobic adhesion.* Phys Chem Chem Phys, 2018. **20**(4): p. 2814-2825.

48. Kim, W.K., et al., *Tuning the Permeability of Dense Membranes by Shaping Nanoscale Potentials.* Physical Review Letters, 2019. **122**(10): p. 10.1103/PhysRevLett.122.108001.

49. Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics. 2009.

Chapter 7. Inverse Design of Metal-Organic Frameworks for C_2H_4/C_2H_6 Separation

Efficient separation of C_2H_4/C_2H_6 mixtures is of paramount importance in the petrochemical industry. Nanoporous materials, especially metal-organic frameworks (MOFs), may serve the purpose owing to their tailorable microscopic structure and pore geometry. In this chapter, we propose a computational framework for high-throughput screening and inverse design of high-performance MOFs that can be utilized for adsorption and membrane separation processes. High-throughput screening of the computational-ready, experimental MOF database (CoRE MOF 2019) leads to materials with exceptionally high ethane-selective adsorption selectivity (LUDLAZ: 7.68) and ethene-selective membrane selectivity (EBINUA02: 2167.3). Moreover, the inverse design enables the exploration of a broader chemical space and identification of MOFs with even higher membrane selectivity and permeability. In addition, a relative membrane performance score (rMPS) has been formulated to evaluate the overall performance of MOF membranes relative to the Robeson boundary. The computational framework offers useful guidelines for the experimental design of MOFs and is generically appliable to materials discovery for gas storage and separation.

7.1 Introduction

The efficiency of C_2H_4/C_2H_6 separation is important for the petrochemical industry because high-purity C_2H_4 is used as the primary feedstock for the synthesis of diverse chemical products including plastics, polyesters and rubber materials.[1, 2] Conventional processes for C_2H_4/C_2H_6 separation are mostly based on high-pressure cryogenic distillation, which requires extensive energy input while suffers from low

separation efficiency. To reduce the energy cost and increase the selectivity, it is desirable to develop alternative approaches such as adsorption or permeation processes based on nanoporous materials.[3-5]

Metal-organic frameworks (MOFs) are ideal candidates for the efficient separation of C_2H_4/C_2H_6 because they have good mechanic stability, large specific surface area, and tailorable pore structure and geometry.[6-8] In particular, such materials show excellent performance for separating molecules with similar size and interaction energy, such as the mixtures of H_2/D_2 isotopes, of noble gases (Ar/Kr/Xe), and of xylene isomers.[9-12] For C_2H_4/C_2H_6 separation, promising MOF candidates have been identified by experiments.[3, 4, 13, 14] Whereas the possible variations of MOFs are virtually unlimited and the separation efficiency is sensitive to the atomic details, it is practically infeasible to explore the design space only through experiment. Previously, computational methods have been used to identify best material candidates for separation process through high-throughput screening.[9, 15, 16] While the adsorption isotherms predicted by the computational methods are found in good agreement with experimental measurement,[17] membrane processes are often considered more efficient to separate C_2H_4 from C_2H_6 by leveraging the difference in both adsorption affinity and gas diffusivity.[18] To the best of our knowledge, previous research on the computational screening of MOF database is mostly concerned with the separation of C_2H_4/C_2H_6 by gas adsorption.[17, 19-22] From the computational perspective, the assessment of MOF materials for membrane separation is much more demanding because the evaluation of gas diffusivity in confined geometry is typically much more time-consuming than that of

gas adsorption. In particular, the strong confinement makes it computationally prohibitive to predict the diffusion coefficients of gas molecules in a large library of nanoporous materials using conventional methods such as molecular simulation (MD) simulation.[23, 24]

Although computational methods (e.g., MD, grand canonical Monte Carlo simulation, and classical density function theory) have been well established for the accurate prediction of gas adsorption and diffusivity,[25-28] the inverse design of nanoporous materials for separation processes remains a theoretical challenge from both computational and practical perspectives. While generative adversarial network (GAN) shows early success in the inverse design of zeolites for methane storage,[29] its computational complexity increases significantly with the number of elements in the crystal structure. Besides, GAN easily breaks down and fails to converge for complicated crystalline materials such as MOFs because a large number of atomic types need to be considered. In contrast, variational autoencoder (VAE) can well accommodate the complex topology and molecular structure of the secondary building blocks (SBUs) by compressing the MOF structure into a text string and projecting it into the latent space.[30] However, VAE requires accurate projection (*viz*., encode and decoder) between the crystal structure and a latent space, and the VAE training would become infeasible when a vast number of SBUs are considered for the MOF design. Alternatively, evolutionary algorithms, such as the genetic algorithm, are promising for the inverse design of MOFs because they can accommodate not only a large number of SBUs for

MOF design, but also find the solution in a nonlinear space consisted of the material topology and SBUs.[31, 32]

In previous work,[25, 26] we developed a computational framework with GPU-acceleration that offers fast and accurate evaluation of sorption and diffusion properties of gas molecules in nanoporous materials. Empowered by the new computational capability, here we perform high-throughput screening of the computational-ready experimental (CoRE 2019) MOF database (over 10k MOFs) for the separation of C_2H_4/C_2H_6 with adsorption and membrane processes. Compared with the state-of-art materials from the literature,[4, 33] the best MOFs identified in this work have much higher separation selectivity. The highest ethane-selective adsorption selectivity in LUDLAZ is up to 7.68, and the highest ethene-selective membrane selectivity in EBINUA02 can reach 2167.3. Leveraging on the high-throughput capability, a genetic algorithm (GA) is incorporated into our computational workflow to achieve the inverse design of MOF membranes with both high membrane selectivity and permeability. The inverse design allows us to explore a broader chemical space in comparison with high-throughput screening and identify MOFs with even higher membrane selectivity and permeability. The structural analyses of top MOFs with excellent separation performance provide insights for the experimental design of MOFs for adsorption and membrane separation.

7.2 Methods and Models

7.2.1 Molecular Models

In this work, ethane (C_2H_6) and ethene (C_2H_4) molecules are modeled as two united-atom groups separated by a fixed bond length.[34] The detailed force field parameters can be found in Table 7.1. Metal-organic frameworks (MOFs) are considered to be rigid with the universal force field (UFF) for all nonbonded interactions.[35] The Lennard-Jones (LJ) 12-6 potential is truncated and shifted to zero at 12.9 Å, and the Lorentz-Berthelot mixing rule is used for the energy and size parameters between different kinds of atoms. The periodic boundary condition is applied to all cell axes. The unit cell is duplicated such that the length along each lattice axis is at least two times the cutoff distance. The structural properties, such as the pore limit diameter, the largest cavity diameter and the void fraction, are calculated with Zeo++.[36]

Table 7.1. The Lennard-Jones (LJ) parameters for ethane and ethene[34]

	σ (Å)	ε/k_B (K)	bond length (Å)
CH_3- (ethane)	3.760	108	1.54
CH_2- (ethene)	3.680	92.8	1.33

7.2.2 Adsorption Separation

Nanoporous materials have been widely used as adsorbent in industrial applications. In the low pressure region, the adsorption selectivity for two chemical species can be measured with the ratio of Henry's constants[37]

$$\alpha^{IM} = \frac{K_{h,2}}{K_{h,1}} \tag{7.1}$$

where $K_{h,i}$ represents the Henry's constant of component i. For a gas molecule with a

rigid conformation, the Henry's constant can be calculated via the integration of the external potential due to its interaction with the nanoporous material[38]

$$K_h = \frac{1}{2\pi RTV}\int d\omega \int d\mathbf{r} \exp\left[-\beta\phi^{ext}(\mathbf{r},\omega)\right] \tag{7.2}$$

where $\beta = 1/(k_B T)$, k_B stands for the Boltzmann constant, T is the absolute temperature, V represents the system volume, ϕ^{ext} is the external potential, i.e., the potential energy due to the interaction of a gas molecule with all atoms from the porous material, **r** represents the cartesian coordinates for the center of mass of the gas molecule, and ω stands for its Euler angles. For each MOF, the Henry's constants for C_2H_6 and C_2H_4 are numerically evaluated via midpoint rule with the step size of 1 Å and 45° for spatial and rotational variables, respectively.

7.2.3 Membrane Separation

According to the solution-diffusion model, the membrane permeability is defined as the Henry's constant multiplied by the gas diffusion coefficient at infinite dilution. The membrane selectivity can thus be calculated from[38]

$$k^{IM} = \frac{K_{h,2}}{K_{h,1}}\frac{D_{0,2}}{D_{0,1}} = \frac{P_2}{P_1} \tag{7.3}$$

where $D_{0,i}$ stands for the diffusion coefficient of component i at infinite dilution , and P_i represent the permeability. In evaluating the membrane selectivity, we use the average of diffusion coefficients along the three lattice vectors (*viz.*, *x*-, *y*-, *z*-axis in cartesian coordinate if the lattice vectors are mutually orthogonal)

$$D_0 = \frac{D_{0,a} + D_{0,b} + D_{0,c}}{3}. \tag{7.4}$$

Along each direction, the diffusivity can be calculated independently according to the transition-state theory (TST)

$$D_{0,\alpha} = \frac{1}{2} k a_\alpha^2 \tag{7.5}$$

where α stands for the direction of lattice vector, k represents the transmission rate (*viz.*, the hopping rate of the gas molecule), and a_α stands for the hopping distance between neighboring unit cells along the direction of lattice vector α. The hopping rate can be obtained from the minimum energy path (MEP) via the Bennett-Chandler formula[39, 40]

$$k = \sqrt{\frac{k_B T}{2\pi m}} \frac{\exp\left[-\beta \phi^{ext}\left(s^*\right)\right]}{\int_0^1 \exp\left[-\beta \phi^{ext}\left(s\right)\right] ds} \tag{7.6}$$

where s is a normalized dimensionless variable along the MEP, and s^* represents the transition state. The diffusion coefficient predicted by eq (7.5) depends heavily on the accuracy of the MEP. The details of accurate MEP calculation for rigid molecules, such as C_2H_4 and C_2H_6, can be found in our previous work.[25] Here it is suffice to state that the diffusion coefficient calculated from the transition-state theory agrees well with that calculated from MD simulation. Therefore, all diffusion coefficients reported in this work are predicted from the MEP obtained from the GPU-accelerated simplified string method.

7.2.4 Genetic Algorithm

For inverse design, we use the genetic algorithm (GA) to construct MOFs with desired properties. Because MOFs can be decomposed into the secondary building blocks (SBUs), each material may be considered as a “chromosome” of different topologies, metal nodes and organic linkers. Because most MOF topologies can only accommodate

less than two types of metal node and three types of organic linker, each chromosome consists of 6 genes and each gene is represented by an integer which corresponds to specific topology, node or linker. The population is set as 2,000 which enables the initial generation to have diverse combinations of topologies, nodes and linkers. Three evolutions are carried after the initial population and a total of 8,000 combinations of topologies, nodes and linkers is explored to find the optimal MOF structure with desired properties. Compared with the MOF structural database used in high-throughput screening, less MOF structures are used in the GA to benchmark its computational performance for the inverse design despite a much larger chemical space is considered. In the initial population, the 2,000 chromosomes are generated by the random selection of topologies, nodes and edges in the SBU database.[41] In each generation, 10 MOFs are used to generate the offspring via single point crossover. The next generation of MOFs are selected by stochastic universal selection to avoid bias towards the SBUs with low fitness values while 30% MOFs would have a random mutation on their genes. In this work, PyGAD library is used for the genetic algorithm.[42] PORMAKE is used to construct MOFs when the chromosome values are assigned.[41]

7.3 Results and Discussion

7.3.1 Screening CoRE MOF 2019 Database

We first perform the high-throughput screening of the computational-ready experimental (CoRE) MOF 2019 database for the separation of C_2H_4/C_2H_6 via adsorption and membrane processes. While high-throughput screening has been commonly used to find the best material candidate for gas adsorption, the procedure is more challenging for

membrane separation because of the steep computational cost in evaluating the diffusion coefficients. Figure 7.1 shows the separation selectivity versus capacity for both adsorption and membrane separations. As shown in Figure 7.1A, the maximum selectivity is less than 30 for ethene-selective MOFs suitable for adsorption separation. It decreases exponentially with the increase of the separation capacity (viz. adsorption amount) because highly confined pores are needed in order to achieve high ethene-selectivity. Such materials offer little pore volume to achieve high adsorption capacity. Table 7.2 lists the properties of top 5 ethene-selective MOFs for the adsorption separation of C_2H_4/C_2H_6 at room temperature (300 K).

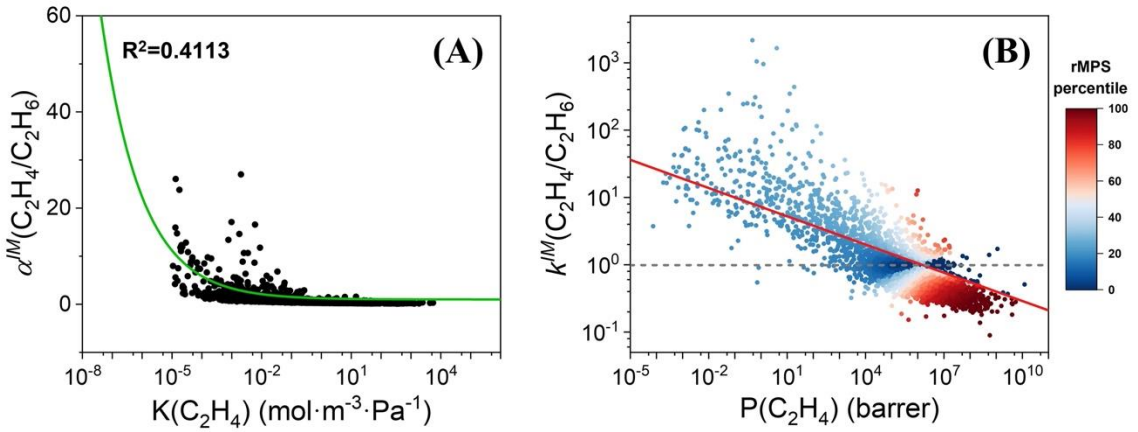

Figure 7.1. Selectivity vs. capacity for CoRE MOFs used in C_2H_4/C_2H_6 separation. (A) Adsorption selectivity. The green line is fitted with $\alpha^{IM}=0.2522e^{-0.7397\log(K)}+1$. (B) Membrane selectivity. The red line denotes the Robeson boundary, and the color stands for the percentile of relative membrane performance score (rMPS): the red, white and blue represent the highest, intermediate, and the lowest rMPS, respectively. The grey dashed line marks the membrane selectivity of 1.

Table 7.2. Henry's constants (K_h), ideal selectivity (α^{IM}) and self-diffusivity (D_0) of top ethene-selective MOFs for adsorption separation of C_2H_4/C_2H_6 at 300 K. The diffusion coefficients are shown only if they are larger than 1×10^{-20} $m^2\cdot s^{-1}$.

MOF	$K_h(C_2H_4)$ ($mol\cdot m^{-3}\cdot Pa^{-1}$)	$K_h(C_2H_6)$ ($mol\cdot m^{-3}\cdot Pa^{-1}$)	α^{IM} (C_2H_4/C_2H_6)	$D_0(C_2H_4)$ ($m^2\cdot s^{-1}$)	$D_0(C_2H_6)$ ($m^2\cdot s^{-1}$)
PIRYOF	5.5×10^{-2}	2.0×10^{-3}	27.01		
BADHIA	3.5×10^{-4}	1.3×10^{-5}	26.07		
BADHOG	4.2×10^{-4}	1.8×10^{-5}	23.81		
EBINUA02	1.7×10^{-2}	9.8×10^{-4}	17.11	9.3×10^{-15}	7.4×10^{-17}
FEDKAB	1.0×10^{-1}	6.0×10^{-3}	16.56		

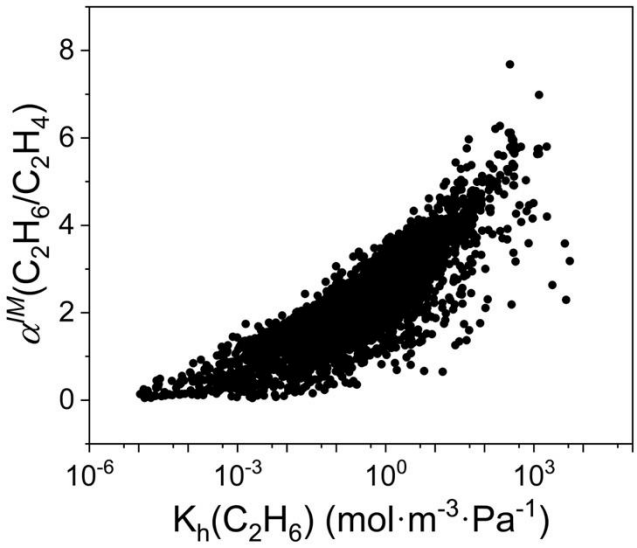

Figure 7.2. Separation selectivity vs. separation capacity for ethane-selective adsorption separation.

Although ethene-selective materials offer a much higher selectivity in adsorption separation of C_2H_4/C_2H_6, industrial applications desire ethane-selective processes because they can significantly reduce the energy cost. Figure 7.2 shows that, consistent with the literature,[4, 5, 20] the highest selectivity of ethane-selective MOFs identified in this work is much smaller than that of ethene-selective MOFs because the stronger

adsorption of smaller molecules (e.g., ethene) yields a large adsorption selectivity of C-$_2$H$_4$/C$_2$H$_6$ via the ultra-small pores of promising MOFs. It is worth noting that, different from ethene-selective MOF for separation of C$_2$H$_4$/C$_2$H$_6$ by adsorption, the selectivity of ethane-selective MOFs increases with the capacity and approaches a limiting value around 8.

Table 7.3. Top ethane-selective MOFs for the separation of C$_2$H$_4$/C$_2$H$_6$ at 300 K by adsorption.

MOF	$K_h(C_2H_4)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$K_h(C_2H_6)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	α^{IM} (C_2H_6/C_2H_4)	$D_0(C_2H_4)$ ($m^2 \cdot s^{-1}$)	$D_0(C_2H_6)$ ($m^2 \cdot s^{-1}$)
LUDLAZ	4.4×10^{1}	3.4×10^{2}	7.68	4.4×10^{-9}	6.3×10^{-9}
EFILUA	1.9×10^{2}	1.3×10^{3}	6.98	2.0×10^{-9}	1.2×10^{-9}
XUJSAY	3.4×10^{1}	2.1×10^{2}	6.27	6.3×10^{-9}	2.1×10^{-9}
ZAZNUL	2.7×10^{1}	1.7×10^{2}	6.20	6.6×10^{-9}	2.1×10^{-9}
KAXQIL	5.7×10^{1}	3.5×10^{2}	6.12	1.4×10^{-9}	1.1×10^{-9}

Table 7.3 lists the top 5 ethane-selective MOFs with the highest adsorption selectivity of C$_2$H$_6$/C$_2$H$_4$ at 300 K. Although excellent MOF candidates for the adsorption separation of C$_2$H$_4$/C$_2$H$_6$ have been reported before,[17, 19-22] the materials identified in this work yield much higher selectivity for both ethene-selective and ethane-selective separations. Previously, the computational screening was carried either on a smaller structural database or a subset of large structural library (e.g., CoRE MOF 2019) that is restricted by certain structural and chemical criteria. While these criteria would speed up the computation by reducing the number of materials to be evaluated, they would also ignore promising candidates due to the complex topology and structure. For the adsorption separation of C$_2$H$_4$/C$_2$H$_6$, the selectivity declines with the increase of loading amount. The reduction in adsorption selectivity can be attributed to the smaller difference

between the adsorbate-adsorbate interactions in comparison to that between adsorbate and adsorbent interactions.

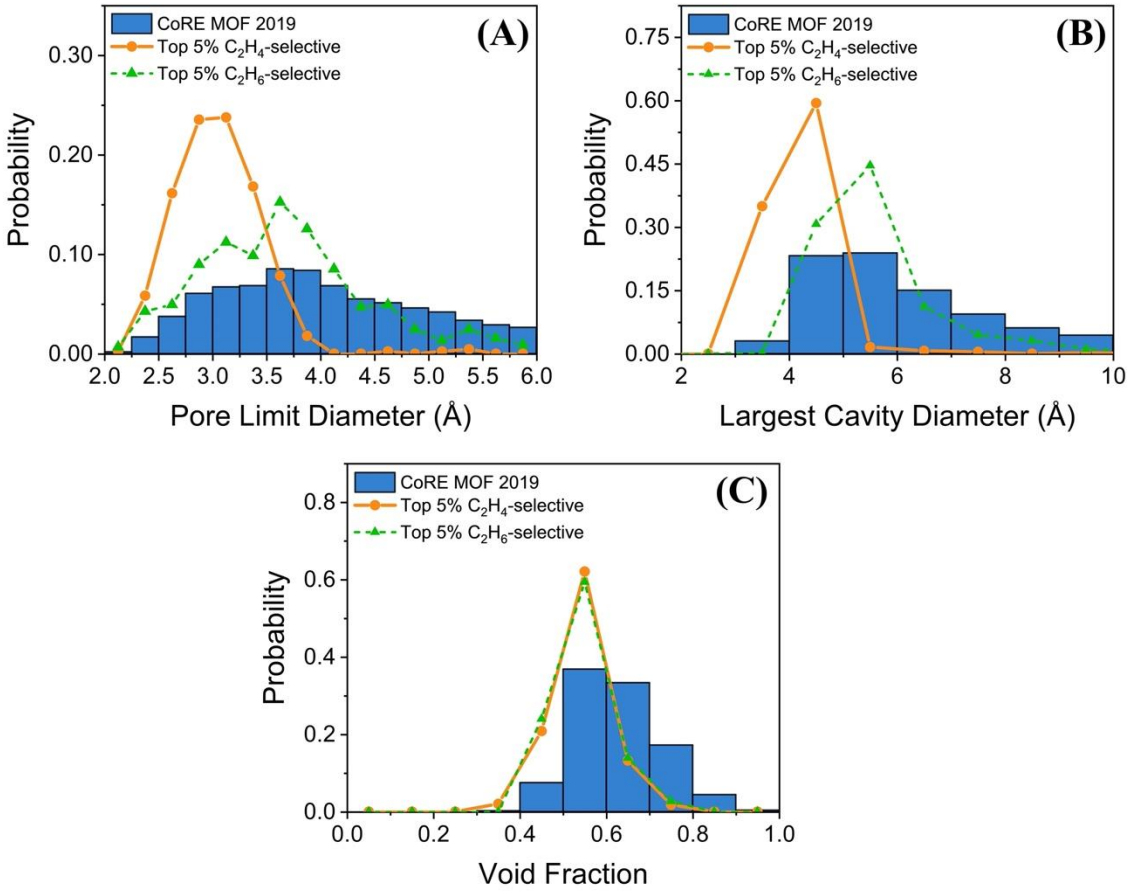

Figure 7.3. Distribution of the pore limit diameter (A), the largest cavity diameter (B) and the void fraction (C) for CoRE MOFs with top 5% ethane-selectivity and ethene-selectivity in the adsorption separation of C_2H_4/C_2H_6.

According to the structural analysis of the promising materials (shown in Figure 7.3), the ethane-selective MOFs have less confined geometry in terms of the pore limit diameter (PLD) and the largest cavity diameter (LCD) in comparison with the ethene-selective MOFs. The increase in pore size leads to a much higher adsorption capacity for those MOFs with higher ethane-selectivity. It is worth mentioning that the highest

adsorption selectivity of ethane-selective MOF [LUDLAZ: $\alpha^{IM}(C_2H_6/C_2H_4)$=7.68] identified in this work is 60% higher than the best nanoporous material found by the previous computational screening[22], a hypothetical zeolite structure [$\alpha^{IM}(C_2H_6/C_2H_4)$=4.86]. It is also more than 70% higher than the state-of-art nanoporous material reported in the experimental literature[4], $Fe_2(O_2)$(dobdc) with C_2H_6/C_2H_4 adsorption selectivity of 4.4. Besides, LUDLAZ has a much larger Henry's constant [$K_h(C_2H_6)$=4.6056 $cm^3 \cdot g^{-1} \cdot Pa^{-1}$] than $Fe_2(O_2)$(dobdc) [$K_h(C_2H_6)$=0.0147 $cm^3 \cdot g^{-1} \cdot Pa^{-1}$], meaning much higher gravimetric adsorption capacity. We note in passing that LUDLAZ was originally synthesized by McKellar et al. to examine how ligand exchange affects the stability and compressibility of the MOF materials.[43] For the top 5 ethane-selective MOFs, their pore size distributions are shown in Figure 7.4. These materials have similar micropores between 5 Å and 7 Å, rendering excellent selectivity of ethane over ethene in adsorption separation.

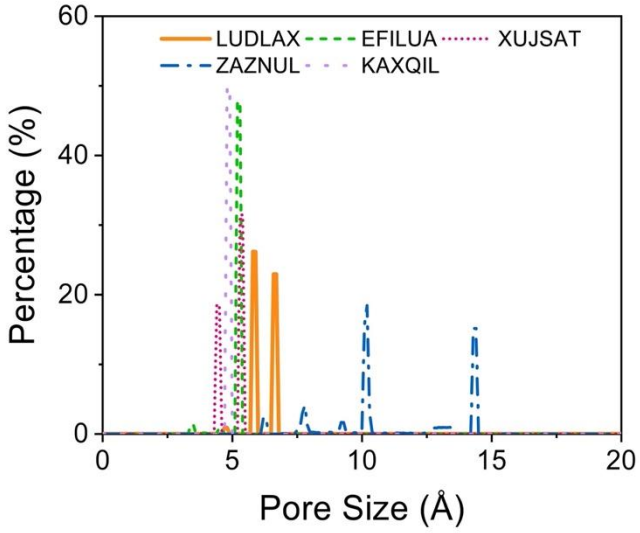

Figure 7.4. Pore size distribution for top 5 CoRE MOFs with highest ethane-selectivity in adsorption separation of C_2H_4/C_2H_6.

Compared with adsorption, gas separation *via* permeation through MOF membranes may achieve not only higher selectivity but also larger separation capacity. In addition, a membrane splits the feed stream into two purified sub-streams (*viz*., retentate stream and permeate stream) such that it does not require a recovery process even for ethene-selective operations. Therefore, the membrane process is often much less energy-intensive in comparison with adsorption. Figure 7.1B shows the membrane selectivity versus membrane permeability in the units of barrer. The red line in Figure 7.1B denotes the Robeson boundary, a semi-empirical upper limit summarized by Rungta et al.[2] based on the state-of-art polymer membranes for specifically separating ethane and ethene. We see that many MOFs in CoRE MOF 2019 database surpass the Robeson boundary, indicating their superior performance compared with the polymer membranes. The highest membrane selectivity of C_2H_4/ C_2H_6 is 2167.3 in MOF – EBINUA02, which

was synthesized by Tian et al.[44] with 1D rhombic channel in the 3D diamond topology network. The selectivity is several orders of magnitude larger than the state-of-art membrane materials discovered by previous computational screening and experimental synthesis.[19, 33] Table 7.4 lists the top 5 MOFs with the highest membrane selectivity of C_2H_4/C_2H_6.

Table 7.4. Top MOFs for C_2H_4/C_2H_6 separation with the highest membrane selectivity (k^{IM}) at 300 K.

MOF	$K_h(C_2H_4)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$K_h(C_2H_6)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$D_0(C_2H_4)$ ($m^2 \cdot s^{-1}$)	$D_0(C_2H_6)$ ($m^2 \cdot s^{-1}$)	k^{IM} (C_2H_4/C_2H_6)
EBINUA02	1.7×10^{-2}	9.8×10^{-4}	9.3×10^{-15}	7.4×10^{-17}	2167.3
HAZGOF	9.3×10^{-4}	2.0×10^{-4}	1.4×10^{-12}	4.2×10^{-15}	1649.1
ALOLES	2.8×10^{-2}	6.7×10^{-2}	8.3×10^{-15}	3.3×10^{-18}	1048.2
EBINUA	1.0×10^{-2}	7.7×10^{-4}	4.0×10^{-14}	5.6×10^{-16}	960.1
EBINUA01	2.1×10^{-2}	2.3×10^{-3}	3.0×10^{-13}	6.1×10^{-15}	438.2

In our previous work,[25, 26] we proposed the membrane performance score (MPS) to evaluate the overall performance of nanoporous materials by combining the membrane selectivity and permeability. Although MPS offers a direct comparison of nanoporous materials with different permeability and selectivity, it does not evaluate the membrane performance relative to the upper limit of the state-of-art polymer membranes (*viz*., the Robeson boundary). Here, we propose a modified version of MPS. The relative membrane performance score (rMPS) is defined as

$$\mathrm{rMPS} = \left(S_{fast/slow} - S_{fast/slow}^{Robeson}\right) \times P_{fast} \tag{7.7}$$

where S represents the membrane selectivity, P stands for the gas permeability, and the fast component refers to the one with higher permeability in the binary mixture. Since rMPS evaluates the overall performance of nanoporous materials relative to the Robeson

boundary, MOFs with a selectivity below the Robeson boundary would have a smaller rMPS compared with MPS. For the separation of C_2H_4/C_2H_6, Figure 7.1B shows that a high rMPS value favors MOFs with high permeability but intermediate selectivity (bottom right) instead of intermediate permeability and high selectivity (top center). Because ethane and ethene have similar molecular size and interaction energy, the increase of diffusion selectivity from an intermediate value requires a larger energy barrier along the minimum energy path (MEP), which leads to a significant reduction of the diffusion coefficient and rMPS. Conversely, the increase of diffusion coefficient results in the decrease of diffusion selectivity due to the absence of a large energy barrier for molecular sieving. As shown in Figure 7.2, only the ideal adsorption selectivity of C_2H_6/C_2H_4 increases with the capacity. As a result, MOFs with high rMPS are mostly ethane-selective, and that the membrane selectivity is mostly attributed to their difference in the adsorption amount. Table 7.5 lists the top 5 MOFs with the highest rMPS.

Table 7.5. Top MOFs with the highest relative membrane performance score (rMPS) for the separation of C_2H_4/C_2H_6 at 300 K.

MOF	$K_h(C_2H_4)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$K_h(C_2H_6)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$D_0(C_2H_4)$ ($m^2 \cdot s^{-1}$)	$D_0(C_2H_6)$ ($m^2 \cdot s^{-1}$)	k^{IM} (C_2H_6/C_2H_4)	rMPS (barrer) $\times 10^{10}$
LUDLAZ	4.4×10^{1}	3.4×10^{2}	4.4×10^{-9}	6.3×10^{-9}	11.1	5.6
PARMIG	9.7×10^{1}	5.6×10^{2}	1.4×10^{-8}	1.3×10^{-8}	5.5	5.4
BEKSAM	4.5×10^{2}	1.9×10^{3}	9.1×10^{-9}	4.2×10^{-9}	2.0	1.3
MIMVEJ	2.9×10^{1}	1.4×10^{2}	1.3×10^{-8}	1.3×10^{-8}	5.0	1.3
MORZID	1.8×10^{2}	8.2×10^{2}	7.6×10^{-9}	6.6×10^{-9}	3.8	1.0

7.3.2 Structural Features of Promising MOF Membranes

As discussed above, the selectivity of MOF membranes is less compromised and much higher at high separation capacity in comparison with MOF adsorbents. To explore

the synergetic effects between adsorption and diffusion, we have further examined the structural features of top MOFs with the highest membrane selectivity and rMPS.

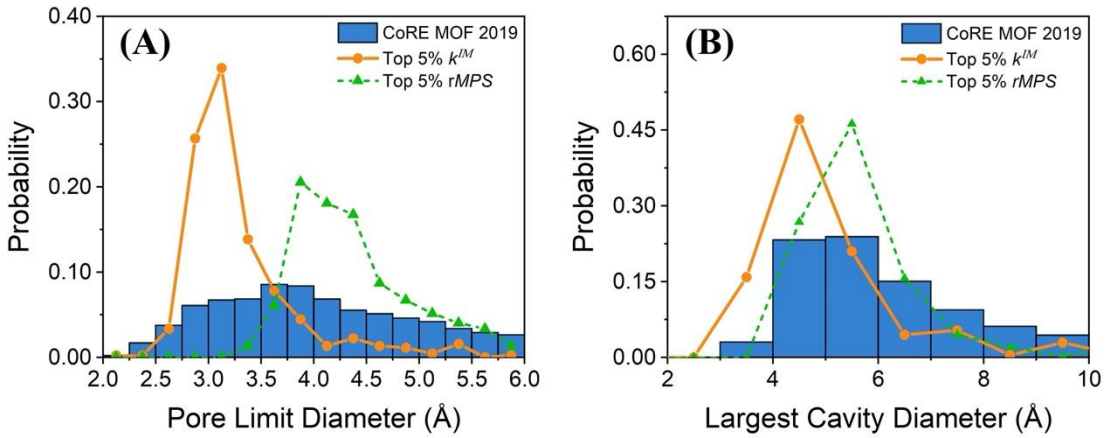

Figure 7.5. Distributions of the pore limit diameter (A) and the largest cavity diameter (B) for all CoRE MOFs and MOFs with top 5% ideal membrane selectivity and relative performance score (rMPS) for C_2H_4/C_2H_6 separation.

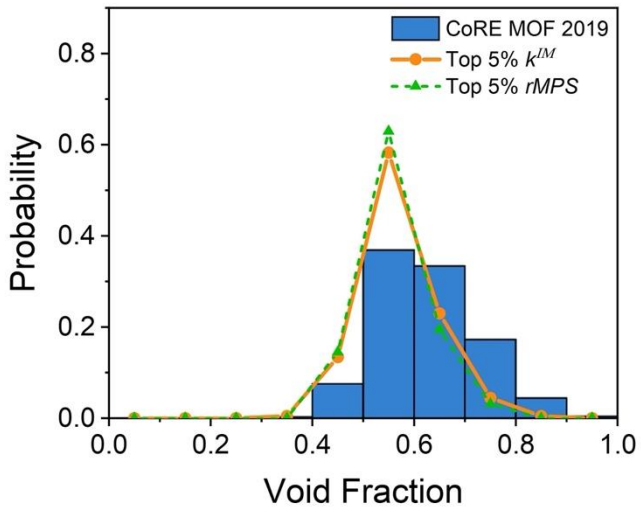

Figure 7.6. Distribution of the void fraction for all CoRE MOFs and MOFs with top 5% ideal membrane selectivity and relative performance score (rMPS) for C_2H_4/C_2H_6 separation.

Figure 7.5 shows the distributions of pore limit diameter (PLD) and the largest cavity diameter (LCD) for all MOFs in CoRE MOF 2019 database and MOFs with top 5% membrane selectivity and rMPS. Compared with the distributions of PLD and LCD for all CoRE MOFs, it is clear that MOFs with top 5% membrane selectivity and rMPS have significantly different structural features. The PLD and LCD distributions suggest that MOFs with top 5% membrane selectivity have much smaller pores than those with top 5% rMPS. However, their void fractions are rather similar, both in the range from 0.4 to 0.7 (shown in Figure 7.6). For MOFs with top 5% membrane selectivity, the PLD mostly distributes between 2.75 Å and 3.5 Å, where the narrow end is even slightly smaller than the Lennard-Jones (LJ) diameter of the methylene group in ethene. The selectivity is maximized because the extremely narrow pore aperture magnifies the difference between C_2H_4 and C_2H_6 in the potential energy of the transition state. By contrast, MOFs with top 5% rMPS have a PLD distribution from 3.75 Å to 4.75 Å. Compared with the MOFs with top 5% membrane selectively, the slightly larger PLD for the MOFs with top 5% rMPS leads to a smaller difference in the potential energy at the transition state between ethane and ethene, and therefore, much higher permeability with intermediate membrane selectivity. Similar to the PLD distribution, the LCD distribution for the MOFs with top 5% membrane selectivity is mostly localized at the smaller pore size than those with top 5% rMPS. For MOFs with top 5% membrane selectivity, the ultra-narrow pore apertures contribute to larger membrane selectivity but smaller diffusion coefficient and lower permeability.

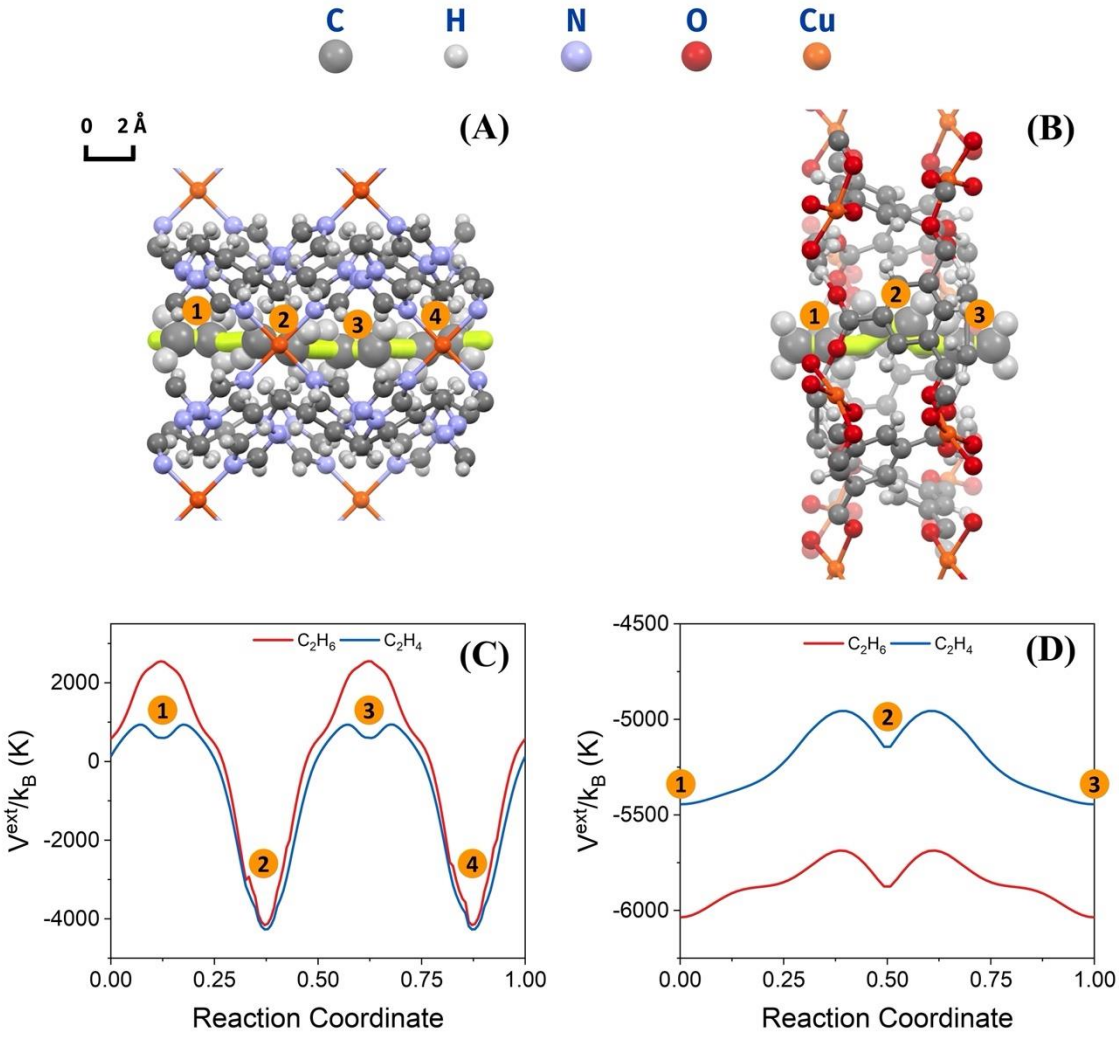

Figure 7.7. The position and orientation of an ethene molecule along the minimum energy path (MEP) in EBINUA02 (A) and LUDLAZ (B). Here the detailed molecular structures are only for illustration. The energy landscape along the MEP for ethane and ethene in EBINUA02 (C) and LUDLAZ (D) at 300 K.

Figure 7.7 shows the minimum energy path (MEP), molecular orientation along the MEP, and the energy landscape for gas diffusion in MOFs with the highest membrane selectivity [EBINUA02: $k^{IM}(C_2H_4/C_2H_6)$=2167.3] and the highest rMPS [LUDLAZ: $k^{IM}(C_2H_6/C_2H_4)$=11.1]. For both EBINUA02 and LUDLAZ, only one direction along the

lattice vector can accommodate the diffusion of ethane or ethene molecule. Figure 7.7A and (B) show that, despite the significant difference between EBINUA02 and LUDLAZ in the energy landscape along the MEP, their MEPs inside MOFs are almost identical. Both EBINUA02 and LUDLAZ yield near straight trajectories for the molecular center of mass on the MEP with a minimal change of the molecular orientation, suggesting that the high membrane selectivity is attributed to extremely narrow pores. It is also worth noticing here the local chemical environments are very similar along the MEP in EBINUA02 and LUDLAZ, and their different pore structures result in the distinct energy landscape along the MEP.

Although EBINUA02 and LUDLAZ have a similar void fraction, their pore structures (e.g., PLD and LCD) are very different thus results in different separation mechanisms. According to the solution-diffusion theory,[18] the membrane selectivity can be improved by increasing the difference in adsorption, diffusion, or a combination of both quantities. EBINUA02 has a much smaller pore aperture (PLD: 2.91 Å and LCD: 3.96 Å) than LUDLAZ (PLD: 4.18 Å and LCD: 5.96 Å), which leads to the preferential diffusion and adsorption (*viz*. solubility in solution-diffusion theory) of C_2H_4 and the extremely high membrane selectivity of C_2H_4 over C_2H_6. Whereas in LUDLAZ, its interaction with the gas molecules is attractive even at the transition state and the difference in energy barrier between C_2H_4 and C_2H_6 is almost negligible (Fig. 3D). The relatively spacious pore structure in LUDLAZ results in a slightly faster diffusion of C_2H_6 than C_2H_4 because ethane experiences a stronger van der Waals attraction. As a result, the membrane selectivity of LUDLAZ is mostly contributed by the difference in

adsorption (*viz*., Henry's constant) between C_2H_4 and C_2H_6. According to the above analysis of the top MOFs with high membrane selectivity and rMPS, a large energy barrier (*viz*., extremely narrow pore aperture) is not preferred for the design of ideal MOF membrane (with both high selectivity and permeability) because it also significantly reduces the diffusion coefficient and permeability. Therefore, for the rational design of ideal MOF membranes for C_2H_4/C_2H_6 separation, the selectivity and permeability need to be harnessed by the adsorption and diffusion, respectively.

7.3.3 Inverse Design of MOF Membranes

Since the membrane process has major advantages in comparison with adsorption in terms of both separation selectivity and capacity, our inverse design is concerned only with MOF membranes. To find materials with the ideal performance (*viz*., high separation selectivity and capacity) for the separation of C_2H_4/C_2H_6, we use a genetic algorithm (GA) with the fitness score of

$$F^{total} = 0.5F^{k^{IM}} + 0.5F^{P} \tag{7.8}$$

In eq (7.8), the total fitness score, F^{total}, is evenly weighted according to the selectivity and permeability. The former, member selectivity fitness score,is formulated as

$$F^{k^{IM}} = \begin{cases} \left(k^{IM} - 5\right)^2 + 2 & k^{IM} \geq 5 \\ e^{k^{IM}-5} + 1 & k^{IM} < 5 \end{cases} \tag{7.9}$$

where k^{IM} stands for the (ethene-selective) membrane selectivity; and the latter,

$$F^{P} = \begin{cases} \log\left(P\right) & P \geq 100 \\ e^{P-100} + 1 & P < 100 \end{cases} \tag{7.10}$$

represents the permeability fitness score. The membrane selectivity of 5 and permeability

of 100 barrers are used as threshold value because most CoRE MOFs have a permeability larger than 100 but few have a membrane selectivity larger than 5.

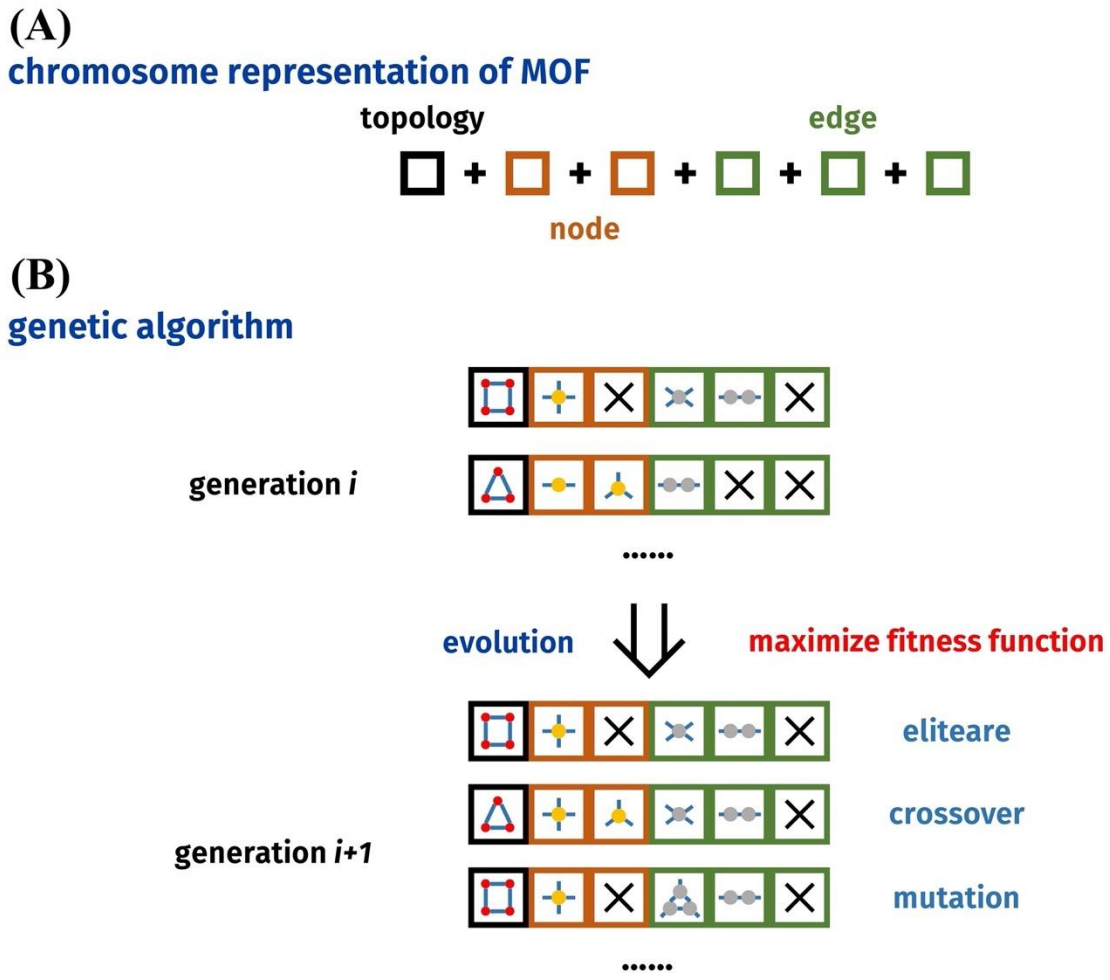

Figure 7.8. (A) Chromosome representation of MOFs investigated in this work where topology, node and edge are treated as genes in the chromosome. (B) Workflow of genetic algorithm for the inverse MOF design. Here square boxes represent the secondary building blocks (SBUs) used for MOF construction. In analogy to the genes in the chromosome, the choice of SBUs directly determines the physiochemical properties and separation performance. Black, brown and green boxes represent the topology, node and edge, respectively.

Figure 7.8 shows how each MOF is represented as a "chromosome", i.e., in terms of the MOF topology, node and edge. The schematic flowchart elucidates the

computational steps in the inverse design of MOF membranes. Compared with CoRE MOF database or other existing MOF structural databases, a much larger chemical space is explored during the inverse design. In this work, the secondary building block (SB) database consists of 1,687 topologies, 648 nodes and 219 edges. As each MOF is defined by a topology, two metal nodes and three edges, a total of $1678\times648^2\times219^3\approx7.4\times10^{15}$ combinations are possible for the MOF construction. It is worth mentioning that not all chromosomes would lead to a successful MOF design because of the unmatched coordination number and bonding distance in the topologies, nodes and linkers.

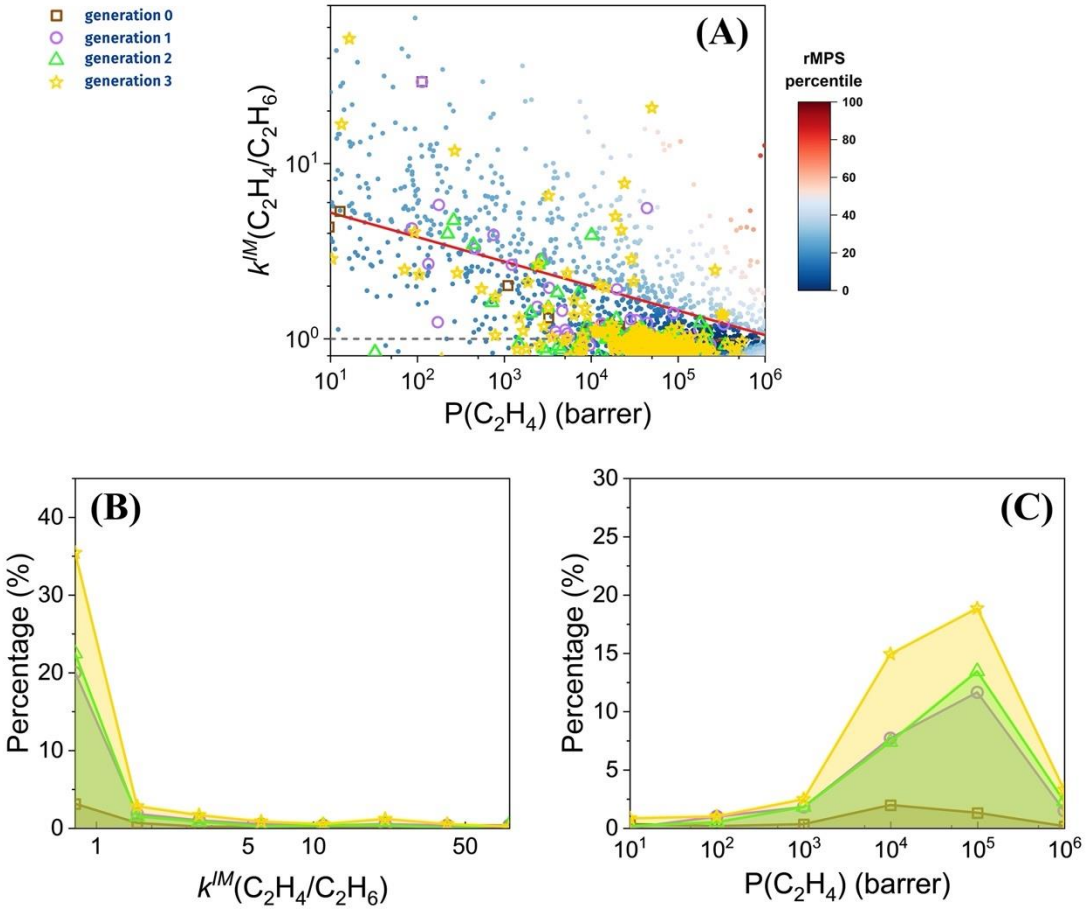

Figure 7.9. (A) Membrane separation selectivity vs. permeability for CoRE MOFs (filled dots) and inverse designed MOFs (open symbols). The distribution of membrane selectivity (B) and permeability (C) for inversed designed MOFs. The red line denotes the Robeson boundary, and color spectrum stands for the percentile of relative membrane performance score (rMPS): the red, white and blue represent the highest, intermediate and the lowest rMPS, respectively. Brown box, purple circle, green triangle and gold star stand for inverse designed MOFs in generation 0, 1, 2, 3, respectively.

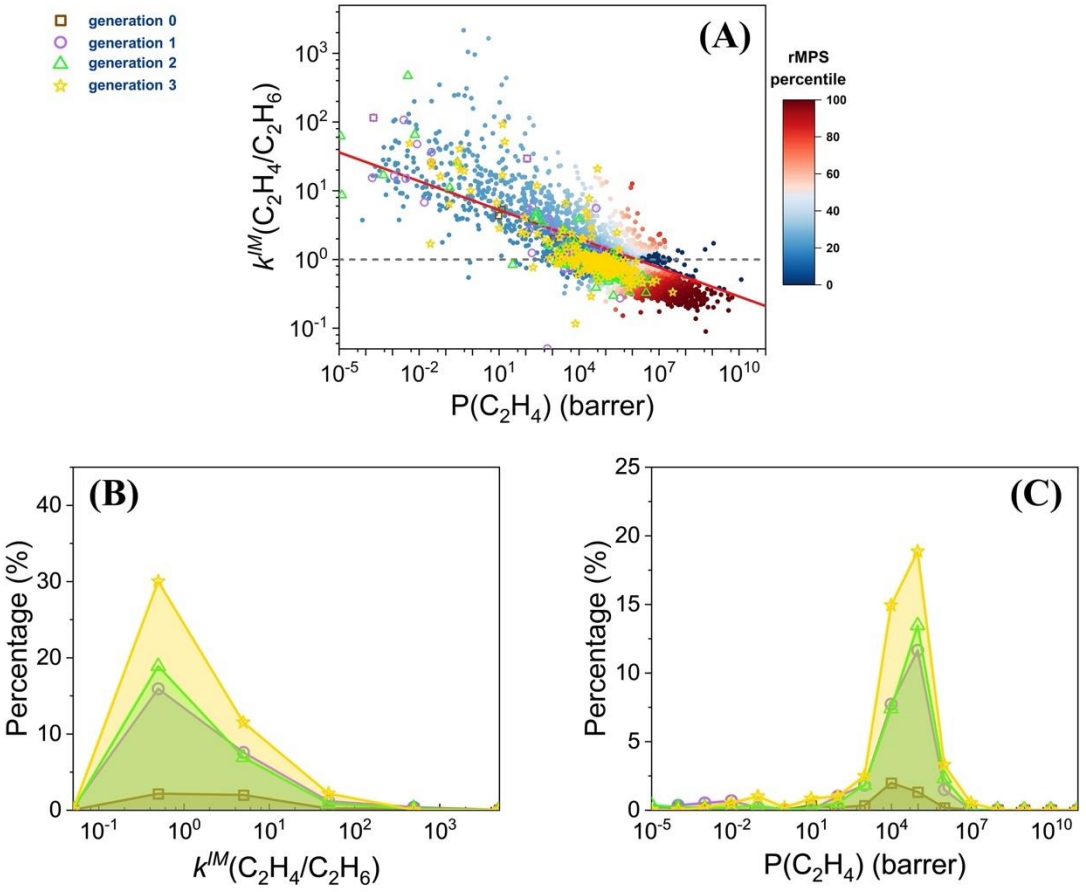

Figure 7.10. (A) Membrane separation selectivity vs. permeability for CoRE MOFs (filled dots) and inverse designed MOFs (open symbols). The distribution of membrane selectivity (B) and permeability (C) for inversed designed MOFs. The red line denotes the Robeson boundary and color spectrum stands for the percentile of relative membrane performance score (rMPS): the red, white and blue represent the highest, intermediate and the lowest rMPS, respectively. Brown box, purple circle, green triangle and gold star stand for inverse designed MOFs in generation 0, 1, 2, 3, respectively.

Figure 7.9 shows the membrane selectivity and permeability of inverse designed MOFs in comparison with CoRE MOFs. Here, we consider materials only in the region where the scale of permeability is similar to the experiment results. Figure 7.10 shows the same figure with the full range of permeability and selectivity. In both Figure 7.9 and 7.10, the distributions of membrane selectivity and permeability for designed MOFs in

each generation are shown in terms of percentage of total designed MOFs. With the evolution of each generation via GA, not only can more MOFs be successfully explored, but MOFs with better targeted properties can be designed as well. As shown in Figure 7.9B and C, the improvement is evident in the increased distribution at higher permeability and membrane selectivity. The final generation (gold stars) contains most candidates surpassed the Robeson boundary and the threshold values of both membrane selectivity and permeability. The best MOF identified by the inverse design (yfk-N379) has a much better overall membrane separation performance in terms of both membrane selectivity and permeability compared to all existing experimental MOFs. Table 7.6 lists the detailed properties of yfk-N379.

Table 7.6. Properties of designed MOFs with both high membrane selectivity and permeability for the separation of C_2H_4/C_2H_6 at 300 K.

MOF	$K_h(C_2H_4)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$K_h(C_2H_6)$ ($mol \cdot m^{-3} \cdot Pa^{-1}$)	$D_0(C_2H_4)$ ($m^2 \cdot s^{-1}$)	$D_0(C_2H_6)$ ($m^2 \cdot s^{-1}$)	$k^{IM}(C_2H_6/C_2H_4)$
yfk-N379	1.682×10^{-2}	9.828×10^{-4}	1.677×10^{-9}	1.467×10^{-9}	20.8

Figure 7.11 shows molecular orientation and energy landscape for the diffusion of gas molecules along the MEP. In yfk-N379, the metal node, $(CO_2\text{-}\kappa^2O)TbO_2(\mu\text{-}CO_2\text{-}\kappa^2O)_4TbO_2(CO_2\text{-}\kappa^2O)$, is connected with the yfk topology network to form a 1D channel for the diffusion of C_2H_4 and C_2H_6 molecules. The energy landscape along the MEP in yfk-N379 is similar to that in LUDLAZ where the intermolecular interaction along the MPE is all attractive. Also, like that in LUDLAZ, the difference in the energy barrier between C_2H_4 and C_2H_6 is relatively small. As discussed above, in order to design an ideal MOF membrane with both high membrane selectivity and permeability, the

separation selectivity shall be harnessed by the difference in the adsorption (*viz*., solubility), and the high permeability should be obtained by fast diffusion. As shown in Figure 7.11B, the 1D channel in yfk-N379 offers strong attraction and relatively moderate energy barrier along the MEP, which results in the extremely fast diffusion of gas molecules. The distinct difference of adsorption properties (*viz*., Henry's constant) between C_2H_4 and C_2H_6 in yfk-N379 leads to an exceptionally high membrane selectivity compared to the CoRE MOFs. As a result, the synergetic effect from adsorption and diffusion helps yfk-N379 achieve both high membrane selectivity and permeability at the same time.

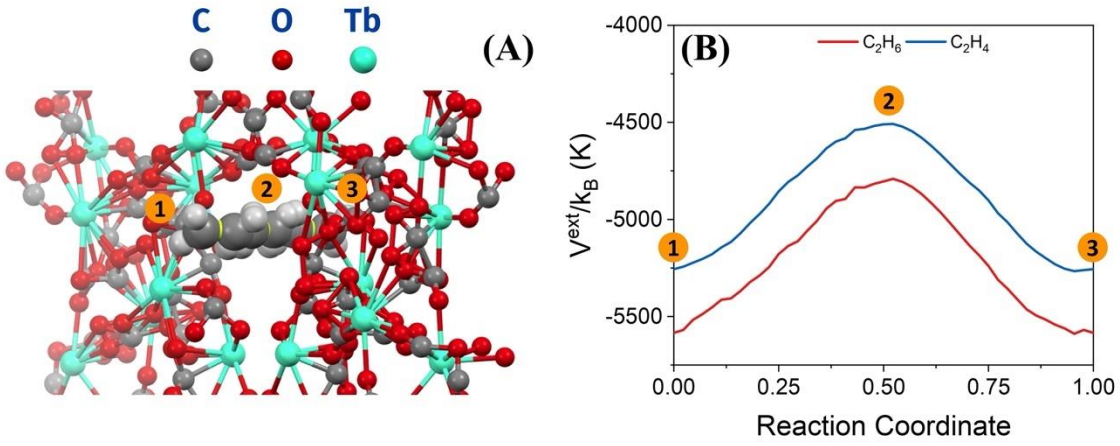

Figure 7.11. (A) The position and orientation of an ethene molecule along the MEP (yellow line) in yfk-N379. (B) Energy landscape along the MEP for ethane and ethene in yfk-N379 at 300 K.

Compared with high-throughput screening, the inverse design via GA is computationally more efficient. For example, the ideal MOF candidate (both high membrane selectivity and permeability) for the membrane separation of C_2H_4/C_2H_6 can be found with only 8,000 attempts in a much larger chemical space. The computational

workflow thus demonstrates that, with theoretical tools for the efficient evaluation of materials performance, how the inverse design can significantly accelerate the material discovery, especially for the construction of reticular materials (e.g., MOFs and COFs) for the gas storage and separation.

7.4 Conclusions

In this work, we used both high-throughput screening and inverse design to find the best metal-organic frameworks (MOFs) for C_2H_4/C_2H_6 separation. Both adsorption and membrane processes have been considered in the high-throughput screening of computational-ready experimental (CoRE 2019) MOF database. For adsorption separation, the separation selectivity of ethene-selective MOF decreases with the increase of separation capacity because highly ethene-selective materials have extremely small pores with low adsorption capacity. While the selectivity of ethane-selective MOF increases with the adsorption capacity, the highest adsorption selectivity [$\alpha^{IM}(C_2H_6/C_2H_4)$=7.68] of ethane-selective MOF (LUDLAZ) is smaller than that [$\alpha^{IM}(C_2H_4/C_2H_6)$=27.01] of ethene-selective MOF (PIRYOF). Nevertheless, LUDLAZ is more than 70% higher than the state-of-art ethane-selective MOF identified by previous work.

Compared with that in the adsorption process, the selectivity of membrane process is less compromised by the increase of the separation capacity. Through high-throughput screening, we find that EBINUA02 yields the highest membrane selectivity [$k^{IM}(C_2H_4/C_2H_6)$=2167.3]. To evaluate the overall membrane performance, we introduced a relative membrane performance score (rMPS) in terms of the selectivity and

permeability with respect to the Robeson boundary. For the separation of C_2H_4/C_2H_6, high rMPS favors MOFs with high permeability and intermediate membrane selectivity because high membrane selectivity requires large energy barrier along the minimum energy path (MEP) and leads to the slow diffusion. According to the structural analysis, MOFs with top 5% membrane selectivity have a much more confined diffusion path in terms of pore limit diameter (PLD) and largest cavity diameter (LCD) than those with top 5% rMPS, despite their similarity in the distribution of void fraction. The separation mechanism is quite different between EBINUA02 and the MOF (LUDLAZ) with the highest rMPS. While a small pore aperture (PLD: 2.91 Å and LCD: 3.96 Å) in EBINUA02 results in faster diffusion and stronger adsorption of C_2H_4 over C_2H_6, the less confined diffusion path in LUDLAZ (PLD: 4.18 Å and LCD: 5.96 Å) leads to negligible difference in the diffusion. In that case, the membrane selectivity is mostly contributed by its ethane-selective solubility.

The computational efficiency of the theoretical tools for predicting the sorption and diffusion properties of nanoporous materials allows us to design MOF membranes with both high membrane selectivity and permeability using the genetic algorithm (GA). Compared with high-throughput screening, not only can GA explore the material design space with targeted properties, but it takes less attempts to identify the most promising candidates as well. The best MOF discovered by GA consists metal node - $(CO_2\text{-}\kappa^2O)TbO_2(\mu\text{-}CO_2\text{-}\kappa^2O)_4TbO_2(CO_2\text{-}\kappa^2O)$ with the yfk topology. The designed material has both permeability and membrane selectivity significantly larger than the threshold values set in the fitness function. Besides, its overall membrane separation performance is better

than all existing experimental MOF candidates. The computational workflow used in the work thus demonstrates the capability of inverse design to accelerate the discovery of nanoporous materials, especially reticular materials (such as MOFs and COFs) for gas storage and separation.

Bibliography

1. Sholl, D.S. and R.P. Lively, *Seven chemical separations to change the world.* Nature, 2016. **532**(7600): p. 435-7.

2. Rungta, M., et al., *Membrane‐based ethylene/ethane separation: The upper bound and beyond.* AIChE Journal, 2013. **59**(9): p. 3475-3489.

3. Ding, Q., et al., *Exploiting equilibrium-kinetic synergetic effect for separation of ethylene and ethane in a microporous metal-organic framework.* Science advances, 2020. **6**(15): p. eaaz4322.

4. Li, L., et al., *Ethane/ethylene separation in a metal-organic framework with iron-peroxo sites.* Science, 2018. **362**(6413): p. 443-446.

5. Lv, D., et al., *Recent advances in adsorptive separation of ethane and ethylene by C2H6-selective MOFs and other adsorbents.* Chemical Engineering Journal, 2021: p. 133208.

6. Rangnekar, N., et al., *Zeolite membranes - a review and comparison with MOFs.* Chem Soc Rev, 2015. **44**(20): p. 7128-54.

7. Li, J.R., J. Sculley, and H.C. Zhou, *Metal-organic frameworks for separations.* Chemical Reviews (Washington, D. C.), 2012. **112**(2): p. 869-932.

8. Tranchemontagne, D.J., et al., *Secondary building units, nets and bonding in the chemistry of metal-organic frameworks.* Chem Soc Rev, 2009. **38**(5): p. 1257-83.

9. Zhou, M., A. Vassallo, and J. Wu, *Toward the inverse design of MOF membranes for efficient D2/H2 separation by combination of physics-based and data-driven modeling.* Journal of Membrane Science, 2020. **598**.

10. Wang, J., et al., *Virtual Screening of Nanoporous Materials for Noble Gas Separation.* ACS Applied Nano Materials, 2022.

11. Bárcia, P.S., et al., *Reverse shape selectivity in the adsorption of hexane and xylene isomers in MOF UiO-66.* Microporous and Mesoporous Materials, 2011. **139**(1-3): p. 67-73.

12. Oh, H., et al., *A cryogenically flexible covalent organic framework for efficient hydrogen isotope separation by quantum sieving.* Angew Chem Int Ed Engl, 2013. **52**(50): p. 13219-22.

13. Chen, G., et al., *M-gallate MOF/6FDA-polyimide mixed-matrix membranes for C2H4/C2H6 separation.* Journal of Membrane Science, 2021. **620**: p. 118852.

14. Bao, Z., et al., *Molecular Sieving of Ethane from Ethylene through the Molecular Cross-Section Size Differentiation in Gallate-based Metal–Organic Frameworks.* Angewandte Chemie International Edition, 2018. **57**(49): p. 16020-16025.

15. Zhou, M.S., et al., *Fractionation of Isotopic Methanes with Metal-Organic Frameworks.* Journal of Physical Chemistry C, 2019. **123**(12): p. 7397-7407.

16. Wang, J.Q., et al., *Computational screening and design of nanoporous membranes for efficient carbon isotope separation.* Green Energy & Environment, 2020. **5**(3): p. 364-373.

17. Kang, M., et al., *High-Throughput Discovery of Ni (IN) 2 for Ethane/Ethylene Separation.* Advanced Science, 2021. **8**(11): p. 2004940.

18. Wijmans, J.G. and R.W. Baker, *The solution-diffusion model: a review.* Journal of membrane science, 1995. **107**(1-2): p. 1-21.

19. Altintas, C. and S. Keskin, *Computational screening of MOFs for C2H6/C2H4 and C2H6/CH4 separations.* Chemical Engineering Science, 2016. **139**: p. 49-60.

20. Tang, H. and J. Jiang, *In silico screening and design strategies of ethane-selective metal–organic frameworks for ethane/ethylene separation.* AIChE Journal, 2021. **67**(3): p. e17025.

21. Halder, P. and J.K. Singh, *High-throughput screening of metal–organic frameworks for ethane–ethylene separation using the machine learning technique.* Energy & Fuels, 2020. **34**(11): p. 14591-14597.

22. Kim, J., et al., *Large-scale computational screening of zeolites for ethane/ethene separation.* Langmuir, 2012. **28**(32): p. 11914-11919.

23. Kärger, J., D.M. Ruthven, and D.N. Theodorou, *Diffusion in Nanoporous Materials*. 2012.

24. Verploegh, R.J., S. Nair, and D.S. Sholl, *Temperature and loading-dependent diffusion of light hydrocarbons in ZIF-8 as predicted through fully flexible molecular simulations.* Journal of the American Chemical Society, 2015. **137**(50): p. 15760-15771.

25. Zhou, M. and J. Wu, *Massively Parallel GPU-Accelerated String Method for Fast and Accurate Prediction of Molecular Diffusivity in Nanoporous Materials.* ACS Applied Nano Materials, 2021. **4**(5): p. 5394-5403.

26. Zhou, M. and J. Wu, *A GPU implementation of classical density functional theory for rapid prediction of gas adsorption in nanoporous materials.* J Chem Phys,

2020. **153**(7): p. 074101.

27. Glaser, J., et al., *Strong scaling of general-purpose molecular dynamics simulations on GPUs.* Computer Physics Communications, 2015. **192**: p. 97-107.

28. Kutzner, C., et al., *More bang for your buck: Improved use of GPU nodes for GROMACS 2018.* Journal of computational chemistry, 2019. **40**(27): p. 2418-2431.

29. Kim, B., S. Lee, and J. Kim, *Inverse design of porous materials using artificial neural networks.* Sci Adv, 2020. **6**(1): p. eaax9324.

30. Yao, Z.P., et al., *Inverse design of nanoporous crystalline reticular materials with deep generative models.* Nature Machine Intelligence, 2021. **3**(1): p. 76-86.

31. Gustafson, J.A. and C.E. Wilmer, *Intelligent selection of metal–organic framework arrays for methane sensing via genetic algorithms.* ACS sensors, 2019. **4**(6): p. 1586-1593.

32. Chung, Y.G., et al., *In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm.* Science advances, 2016. **2**(10): p. e1600909.

33. Dou, H., et al., *Boron nitride membranes with a distinct nanoconfinement effect for efficient ethylene/ethane separation.* Angewandte Chemie, 2019. **131**(39): p. 14107-14113.

34. Wu, Y., et al., *Adsorption and separation of ethane/ethylene on ZIFs with various topologies: Combining GCMC simulation with the ideal adsorbed solution theory (IAST).* Chemical Engineering Science, 2015. **124**: p. 144-153.

35. Rappe, A.K., et al., *Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations.* J Am Chem Soc, 1992. **114**(25): p. 10024-10035.

36. Willems, T.F., et al., *Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials.* Microporous and Mesoporous Materials, 2012. **149**(1): p. 134-141.

37. Cotterman, R.L., B.J. Schwarz, and J.M. Prausnitz, *Molecular thermodynamics for fluids at low and high densities. Part I: Pure fluids containing small or large molecules.* AIChE J, 1986. **32**(11): p. 1787-1798.

38. Haldoupis, E., S. Nair, and D.S. Sholl, *Efficient calculation of diffusion limitations in metal organic framework materials: a tool for identifying materials*

for kinetic separations. J Am Chem Soc, 2010. **132**(21): p. 7528-39.

39. Bennett, C. *Algorithms for chemical computations*. in *ACS symposium Series*. 1977. Am. Chem. Soc Washington.

40. Chandler, D., *Statistical mechanics of isomerization dynamics in liquids and the transition state approximation.* The Journal of Chemical Physics, 1978. **68**(6): p. 2959-2970.

41. Lee, S., et al., *Computational Screening of Trillions of Metal–Organic Frameworks for High-Performance Methane Storage.* ACS Applied Materials & Interfaces, 2021. **13**(20): p. 23647-23654.

42. Gad, A.F., *Pygad: An intuitive genetic algorithm python library*. arXiv preprint arXiv:2106.06158, 2021.

43. McKellar, S.C., et al., *The effect of pressure on the post-synthetic modification of a nanoporous metal–organic framework.* Nanoscale, 2014. **6**(8): p. 4163-4173.

44. Tian, L., et al., *Novel flexible bis-triazole bridged copper (II) coordination polymers varying from one-to three-dimensionality*. CrystEngComm, 2012. **14**(6): p. 2032-2039.

Chapter 8. Insights from Machine Learning of Carbon Electrodes for Electric Double Layer Capacitors

Recent years have witnessed the broad use of carbon electrodes for electric double layer capacitors (EDLCs) because of large surface area, high porosity and low cost. Whereas experimental investigations are mostly focused on the device performance, computational studies have been rarely concerned with electrochemical properties at conditions remote from equilibrium, limiting their direct applications to materials design. In this chapter, through a comprehensive analysis of extensive experimental data with various machine learning methods, we report herein quantitative correlations between the structural features of carbon electrodes and the *in-operando* behavior of EDLCs including energy and power density. Machine learning models allow us to identify important characteristics of activated carbons useful to optimize their efficiency in energy storage.

8.1 Introduction

Electric double layer capacitors (EDLCs, a.k.a. supercapacitors) have attracted tremendous interest in recent years due to their great potential for energy storage with long maintenance-free lifetime, high cycle efficiency and high power density[1-3]. EDLCs may well bridge the existing energy-storage gap between conventional capacitors and electrochemical batteries in terms of energy and power density. It is generally believed that EDLCs will play a vital role for energy storage not only in industrial sectors but also in consumer electronics such as forklifts, electric vehicles, and memory backup

batteries. A critical issue amid many promising applications is low energy density, which has been a prominent limiting factor for more widespread usage of EDLCs.

EDLCs harness electrical energy by reversible adsorption/desorption of ionic species. Because electrode materials with large surface areas yield large capacitances, porous carbons are an excellent choice for high-density energy storage. In recent years, carbon-based materials have been widely used as the EDLC electrodes because of large specific surface area, high porosity, high conductivity, low cost, easy availability, and environmental friendliness[4, 5]. Despite extensive computational studies[6], yet-to-be-established are quantitative correlations between the characteristics of carbon-based materials (i.e., pore connectivity and size/shape distributions) and the EDLC performance at conditions of practical interest. A conventional wisdom is that the higher surface area an electrode has, the larger capacitance it would achieve. Accordingly, one might expect that micropores (< 2 nm) would be most desirable as they offer significantly more specific surface areas than mesopores (> 2 nm and < 50 nm) or macropores (> 50 nm). However, it has been well documented that an electrode with only micropores suffers not only from poor power density but from low capacitance as well[4, 7-9]. Micropores often lead to unsatisfactory power density due to the increased resistance of ion diffusion[4, 9]. Besides, the reduction in the capacitance might be attributed to inaccessibility of micropores due to sieving and/or transport effects[7, 8]. Because a precise characterization of pore size and shape distribution is experimentally challenging for amorphous porous materials, inconsistent results have been reported concerning the pore size effects on the EDLC performance in particular when the dimension of micropores is

comparable to that of ionic species[10-15]. One consensus is, however, that an interconnected 3D pore network may lead to both satisfactory capacitance and power density by optimizing the surface areas of mesopores and micropores simultaneously[4]. To tailor carbon materials for better performance of EDLCs, we need quantitative correlations of their structural features with the power and energy density at *in-operando* conditions.

Equivalent circuit models (ECMs) are conventionally used to describe the kinetics of charging and discharging for energy storage devices[16]. While ECMs are appealing for their simplicity and effectiveness to correlate experimental data, such models are not applicable to electrodes with micropores and provide little insights on the microscopic details of physical processes underlying ion transport and energy storage[17]. Conversely, molecular models are powerful to describe the ionic behavior, but their applicability to EDLCs is severely limited not only by the system size but by near equilibrium conditions. In recent years, machine learning (ML) methods are emerging as a powerful alternative to physics-based approaches[18-20]. Provided that sufficient training data are variable, ML methods allow us to establish useful correlations between materials properties and their performance without evoking the physical details. Different from physics-based modeling, ML methods hinge on computational algorithms to self-educate and extract quantitative relationships between input and output variables from existing data. Although justification of a ML algorithm for practical applications often requires physics knowledge and, sometimes, trial and error, once calibrated ML methods are able to describe complicated physical processes solely based on the input data.

Previously, ML methods have been used to predict the performance of porous material as effective media for gas adsorption or separation[20-22] and as electrodes for energy storage[23-26]. For example, Zhu et al. complied over 10,000 data points for the capacitance of carbon-based supercapacitors from more than 1000 publications[23]. A useful correlation was established between the EDLC capacitance and the physical and chemical features of carbon materials including specific surface area, pore volume, ID/IG ratio, and doping elements by using an artificial neutron network (ANN). To identify the dependence of the EDLC capacitance on the individual features of carbon materials, Su et al. established quantitative correlations using regression trees (RT) and multi-layer perception (MLP) models[24]. Although good correlations have been shown, it remains unclear how ML methods would lead to a better understanding of the relation between the capacitance and the structural features of carbon materials. In particular, the performance of EDLCs is dependent on operational conditions but the effect of scan rate on the capacitance and power density was not explicitly considered in previous investigations.

In this work, we propose a physics-informed ML method to unravel the structural features of carbon electrodes that may have the most significant impacts on the capacitance and power density of EDLCs. To minimize the number of input variables, the computational analysis is focused on the *in-operando* performance of EDLCs consisting of pristine activated carbon materials and an aqueous electrolyte solution (6 M KOH) that is most commonly used in practical applications.

8.2 Methods and Models

The EDLC performance depends on, in addition to electrode materials, the properties of electrolytes and operational conditions such as the electrochemical potential window and charging discharging rates. In order to minimize the number of input variables affecting the performance of an EDLC device, our machine-leaning analysis is focused on experimental results for the specific capacitance and the power density of pristine activated carbon materials. Besides, we consider only the results measured with three-electrode cell in 6 M KOH aqueous solution with the voltage window of 1 V. In comparison to that in two-electrode measurements, a three-electrode cell provides a more precise control of both potential and current, and thus is able to better distinguish the electrochemical properties of different electrode materials[3]. Because experimental data are scarce for materials with low specific surface areas, we added three points to replenish the results at zero surface area. All data points are from the literature[27-33] and listed in Table 8.1. The specific integral capacitance is given by

$$C_{sp} = \frac{\int_{V_{start}}^{V_{end}} i(V)dV}{2\upsilon m\Delta V} = \frac{\bar{I}\dfrac{\Delta V}{\upsilon}}{m\Delta V} = \frac{\bar{I}\Delta t}{m\Delta V} \tag{8.1}$$

where υ is the scan rate (V/s), i is the electrical current, m is the electrode mass, ΔV is the potential window, $\bar{I}$ is the average current, $\Delta t=\Delta V/\upsilon$ is charging/discharging time, and C_{sp} stands for specific integral capacitance of the electrode in a three-electrode system. Following the experimental literature, the energy density is defined as

$$E = \frac{C_{cell}\Delta V^2}{2} = \frac{C_{sp}\Delta V^2}{8} = \frac{\bar{I}\Delta t\Delta V}{8m} \tag{8.2}$$

where C_{cell} is the specific capacitance of a two-electrode symmetrical supercapacitor. Accordingly, the power density is calculated from

$$P = \frac{E}{\Delta t} = \frac{\bar{I}\Delta V}{8m} \tag{8.3}$$

It should be emphasized that none of the quantities calculated above correspond to equilibrium values. Nevertheless, these quantities are most relevant to evaluating the performance of EDL capacitors under practical conditions.

Table 8.1. Data used in this work.

#	C_{sp} (F/g)	E (Wh/kg)	P (kW/kg)	SA_{micro} (m^2/g)	SA_{meso} (m^2/g)	v (mV/s)
1	0	0	0	0	0	0
2	0	0	0	0	0	5
3	0	0	0	0	0	10
4	188.58	6.548	0.118	1990	879	5
5	232.27	8.065	0.145	636	442	5
6	222.77	7.735	0.278	636	442	10
7	202.29	7.024	0.506	636	442	20
8	185.15	6.429	1.157	636	442	50
9	155.41	5.396	1.943	636	442	100
10	185.11	6.428	0.116	713	290	5
11	170.51	5.921	0.213	457	126	10
12	101.47	3.523	1.268	457	126	100
13	160.84	5.585	0.201	429	188	10
14	115.52	4.011	1.444	429	188	100
15	175.29	6.086	0.219	481	193	10
16	141.55	4.915	1.769	481	193	100
17	253.90	8.816	0.317	1118	504	10
18	203.05	7.050	2.538	1118	504	100
19	224.15	7.783	0.056	735	1200	2
20	202.99	7.048	0.127	735	1200	5
21	189.89	6.593	0.237	735	1200	10
22	176.24	6.119	0.441	735	1200	20
23	144.14	5.005	0.901	735	1200	50
24	113.60	0.394	0.142	735	1200	100
25	241.54	8.387	0.151	1506	269	5
26	212.44	7.376	0.266	1506	269	10
27	207.12	7.192	0.518	1506	269	20
28	197.94	6.873	1.237	1506	269	50
29	198.00	6.875	2.475	1506	269	100
30	182.58	6.340	0.114	437	10	5
31	161.70	5.615	1.011	437	10	50
32	158.97	5.520	1.987	437	10	100
33	221.86	7.703	0.139	501	25	5
34	191.76	6.658	1.198	501	25	50
35	182.59	6.340	2.282	501	25	100
36	159.09	5.524	0.099	579	83	5
37	139.68	4.850	0.873	579	83	50
38	136.66	4.745	1.708	579	83	100
39	116.85	4.057	0.029	0	24	2
40	79.11	2.747	0.049	0	24	5
41	68.03	2.362	0.085	0	24	10

42	61.20	2.125	0.153	0	24	20
43	53.48	1.857	0.334	0	24	50
44	46.58	1.617	0.582	0	24	100
45	41.30	1.434	1.033	0	24	200
46	31.42	1.091	1.964	0	24	500
47	257.94	8.956	0.064	115	1158	2
48	244.31	8.483	0.153	115	1158	5
49	238.34	8.276	0.298	115	1158	10
50	232.37	8.068	0.581	115	1158	20
51	224.65	7.800	1.404	115	1158	50
52	216.90	7.531	2.711	115	1158	100
53	207.36	7.200	5.184	115	1158	200
54	187.26	6.502	11.704	115	1158	500
55	179.60	6.236	0.022	120	216	1
56	172.40	5.986	0.043	120	216	2
57	166.30	5.774	0.104	120	216	5
58	155.00	5.382	0.194	120	216	10
59	211.60	7.347	0.026	107	315	1
60	201.60	7.000	0.050	107	315	2
61	184.20	6.396	0.115	107	315	5
62	172.60	5.993	0.216	107	315	10
63	277.00	9.618	0.035	153	553	1
64	259.60	9.014	0.065	153	553	2
65	229.50	7.969	0.143	153	553	5
66	198.10	6.878	0.248	153	553	10
67	280.10	9.726	0.035	200	900	1
68	273.5	9.497	0.068	200	900	2
69	265.2	9.208	0.166	200	900	5
70	250.1	8.684	0.313	200	900	10

Note: 4-10[27]; 11-18[28]; 19-24[29]; 25-29[30]; 30-38[31]; 39-54[32]; 55-70[33];

To analyze the experimental data with ML methods, we have calibrated the performance of four regression models through *leave-one-out* cross validation. One of the most elementary ML models is *linear regression*, which expresses the experimental results as a linear function of the model parameters[34]. The *generalized linear regression* (GLR) method contains not only the intercept and linear terms, but also the products of different features[35]. Because most features used for data-based modelling do not affect the output independently, we expect that GLR is able to better extract

correlation between different features from the input data. In the *support-vector machine* (SVM) model, a hyperplane is constructed to map the input data from the original finite-dimensional space to that corresponding to a higher-dimensional space[36, 37]. By doing so, we can better distinguish data with different characteristics because the kernel trick used in SVM is able to well recognize data patterns. SVM can be used for both regression and classification; it represents one of the most popular classification algorithms. The *random forest* (RF) model is based on the *decision tree* (DT) model, it employs many decision trees ("if-then" logic consequence) to extract useful information from the input data[38, 39]. The output of RF is determined by the vote of all decision trees. Compared with DT, the RF model can avoid overfitting as every decision tree has an equal weight. Finally, the *artificial neural network* (ANN) mimics neutral cells in a human brain by constructing a layer of neurons (transfer function) between input and output[40, 41]. In this work, the ANN model is trained with the Bayesian regularization backpropagation training function as it avoids overfitting with scarce and sparse data[42].

For all ML models, the parameters are optimized with the *k*-fold validation method using a *k*-fold value of 10. The performance of each ML model is evaluated through *leave-one-out* cross validation. Both coefficient of determination (R^2) and root mean square error (RMSE) are used as the metrics to benchmark different ML models:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - u_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{u}\right)^2} \tag{8.4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - u_i\right)^2}{n}} \tag{8.5}$$

where n, y_i, u_i and $\bar{u}$ are, respectively, the number of data points, the value calculated from the ML correlation, the experimental value, and the average of all experimental data.

Table 8.2. Summary of inputs and outputs used for different ML models

Inputs	Output
SA_{micro}	Specific Capacitance
SA_{meso}	
Scan Rate	Power Density

Table 8.2 summarizes the input and output variables used for different ML models. Because we are concerned only with pristine activated carbons, there is no need to consider chemical information (e.g. doping element and percentage) explicitly. As the EDLC energy is manifested at electrolyte-electrode interface, we conjecture that surface area would be more important than pore volume to characterize the porous electrodes. To account for the pore-size effect on capacitance, we distinguish surface areas from micropores (< 2 nm) and mesopores (> 2 nm and <50 nm). For each type of pores, the specific capacitance per surface area is approximately a constant for the same electrolyte[43, 44]. For the aqueous electrolyte considered in this work (6M KOH), the size of the hydrated ions is comparable to that of N_2 molecules[7]. As a result, the surface area measured from N_2 adsorption isotherm at 77 K can well characterize the surface where the ionic adsorption/desorption takes place[11-13]. The BET surface areas of micropores and mesopores thus represent a good choice for the input features of carbon electrodes.

In previous ML studies [23, 24], the scan rate was not taken into consideration for the correlation between capacitance and materials features. However, the scan rate determines the duration of the EDLC charging/discharging and is directly related to the

integral capacitance measured by experiments (see eq [8.1]). In general, the EDLC capacitance measured by cyclic voltammetry at a low scan rate would be higher than that at a high scan rate, and the equilibrium capacitance is achieved only in the limit of infinitely slow charging. For example, an increase in the integral capacitance was observed experimentally when the scan rate is even below 0.2 mV/s [45]. For better comparison of different features of the carbon electrodes influencing the EDLC performance, it is critical to include the scan rate as an input variable. This kinetic parameter is particularly important for evaluating the EDLC performance at *in-operando* conditions[46]. It should be noted that the scan rate used in cyclic voltammetry experiments is not the same as the charging rate as often considered in theoretical studies. The former is referred to the duration of EDL charging/discharging during a cyclic process while the latter is defined as how fast a voltage change is applied to the system without cyclic changes. In terms of charging kinetics, recent MD simulations indicate that a rapid charging rate may lead to the formation of an electroneutral ionic liquid region or co-ion trapping thereby slow charging kinetics [17, 47].

8.3 Results and Discussion

8.3.1 Model Evaluation

We first present results for validating the four ML algorithms discussed above: (A) generalized linear regression (GLR), (B) support vector machine (SVM), (C) random forest (RF), and (D) artificial neural network (ANN). Figure 8.1 shows correlations between the experimental data and results from different ML models for the specific capacitance of various activated carbons. Here 70 data points were used in training the

ML models. In each panel, the diagonal line stands for the perfect correlation between experimental and predicted values.

Figure 8.1 shows that, although all ML methods are able to capture the right trends, their efficiency in representing the experimental results is apparently different. Among the four ML methods considered in this work, GLR results in the worst statistics as is evident from significant deviations of many data points from the diagonal line and as indicated by the lowest R^2 and highest RMSE values. Apparently, multiplication of different features or the quadratic approximation is insufficient to capture the correlation between different input parameters. Both SVM and RF can provide much better fittings of the experimental data since they both have much higher R^2 and lower RMSE values. However, they fail to reproduce the experimental results when the capacitance approaches zero. Overall, ANN provides the best fitting because it gives the highest R^2 and lowest RMSE values.

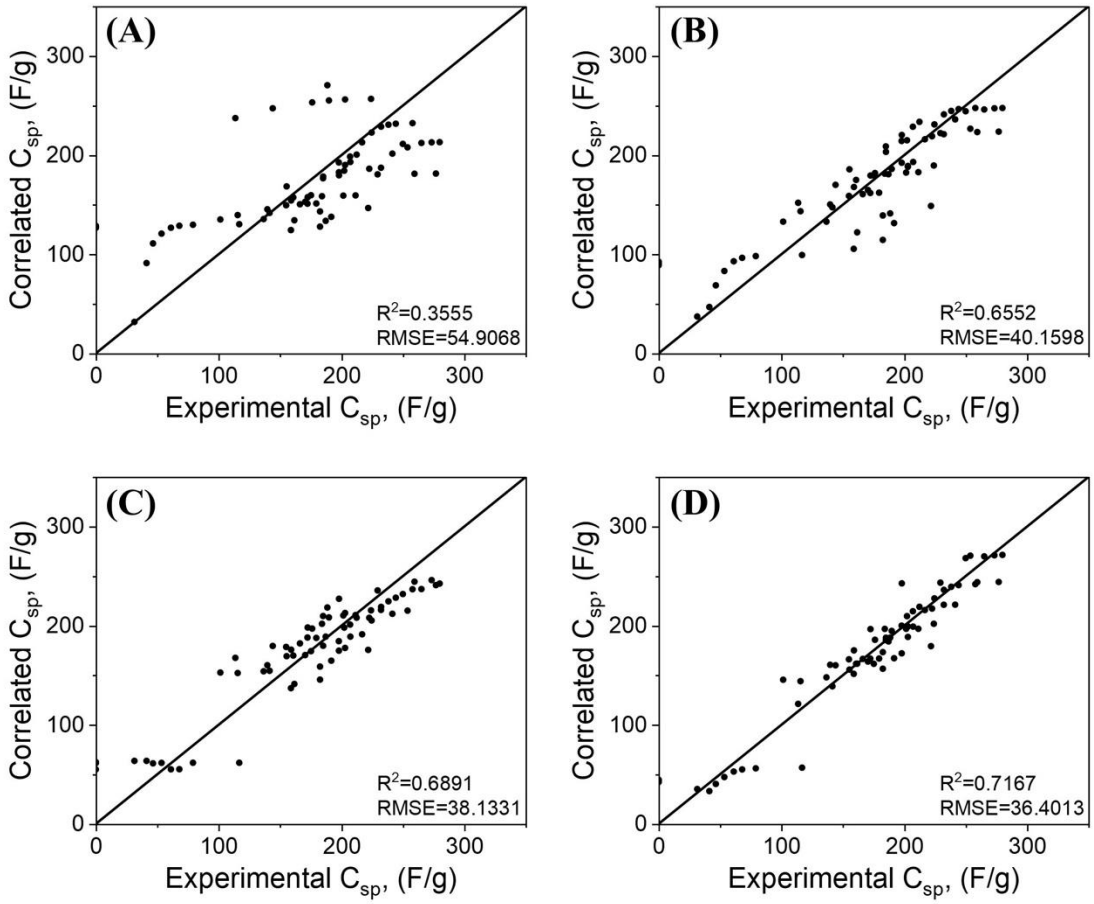

Figure 8.1. Correlations between experimental and different machine learning models for the specific capacitance (C_{sp}, F/g) of activated carbons: (A) generalized linear regression (GLR), (B) support vector machine (SVM), (C) random forest (RF), (D) artificial neural network (ANN). In each panel, the diagonal line represents the perfect correlation between experimental and machine-learning results.

As ANN is able to capture the correlation between the structural features of activated carbons and their influence on EDLC capacitance, we further utilize it to analyze the power density based on the mesopore surface area, the micropore surface area, and the scan rate. Figure 8.2 shows that ANN also provides satisfactory correlation of the power density data with experimental measurements. It gives an excellent correlation over the entire range of the power density. Correlations using other ML models are

provided in Supporting Information (Figure 8.3). It should be noted that ANN yields a larger RMSE for capacitance than that for power density because the absolute values of the capacitance are larger. If a dimensionless metric was used to benchmark ANN for power density and energy density, their performances are rather similar (0.7167 vs 0.6382).

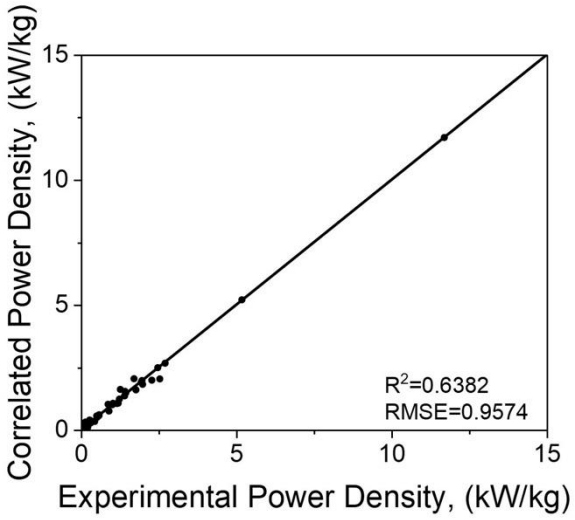

Figure 8.2. Artificial neural network (ANN) model for correlating the power density of activated carbons. The diagonal line represents the perfect correlation between experimental and machine-learning results.

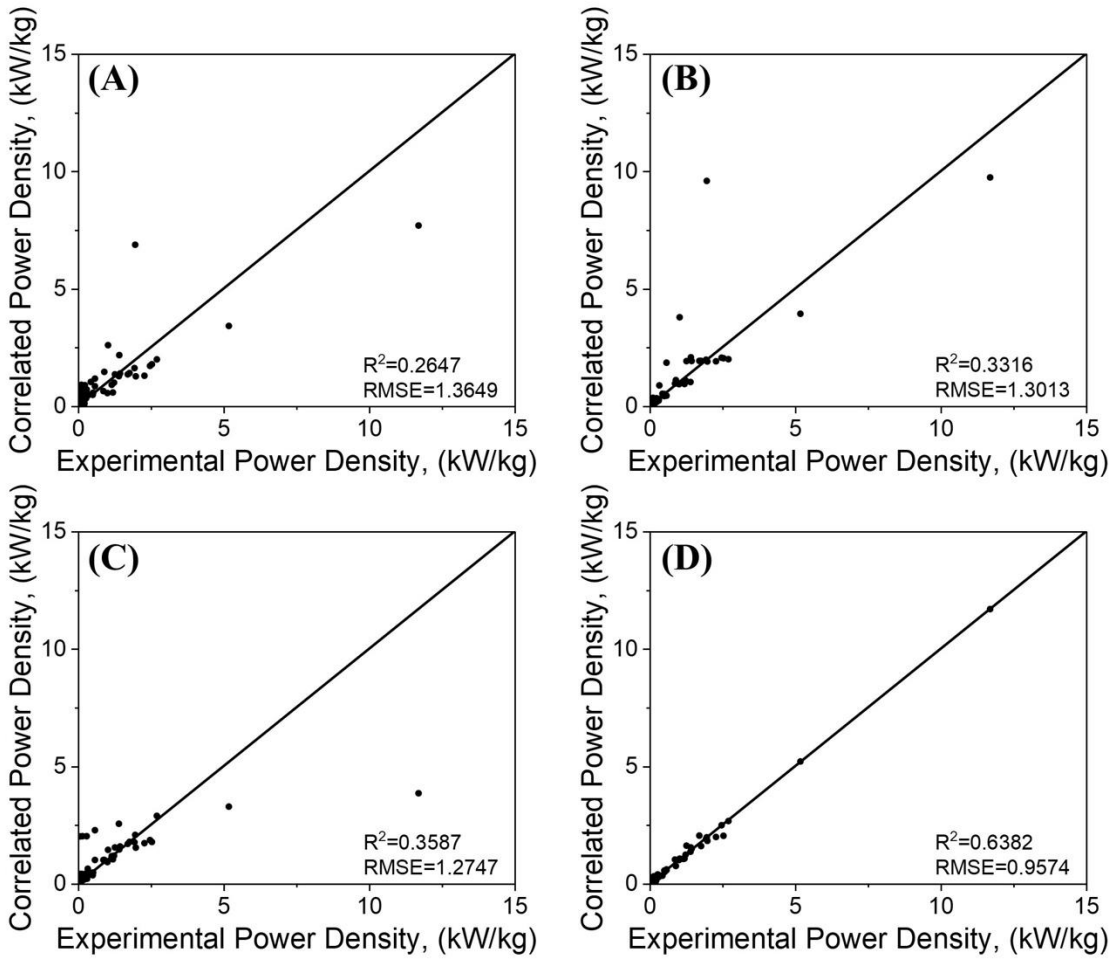

Figure 8.3. Power density from experiment and different machine learning models (A) generalized linear regression, (B) support vector machine, (C) random forest and (D) artificial neural network. The diagonal line indicates perfect correlation between experimental and machine learning results.

One of the biggest challenges in physics-based modeling of energy storage devices is that most theoretical models are valid only for systems at or near equilibrium while the device performance depend on properties remote from equilibrium. As a result, substantial gaps exist between the scopes of theoretical and experimental investigations. For example, the specific capacitance corresponding to the equilibrium value can be measured only in the limit of zero scan rate. However, reliable results can hardly be

obtained in experiment at an extremely low scan rate. Besides, the charge/discharge processes typically take place at high rates in order to fully utilize of the unique properties of supercapacitors. Different from physics-based modeling, ML methods can be used to correlate input and output variables regardless of physical conditions. As shown in Figure 8.4, the ANN model is able to predict the specific capacitance as a function of the scan rate once such a correlation is established. We see that ANN can well reproduce the experimental data for carbon electrodes over a large range of the scan rate (from 2 mV to 500 mV).

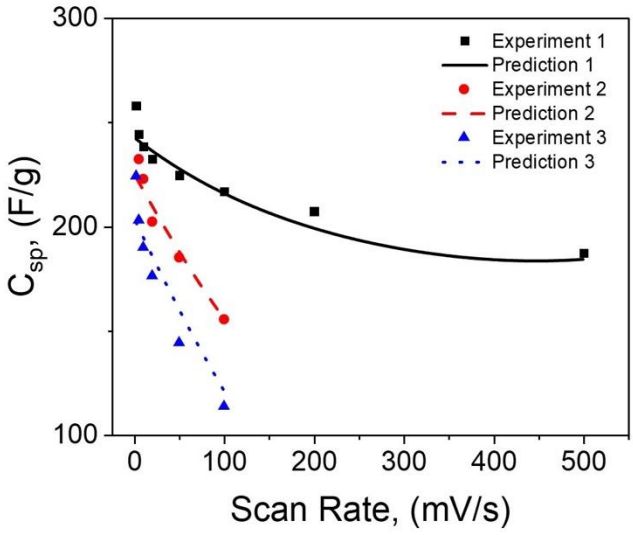

Figure 8.4. The specific capacitance versus the scan rate predicted by the ANN model. Here black squares are experimental data for an electrode with SA_{micro}=115 m^2/g and SA_{meso}=1158 m^2/g, red circles are electrodes with SA_{micro}=636 m^2/g and SA_{meso}=442 m^2/g, and blue triangles for an electrode with SA_{micro}=735 m^2/g and SA_{meso}=1200 m^2/g. Because an electrode with high mesopore surface area and low micropore surface area minimizes the resistance of ion transport, its specific capacitance decays much slower with the increase of the scan rate.

8.3.2 Prediction of EDLC Performance

Figure 8.5 shows variation of the specific capacitance with micropore and mesopore surface areas at three representative scan rates, 1 mV/s, 50 mV/s and 100 mV/s. If charging/discharging is carried out reversibly (i.e., near equilibrium), materials with only micropores would be favored to have high specific capacitance, provided that the micropores are sufficiently large to accommodate the ionic species. Intuitively, we also expect that a large surface area may result in high capacitance because it facilitates more ion adsorption on the electrode surface. However, the equilibrium condition can be hardly satisfied during charging/discharging, and thus maximizing the surface area of the micropores is not necessarily practical for the optimal performance. At a finite charging/discharging rate, mesopores are expected to play an important role because they allow ionic adsorption/desorption at high speed. Besides, it has been recently shown that the presence of mesopores increases the connectivity between micropores and thereby the energy storage[48]. Figure 8.5A shows that, even at a relatively small scan rate (1 mV/s), active carbons with both moderately high micropore and mesopore surface areas achieve the highest capacitance. At a larger scan rate (100 mV/s), the capacitance is mainly determined by mesopores because slow ion transport makes small pores hardly accessible. We can see from Figure 8.5A to Figure 8.5C a clear transition of different contributions to the specific capacitance from micropores to mesopores upon a gradual increase of the scan rate. A balance of specific surface area and ion transport resistance effects explains why the EDLC capacitance cannot be optimized by increasing the micropore or mesopore surface area alone.

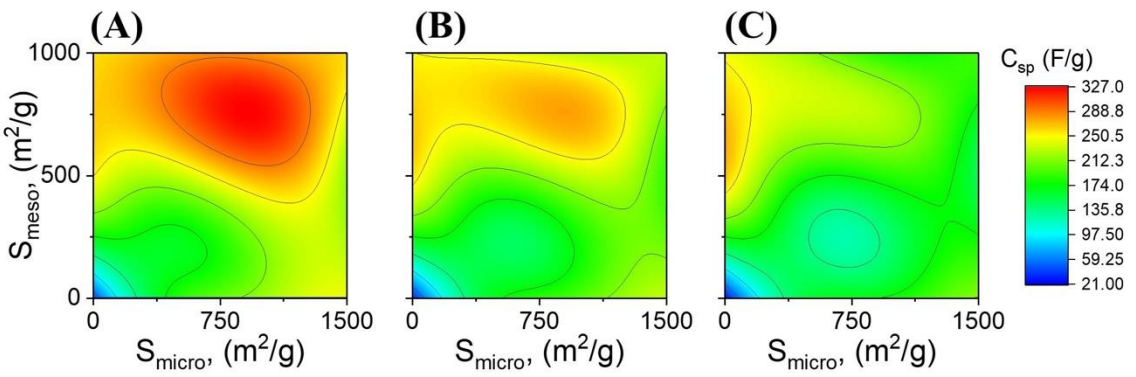

Figure 8.5. Specific capacitance versus the surface areas of micropores and mesopores predicted by the artificial neural network model. The scan rate is (A) 1 mV/s, (B) 50 mV/s and (C) 100 mV/s.

In accordance with experimental observations[49], Figure 8.5 shows that increasing the surface area does not raise the specific capacitance when the total surface area exceeds about 1500 m^2/g. The decline in capacitance at large surface areas might result from the limiting thickness of the micropore surfaces[49]. At large scan rates, the EDLC operates further away from equilibrium leading to a drastic reduction of the capacitance. Compared to that for mesopores, the capacitance affiliated with micropores is more sensitive to the scan rate. It is worth noting that the capacitance does not always rise with the micropore surface area even at a relatively low scan rate (1 mV/s). Instead, it reaches a plateau and then decreases at extremely high micropore surface area when the micropores are no more accessible to ion adsorption[49]. As the surface area per mass increases, the thickness of carbon flakes would decrease, regardless of the pore size, leading to interactions between electrolytes in neighboring pores [48].

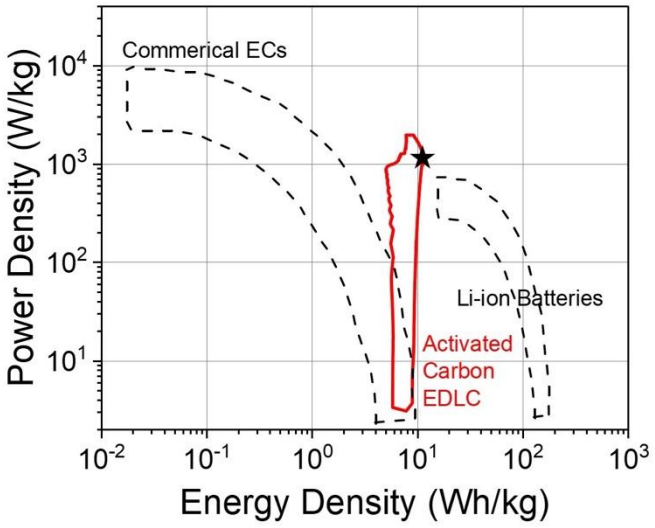

Figure 8.6. The Ragone plot for activated-carbon EDLCs. Here the region enclosed by the red solid line is predicted by the artificial neural network (ANN). The symbol star represents an electrode with SA_{micro}=920 m^2/g, SA_{meso}=770 m^2/g at the scan rate of 1 mV/s. The regions enclosed by the dashed lines correspond to the commercial electrical capacitors (EC) and lithium ion batteries.

With the scan rate and the surface areas of micropore and mesopore as the input, ANN allows us to construct the Ragone plot, i.e., relations between the energy density and power density of energy storage devices. As shown in Figure 8.6, activated-carbon-based EDLCs have a great potential to improve the performance of commercial electrical capacitors (EC). By optimizing the micropore and mesopore surface areas, we may have an EDLC with the energy density comparable to that of the Li-ion battery while exhibiting a much higher power density. The ANN model predicts that the energy density can be maximized with SA_{micro}=920 m^2/g and SA_{meso}=770 m^2/g for the micro- and mesopore surface areas, respectively. While the optimal characteristics of the carbon electrode depends on the scan rate, the EDLC performance may be further improved by chemical modifications such as doping with nitrogen or oxygen.

8.4 Conclusions

In this work, we establish quantitative correlations between the structural features of activated carbon electrodes and the EDLC performance using physics-informed machine-learning (ML) models: generalized linear regression (GLR), support vector machine (SVM), random forest (RF), and artificial neural network (ANN). Among the four ML algorithms tested, we find that ANN results in the best performance and is able to capture the capacitance dependence on micropore and mesopore surface areas over a broad range of the scan rate. Besides, it can be used to construct a Ragone plot that predicts the desirable features of activated carbons that lead to the highest energy and power density. While similar correlations were established before for EDLC capacitance, we provide the first comprehensive analysis of the EDLC performance directly relevant to practical applications. While the statistics of ML correlations can be further improved with more experimental data, we hope that the new insights gained from this work would be useful to guide future studies in design and synthesis of carbon electrodes to improve the supercapacitor performance.

Bibliography

1. Berrueta, A., et al., *Supercapacitors: Electrical Characteristics, Modeling, Applications, and Future Trends.* Ieee Access, 2019. **7**: p. 50869-50896.

2. Simon, P. and Y. Gogotsi, *Capacitive energy storage in nanostructured carbon-electrolyte systems.* Acc Chem Res, 2013. **46**(5): p. 1094-103.

3. Stoller, M.D. and R.S. Ruoff, *Best practice methods for determining an electrode material's performance for ultracapacitors.* Energy & Environmental Science, 2010. **3**(9): p. 1294-1301.

4. Ghosh, A. and Y.H. Lee, *Carbon-based electrochemical capacitors.* ChemSusChem, 2012. **5**(3): p. 480-99.

5. Zhai, Y., et al., *Carbon materials for chemical capacitive energy storage.* Adv Mater, 2011. **23**(42): p. 4828-50.

6. Zhan, C., et al., *Computational Insights into Materials and Interfaces for Capacitive Energy Storage.* Adv Sci (Weinh), 2017. **4**(7): p. 1700059.

7. Hasegawa, G., et al., *New Insights into the Relationship between Micropore Properties, Ionic Sizes, and Electric Double-Layer Capacitance in Monolithic Carbon Electrodes.* Journal of Physical Chemistry C, 2012. **116**(50): p. 26197-26203.

8. Eliad, L., et al., *Ion Sieving Effects in the Electrical Double Layer of Porous Carbon Electrodes: Estimating Effective Ion Size in Electrolytic Solutions.* The Journal of Physical Chemistry B, 2001. **105**(29): p. 6880-6887.

9. Lee, G.J. and S.I. Pyun, *Theoretical approach to ion penetration into pores with pore fractal characteristics during double-layer charging/discharging on a porous carbon electrode.* Langmuir, 2006. **22**(25): p. 10659-65.

10. Garcia-Gomez, A., et al., *Constant capacitance in nanopores of carbon monoliths.* Phys Chem Chem Phys, 2015. **17**(24): p. 15687-90.

11. Chmiola, J., et al., *Anomalous increase in carbon capacitance at pore sizes less than 1 nanometer.* Science, 2006. **313**(5794): p. 1760-3.

12. Feng, G. and P.T. Cummings, *Supercapacitor Capacitance Exhibits Oscillatory Behavior as a Function of Nanopore Size.* Journal of Physical Chemistry Letters, 2011. **2**(22): p. 2859-2864.

13. Jiang, D.E., Z. Jin, and J. Wu, *Oscillation of capacitance inside nanopores.* Nano Lett, 2011. **11**(12): p. 5373-7.

14. Stoeckli, F. and T.A. Centeno, *Optimization of the characterization of porous carbons for supercapacitors.* Journal of Materials Chemistry A, 2013. **1**(23): p. 6865-6873.

15. Kondrat, S., et al., *A superionic state in nano-porous double-layer capacitors: insights from Monte Carlo simulations.* Phys Chem Chem Phys, 2011. **13**(23): p. 11359-66.

16. Newman, J.S., *Electrochemical systems*. 3rd ed. ed, ed. K.E. Thomas-Alyea. 2004, Hoboken, N.J: J. Wiley.

17. Breitsprecher, K., C. Holm, and S. Kondrat, *Charge Me Slowly, I Am in a Hurry: Optimizing Charge-Discharge Cycles in Nanoporous Supercapacitors.* ACS Nano, 2018. **12**(10): p. 9733-9741.

18. Severson, K.A., et al., *Data-driven prediction of battery cycle life before capacity degradation.* Nature Energy, 2019. **4**(5): p. 383-391.

19. Wu, B., et al., *Application of artificial neural networks in design of lithium-ion batteries.* Journal of Power Sources, 2018. **395**: p. 128-136.

20. Pardakhti, M., et al., *Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs).* ACS Comb Sci, 2017. **19**(10): p. 640-645.

21. Zhang, Z., et al., *Prediction of Carbon Dioxide Adsorption via Deep Learning.* Angew Chem Int Ed Engl, 2019. **58**(1): p. 259-263.

22. Dureckova, H., et al., *Robust Machine Learning Models for Predicting High CO2 Working Capacity and CO2/H2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture.* The Journal of Physical Chemistry C, 2019. **123**(7): p. 4133-4139.

23. Zhu, S., et al., *Artificial neural network enabled capacitance prediction for carbon-based supercapacitors.* Materials Letters, 2018. **233**: p. 294-297.

24. Su, H., et al., *Predicting the Capacitance of Carbon-based Electric Double Layer Capacitors by Machine Learning.* Nanoscale Advances, 2019.

25. Dawson-Elli, N., et al., *Data Science Approaches for Electrochemical Engineers: An Introduction through Surrogate Model Development for Lithium-Ion Batteries.* Journal of the Electrochemical Society, 2018. **165**(2): p. A1-A15.

26. Farsi, H. and F. Gobal, *Artificial neural network simulator for supercapacitor performance prediction.* Computational Materials Science, 2007. **39**(3): p. 678-683.

27. Li, Y.T., et al., *Hierarchical porous active carbon from fallen leaves by synergy of K2CO3 and their supercapacitor performance.* Journal of Power Sources, 2015. **299**: p. 519-528.

28. Zhang, D.D., et al., *Scalable synthesis of hierarchical macropore-rich activated carbon microspheres assembled by carbon nanoparticles for high rate performance supercapacitors.* Journal of Power Sources, 2017. **342**: p. 363-370.

29. Zhang, J.T., et al., *Preparation of activated carbon from waste Camellia oleifera shell for supercapacitor application.* Journal of Solid State Electrochemistry, 2012. **16**(6): p. 2179-2186.

30. Jiang, L., et al., *High rate performance activated carbons prepared from ginkgo shells for electrochemical supercapacitors.* Carbon, 2013. **56**: p. 146-154.

31. Yang, W., et al., *Template-free synthesis of ultrathin porous carbon shell with excellent conductivity for high-rate supercapacitors.* Carbon, 2017. **111**: p. 419-427.

32. Jiang, L.L., et al., *Construction of nitrogen-doped porous carbon buildings using interconnected ultra-small carbon nanosheets for ultra-high rate supercapacitors.* Journal of Materials Chemistry A, 2016. **4**(29): p. 11388-11396.

33. Wu, H., et al., *The effect of activation technology on the electrochemical performance of calcium carbide skeleton carbon.* Journal of Solid State Electrochemistry, 2012. **16**(9): p. 2941-2947.

34. Freedman, D.A., *Statistical Models*. 2009.

35. McCullagh, P. and J.A. Nelder, *Generalized Linear Models*. 2019.

36. Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics. 2009.

37. Smola, A.J. and B. Schölkopf, *A tutorial on support vector regression.* Statistics and Computing, 2004. **14**(3): p. 199-222.

38. Tin Kam, H., *The random subspace method for constructing decision forests.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(8): p. 832-844.

39. Rokach, L. and O. Maimon, *Data Mining with Decision Trees*. Series in Machine Perception and Artificial Intelligence. 2014.

40. Waszczyszyn, Z. *Fundamentals of Artificial Neural Networks*. 1999. Vienna: Springer Vienna.

41. Towell, G.G. and J.W. Shavlik, *Knowledge-based artificial neural networks.* Artificial Intelligence, 1994. **70**(1-2): p. 119-165.

42. Xiong, H.Y., Y. Barash, and B.J. Frey, *Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context.* Bioinformatics, 2011. **27**(18): p. 2554-2562.

43. Shi, H., *Activated carbons and double layer capacitance.* Electrochimica Acta, 1996. **41**(10): p. 1633-1639.

44. Lu, Y., et al., *What are the practical limits for the specific surface area and capacitance of bulk sp2 carbon materials?* Science China Chemistry, 2015. **59**(2): p. 225-230.

45. Dyatkin, B., et al., *Electrolyte cation length influences electrosorption and dynamics in porous carbon supercapacitors.* Electrochimica Acta, 2018. **283**: p. 882-893.

46. Allagui, A., et al., *Reevaluation of Performance of Electric Double-layer Capacitors from Constant-current Charge/Discharge and Cyclic Voltammetry.* Sci Rep, 2016. **6**: p. 38568.

47. Pak, A.J. and G.S. Hwang, *Charging Rate Dependence of Ion Migration and Stagnation in Ionic-Liquid-Filled Carbon Nanopores.* Journal of Physical Chemistry C, 2016. **120**(43): p. 24560-24567.

48. Vasilyev, O.A., A.A. Kornyshev, and S. Kondrat, *Connections Matter: On the Importance of Pore Percolation for Nanoporous Supercapacitors.* ACS Appl. Energy Mater., 2019. **doi.org/10.1021/acsaem.9b01069**: p. 4523-4527.

49. Barbieri, O., et al., *Capacitance limits of high surface area activated carbons for double layer capacitors.* Carbon, 2005. **43**(6): p. 1303-1310.

Chapter 9. A Data-Driven Approach to Understanding the In-Operando Performance of Heteroatom-Doped Carbon Electrodes

Doping with heteroatoms such as nitrogen and oxygen has been widely practiced to improve the capacitance of carbon electrodes for supercapacitor. However, the role of different heteroatoms and their local atomic configurations on the supercapacitor performance remains elusive, which hampers the rational design of carbon electrodes to achieve high energy density and power density. In this chapter, machine learning models are applied to interpret how the surface chemistry affects the *in-operando* behavior of various carbon electrodes with different structural features such as the specific surface areas of micro- and mesopores. Quantitative descriptions have been established to predict how the configurations of nitrogen-doping and oxygen-doping influence the capacitance and retention rate. The machine learning models provide insights into the design and possible routes to the synthesis of nitrogen and oxygen co-doped carbon electrodes that maximize the energy storage capacity.

9.1 Introduction

Supercapacitors have been recognized as a viable approach for energy storage with the advantages of high power density and long lifespan in comparison to alternative methods such as batteries and fuel cells.[1-3] However, current applications of supercapacitors are often impaired by their low energy density. One focal point of the ongoing supercapacitor research has thus been directed at improving the energy density while maintaining high power density.

Porous carbons have been widely used as the electrode material for supercapacitors because of high porosity, large surface area, good conductivity, and modest cost.[4, 5] As the energy is stored through the formation of electrical double layer (EDL) at the electrode-electrolyte interface, the more surface area a carbon electrode has, the more energy can be stored in supercapacitor. While the capacitance, and thus the device performance, can be in general improved by increasing the surface area, the specific capacitance reaches a plateau when the specific surface area is larger than about 1500 m^2/g. The asymptotic behavior arises from the reduction of the electrode conductivity and the interference of ion adsorption in neighboring pores.[6,7] Besides, the increase of surface area through adding more micropores limits ion accessibility and transport, which becomes more substantial when the ionic charge needs to be extracted at a fast rate. To improve the retention rate at high-speed charging/discharging processes, heteroatoms such as nitrogen, oxygen and sulfur, have been routinely introduced into the carbon electrodes by using various doping procedures. The heteroatom doping improves the electrical conductivity of the carbon electrodes and electrolyte wettability. The improved capacitive performance may also be attributed to faradaic reactions and to the increased density of states at the electrode surface.[6, 7] Whereas much progress has been made explaining the heteroatom effects on the equilibrium properties of carbon electrodes, it remains unclear how they boost the supercapacitor performance under realistic conditions far away from equilibrium.

Physics-based models have been used to elucidate the origin of improved capacitance due to heteroatom doping.[8, 9] However, existing models are unable to

make a quantitative prediction of the capacitance under conditions pertinent to the device performance or electrochemical characterization. Unlike equilibrium properties as typically described in physical models, the capacitance most relevant to the device performance is a dynamic variable depending on the charging/discharging kinetics and, for electrochemical characterization experiments, on the scan rate and electrical current. Because of limitations in mass transport and reaction kinetics, the dynamic capacitance decreases with the increase of charging/discharging rate. The missing link between the charging behavior and the microscopic details of ion transport in porous electrodes at *in-operando* conditions hampers the direct applications of physics-based methods to the rational design of electrode materials.

In contrast to physics-based modelling, machine learning models allow us to establish quantitative correlations between the important features of electrode materials and their performance based on extensive data measured experimentally.[10-12] The data-driven approach provides an alternative route to theoretical modeling for understanding the device performance at conditions far away from equilibrium and for the inverse design of materials toward tailored applications. Previously, several studies have been reported using machine learning models to study the supercapacitor performance.[13, 14] Although good correlations have been established between the capacitance and a few selected features of pristine carbon electrodes, the interplay of the surface chemistry with the structural features of carbon electrodes remains poorly understood. In this work, we employ machine learning models to unveil the relation between the dynamic performance of carbon electrodes and heteroatom doping (nitrogen-

and oxygen-doping) based on extensive experimental data from the literature. The structural features of the carbon materials are characterized in terms of the micropore and mesopore surface areas similar to our previous work.[15] To confine the parameter space, we consider the supercapacitor performance only for carbon electrodes in a specific aqueous electrolyte (6 M KOH) as commonly used in experiments. While similar machine learning models are used in this study, our focus here is to explore how the chemical composition of carbon electrodes affects the supercapacitor performance. A special attend is given to the supercapacitor performance at high scan rates of cyclic voltammetry (CV) measurements as such condition is most closely related to real-life applications of supercapacitors. With extra descriptors, the machine learning models can not only well reproduce the results for pristine carbon electrodes, but also allow us to map a much larger chemical space with an explicit consideration of different ways of chemical doping.

9.2 Methods and Models

All data used in this work is collected from the literature.[7, 16-30] Although experimental studies have extensively been reported on the capacitance of carbon electrodes in aqueous electrolytes, only a few of them provide information with detailed pore characterization and chemical composition. After extracting the raw data from CV measurements, we calculate the capacitance, energy density and power density of electrodes by using the standard equations (eqs [9.1] ~ [9.3]). The specific integral capacitance is given by

$$C_{sp} = \frac{\int_{V_{start}}^{V_{end}} i(V)dV}{2\upsilon m\Delta V} = \frac{\bar{I}\frac{\Delta V}{\upsilon}}{m\Delta V} = \frac{\bar{I}\Delta t}{m\Delta V} \tag{9.1}$$

where υ is the scan rate (V/s), i is the electrical current, m is the electrode mass, ΔV is the potential window, $\bar{I}$ is the average current, $\Delta t=\Delta V/\upsilon$ is charging/discharging time, and C_{sp} stands for the specific integral capacitance of the electrode in a three-electrode system. Following the experimental literature, we define the energy density as

$$E = \frac{C_{cell}\Delta V^2}{2} = \frac{C_{sp}\Delta V^2}{8} = \frac{\bar{I}\Delta t\Delta V}{8m} \tag{9.2}$$

where C_{cell} corresponds to the specific capacitance of a supercapacitor with two symmetric electrodes. Accordingly, the power density is calculated from

$$P = \frac{E}{\Delta t} = \frac{\bar{I}\Delta V}{8m} \tag{9.3}$$

It should be emphasized that none of the quantities calculated above correspond to the equilibrium values. Nevertheless, these quantities are most relevant to evaluating the performance of supercapacitors under practical conditions.

The dataset also includes key information of the electrode materials including surface areas of micropores (pore diameters less than 2 nm) and mesopores (pore diameters between 2 nm and 50 nm) and the detailed percentages of the doping configurations for different heteroatoms. All data used in this work is provided in Table 9.1.

Table 9.1. Input and output data used in different machine learning models.

#	C_{sp} (F/g)	*E* (Wh /kg)	*P* (kW /kg)	*v* (mV/s)	SA_{micro} (m^2/g)	SA_{meso} (m^2/g)	O (%)	N-5 (%)	N-6 (%)	N-Q (%)	N-other (%)
1	0	0	0	1	0	0	0	0	0	0	0
2	0	0	0	300	0	0	0	0	0	0	0
3	0	0	0	500	0	0	0	0	0	0	0
4	0	0	0	1	0	0	17	4	3	7	2
5	0	0	0	300	0	0	17	4	3	7	2
6	0	0	0	500	0	0	17	4	3	7	2
7	0	0	0	1	0	0	9	2	2	3	1
8	0	0	0	300	0	0	9	2	2	3	1
9	0	0	0	500	0	0	9	2	2	3	1
10	180	7	0	1	120	216	0	0	0	0	0
11	172	8	0	2	120	216	0	0	0	0	0
12	166	8	0	5	120	216	0	0	0	0	0
13	155	7	1	10	120	216	0	0	0	0	0
14	212	6	1	1	107	315	0	0	0	0	0
15	202	5	2	2	107	315	0	0	0	0	0
16	184	6	0	5	107	315	0	0	0	0	0
17	173	6	0	10	107	315	0	0	0	0	0
18	277	4	1	1	153	553	0	0	0	0	0
19	260	6	0	2	153	553	0	0	0	0	0
20	230	4	1	5	153	553	0	0	0	0	0
21	198	9	0	10	153	553	0	0	0	0	0
22	280	7	3	1	200	900	0	0	0	0	0
23	274	8	0	2	200	900	0	0	0	0	0
24	265	7	0	5	200	900	0	0	0	0	0
25	250	7	0	10	200	900	0	0	0	0	0
26	224	6	0	2	735	1200	0	0	0	0	0
27	203	5	1	5	735	1200	0	0	0	0	0
28	190	4	1	10	735	1200	0	0	0	0	0
29	176	6	0	20	735	1200	0	0	0	0	0
30	144	6	1	50	735	1200	0	0	0	0	0
31	114	6	2	100	735	1200	0	0	0	0	0
32	143	8	0	10	417	645	6	3	3	2	0
33	118	7	1	50	417	645	6	3	3	2	0
34	97	6	2	100	417	645	6	3	3	2	0
35	36	6	0	400	417	645	6	3	3	2	0
36	164	5	1	10	219	840	6	2	2	1	0
37	141	5	2	50	219	840	6	2	2	1	0
38	123	4	0	100	219	840	6	2	2	1	0
39	63	3	0	400	219	840	6	2	2	1	0
40	140	2	0	10	173	944	5	2	1	1	0

41	124	2	0	50	173	944	5	2	1	1	0
42	110	2	0	100	173	944	5	2	1	1	0
43	65	2	1	400	173	944	5	2	1	1	0
44	276	1	1	5	1283	764	8	0	0	0	0
45	250	1	2	10	1283	764	8	0	0	0	0
46	239	9	0	20	1283	764	8	0	0	0	0
47	222	8	0	50	1283	764	8	0	0	0	0
48	204	8	0	100	1283	764	8	0	0	0	0
49	190	8	1	150	1283	764	8	0	0	0	0
50	179	8	1	200	1283	764	8	0	0	0	0
51	168	8	3	250	1283	764	8	0	0	0	0
52	286	7	5	5	1574	187	12	1	1	1	0
53	268	7	12	10	1574	187	12	1	1	1	0
54	249	11	0	20	1574	187	12	1	1	1	0
55	208	11	0	50	1574	187	12	1	1	1	0
56	160	10	0	100	1574	187	12	1	1	1	0
57	131	10	1	150	1574	187	12	1	1	1	0
58	282	10	2	5	1556	205	12	1	1	1	1
59	282	10	3	5	1320	621	11	0	0	0	0
60	205	9	7	5	1072	661	8	0	0	0	0
61	189	9	16	5	1990	879	0	0	0	0	0
62	232	6	0	5	636	442	0	0	0	0	0
63	223	6	0	10	636	442	0	0	0	0	0
64	202	6	0	20	636	442	0	0	0	0	0
65	185	5	0	50	636	442	0	0	0	0	0
66	155	7	0	100	636	442	0	0	0	0	0
67	185	7	0	5	713	290	0	0	0	0	0
68	294	6	0	5	971	282	8	1	1	1	0
69	277	6	0	5	633	1394	5	0	0	0	0
70	332	10	0	5	1227	1170	8	1	1	1	0
71	316	9	0	10	1227	1170	8	1	1	1	0
72	298	8	0	20	1227	1170	8	1	1	1	0
73	261	7	0	50	1227	1170	8	1	1	1	0
74	220	10	0	100	1227	1170	8	1	1	1	0
75	205	10	0	200	1227	1170	8	1	1	1	0
76	117	9	0	2	0	24	0	0	0	0	0
77	79	9	0	5	0	24	0	0	0	0	0
78	68	5	0	10	0	24	0	0	0	0	0
79	61	2	0	20	0	24	0	0	0	0	0
80	53	1	0	50	0	24	0	0	0	0	0
81	47	0	0	100	0	24	0	0	0	0	0
82	41	3	0	200	0	24	0	0	0	0	0
83	31	3	0	500	0	24	0	0	0	0	0
84	258	3	0	2	115	1158	0	0	0	0	0

85	244	3	0	5	115	1158	0	0	0	0	0
86	238	4	0	10	115	1158	0	0	0	0	0
87	232	4	0	20	115	1158	0	0	0	0	0
88	225	5	0	50	115	1158	0	0	0	0	0
89	217	4	1	100	115	1158	0	0	0	0	0
90	207	3	1	200	115	1158	0	0	0	0	0
91	187	1	2	500	115	1158	0	0	0	0	0
92	326	6	0	2	327	1280	7	1	1	2	0
93	309	5	1	5	327	1280	7	1	1	2	0
94	302	4	2	10	327	1280	7	1	1	2	0
95	294	2	3	20	327	1280	7	1	1	2	0
96	286	5	0	50	327	1280	7	1	1	2	0
97	278	4	1	100	327	1280	7	1	1	2	0
98	269	4	1	200	327	1280	7	1	1	2	0
99	251	2	3	500	327	1280	7	1	1	2	0
100	142	10	0	10	250	52	12	0	0	0	0
101	220	10	0	5	275	349	8	0	0	0	0
102	207	12	0	10	275	349	8	0	0	0	0
103	180	11	0	50	275	349	8	0	0	0	0
104	161	10	1	100	275	349	8	0	0	0	0
105	183	9	2	5	437	10	0	0	0	0	0
106	162	8	3	50	437	10	0	0	0	0	0
107	159	7	5	100	437	10	0	0	0	0	0
108	222	8	0	5	501	25	0	0	0	0	0
109	192	11	0	50	501	25	0	0	0	0	0
110	183	10	0	100	501	25	0	0	0	0	0
111	159	9	1	5	579	83	0	0	0	0	0
112	140	8	1	50	579	83	0	0	0	0	0
113	137	7	2	100	579	83	0	0	0	0	0
114	171	9	0	10	457	126	0	0	0	0	0
115	101	6	0	100	457	126	0	0	0	0	0
116	161	8	0	10	429	188	0	0	0	0	0
117	116	7	0	100	429	188	0	0	0	0	0
118	254	8	0	10	1118	504	6	0	0	0	0
119	203	10	0	100	1118	504	6	0	0	0	0
120	116	9	0	30	454	98	14	3	3	0	0
121	130	9	1	30	160	497	15	2	2	0	0
122	218	7	1	10	259	633	17	1	2	0	0
123	204	6	2	20	259	633	17	1	2	0	0
124	190	8	0	30	259	633	17	1	2	0	0
125	169	8	0	50	259	633	17	1	2	0	0
126	122	7	0	100	259	633	17	1	2	0	0
127	220	7	0	10	1052	253	8	1	1	1	1
128	286	7	0	5	1462	327	5	1	1	2	2

129	271	7	0	10	1462	327	5	1	1	2	2
130	253	3	1	20	1462	327	5	1	1	2	2
131	215	4	1	50	1462	327	5	1	1	2	2
132	170	3	0	100	1462	327	5	1	1	2	2
133	243	6	0	10	1133	279	8	1	1	1	1
134	218	6	1	10	897	233	8	1	1	1	1
135	211	5	2	10	567	293	8	1	0	1	1
136	204	4	0	10	824	298	7	2	2	2	2
137	210	5	0	10	1307	258	6	1	1	1	1
138	193	8	0	10	1099	182	6	1	1	1	1
139	132	7	1	10	1412	66	8	3	3	2	1
140	56	7	1	20	1412	66	8	3	3	2	1
141	16	6	1	50	1412	66	8	3	3	2	1
142	6	4	2	100	1412	66	8	3	3	2	1
143	91	5	0	10	889	51	11	3	2	2	2
144	79	8	0	10	952	74	8	3	2	2	1
145	93	7	0	10	1260	41	10	2	2	2	2
146	93	6	1	10	752	16	13	4	3	2	1
147	107	6	2	10	2316	115	5	1	1	1	1
148	110	2	0	10	1534	63	7	1	1	1	1
149	241	9	0	10	1809	319	12	3	2	3	0
150	308	9	1	5	1976	403	8	4	2	1	0
151	287	8	1	10	1976	403	8	4	2	1	0
152	264	7	2	20	1976	403	8	4	2	1	0
153	235	5	3	50	1976	403	8	4	2	1	0
154	199	10	0	100	1976	403	8	4	2	1	0
155	252	9	0	10	1552	369	8	3	2	1	0
156	170	8	1	10	562	128	11	3	3	3	0
157	227	8	1	10	779	1003	8	1	1	2	0
158	211	7	3	10	443	804	10	1	0	1	0
159	94	7	4	50	517	61	5	2	3	7	2
160	128	6	4	50	640	184	6	3	3	2	1
161	76	6	5	50	563	120	8	2	1	1	1
162	178	10	0	20	680	641	8	3	2	1	1
163	163	9	0	50	680	641	8	3	2	1	1
164	130	9	1	100	680	641	8	3	2	1	1
165	64	7	1	50	0	1082	0	0	0	0	0
166	269	6	2	10	1590	1030	15	0	0	0	0
167	260	5	2	20	1590	1030	15	0	0	0	0
168	230	10	0	50	1590	1030	15	0	0	0	0
169	197	10	0	100	1590	1030	15	0	0	0	0
170	136	7	0	200	1590	1030	15	0	0	0	0

Note: 10-25[16]; 26-31[7]; 32-43[17]; 44-60[18]; 61-67[19]; 68-75[20]; 76-99[21]; 100-104[22]; 105-113[23]; 114-119[24]; 120-126[25]; 127-138[26]; 139-148[27]; 149-

158[28]; 159-164[29]; 165-170[30];

The selection of features or parameters is key for the application of machine learning models. Redundant input parameters would add unnecessary burden for the amount of data required to train the machine learning models without improving the accuracy, while missing essential features would mislead the training and give problematic predictions. In this work, we take the structural properties of carbon materials, the chemical compositions of the electrode surfaces, and the scan rate in CV measurements as the input parameters. As all experimental data were measured in 6M KOH aqueous solution with the three-electrode configuration within the same range of the potential window (1V),[7, 16, 18-31] we need not consider parameters such as the electrolyte composition, the solution pH, and the potential window that otherwise must be included. The supercapacitor performance is measured in terms of the capacitance and power density. Since all the data are affiliated with the same potential window, the energy density is simply proportional to the capacitance. Table 9.2 lists all input and output parameters employed in this work.

Table 9.2. Summary of the input and output parameters used for training machine learning models.

Input	Output
SA_{micro} (m^2/g)	Specific Capacitance (F/g)
SA_{meso} (m^2/g)	
Scan Rate (mV/s)	
Oxygen (at %)	
Pyrrolic Nitrogen – N-5 (at %)	Power Density (kW/kg)
Pyridinic Nitrogen – N-6 (at %)	
Quaternary Nitrogen – N-Q (at %)	
Other Nitrogen type (at %)	

We use the surface areas of micropores and mesopores characterized by N_2 adsorption isotherm as the primary features of carbon electrodes. Similar features were used in our previous work for pristine carbon electrodes.[15] The micropore surface area provides a reasonable description of the accessible area for electrosorption of hydrated ions from the KOH solution because the pore size is comparable to the ion diameters. All experimental data for the percentage of heteroatom doping were obtained from X-ray photoelectron spectroscopy (XPS) measurements. The XPS data are highly reliable for characterizing the surface composition of the heteroatoms doped at the carbon surface.[32] For nitrogen doping, several stable atomic configurations have been reported including pyrrolic nitrogen (N-5), pyridinic nitrogen (N-6) and quaternary nitrogen (N-Q). As different atomic configurations of nitrogen doping could play different roles in determining the supercapacitor behavior, we describe the surface chemical composition in terms of the percentage of each atomic configuration as an input parameter. Oxygen atoms are often co-doped on the carbon surface along with the nitrogen atoms because they may preexist in the nitrogen precursors and could be introduced to the carbon surface during the activation procedure. While the nitrogen-doping effects are relatively well understood, the role of oxygen atoms on the supercapacitor performance in base electrolytes is not clear. As a matter of fact, the local configurations of oxygens have been rarely reported in the literature. In this work, we consider only the total atomic percentage of oxygen atoms doped at the carbon surface as the input parameter.

To correlate the experimental data for the specific capacitance and the power density in terms of 8 input parameters (features) as listed in Table 9.2, we have tested

four different machine learning models – generalized linear regression (GLR), support vector machine (SVM), random forest (RF), and artificial neural network (ANN). These machine-learning models were selected because they are commonly used in supervised learning and quantitative analysis. The basic ideas underlying these machine-learning models have been discussed in our previous work[15]. In this work, four different machine learning models are tested to correlate supercapacitor performance with the structural features and surface compositions (Table 9.1). The *k-fold* validation is used to optimize the parameters of all four machine learning models with the *k* value set as 17. The percentages of training, validation and test data are 70%, 15% and 15%, respectively. For correlation with *generalized linear regression* (GLR), the model contains an intercept term, a linear term, and a squared term for each input to correlate the capacitance, while one addition term consisting of all products of the pairs of distinct inputs is used to correlate the power density. The second order polynomial kernel function is used in application of *support vector machine* (SVM) to correlating both the specific capacitance and power density. In the *random forest* (RF) model, we used 40 classification trees to correlate the capacitance, and 70 classification trees, with a minimum of 5 observations per tree leaf, to correlate the power density. The *artificial neural network* (ANN) includes a layer of six neurons as hidden layer with a hyperbolic tangent sigmoid transfer function. The backpropagation employs the Bayesian regularization, which makes the ANN more robust and more generalized without overfitting.[33, 34]

All parameters for the machine learning models are optimized by *k-fold* validation with the *k* value set as 17. The coefficient of determination (R^2) and the root-mean-square

error (RMSE) are used to benchmark the performance of different machine learning models:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - u_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{u}\right)^2} \tag{9.4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - u_i\right)^2}{n}} \tag{9.5}$$

where n, y_i, u_i and $\bar{u}$ are, respectively, the number of data points, the value calculated from the correlation, the experimental value, and the average of the corresponding experimental quantities.

9.3 Results and Discussion

9.3.1 Model Evaluation

We first calibrated the machine learning (ML) models considered in this work. As shown in Figure 9.1 and Figure 9.2, ANN has the best overall performance (the highest R^2 and lowest RMSE) for correlating both the capacitance and power density. ANN correlation performance of capacitance and power density is summarized in Figure 9.3. Therefore, this method is selected to predict the optimal performance of supercapacitors in terms of the pore structure and surface compositions of N/O co-doped carbon electrodes. In comparison to the alternative methods, ANN leads to the outstanding performance because it employs the Bayesian regularization in backpropagation. As only six neurons are used in the hidden layer, ANN provides a robust correlation of the experimental data without overfitting.

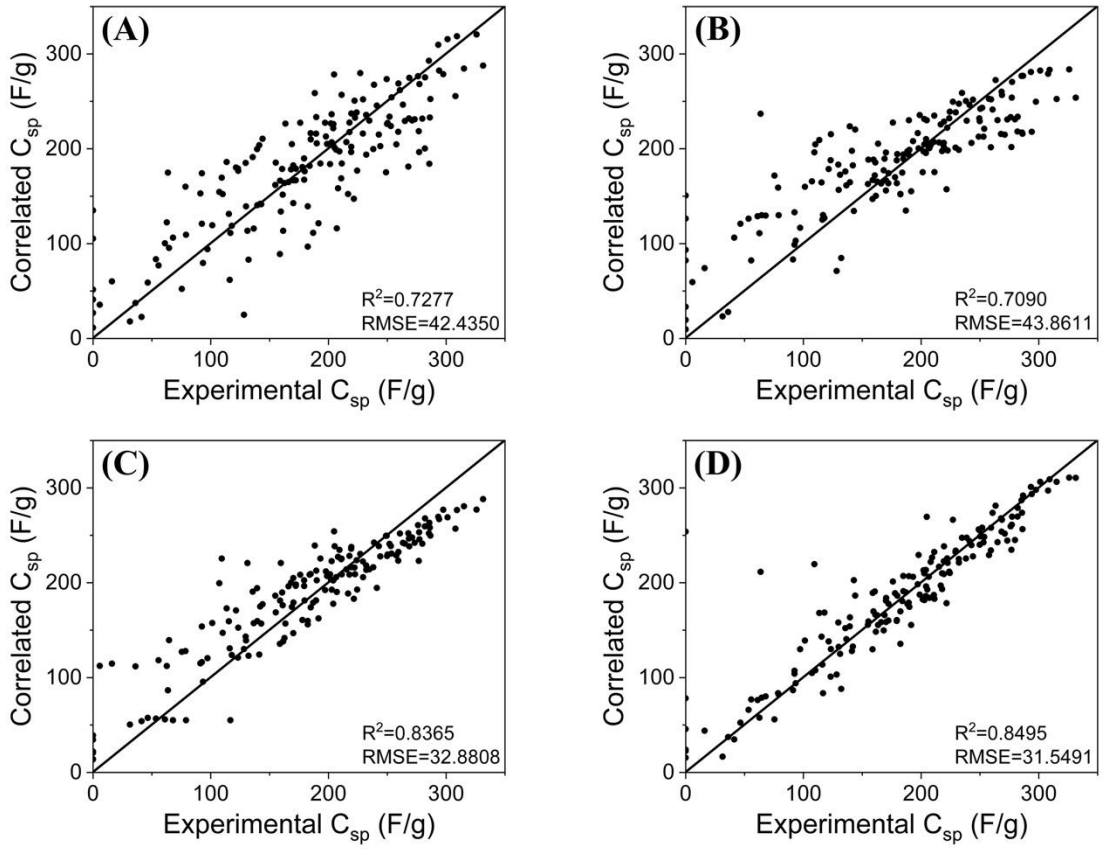

Figure 9.1. Capacitance from experiment and different machine learning models (A) generalized linear regression, (B) support vector machine, (C) random forest and (D) artificial neural network. The diagonal line indicates perfect correlation between experimental and machine learning results.

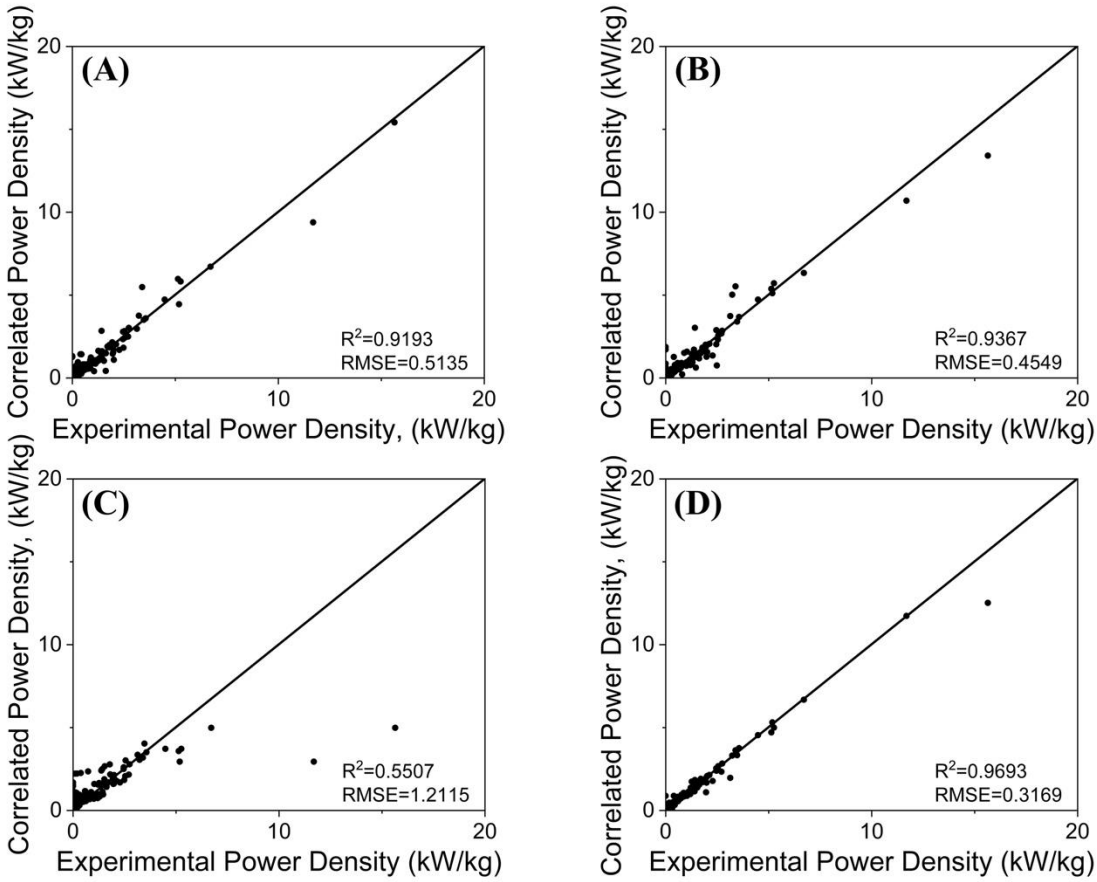

Figure 9.2. Power density from experiment and different machine learning models (A) generalized linear regression, (B) support vector machine, (C) random forest and (D) artificial neural network. The diagonal line indicates perfect correlation between experimental and machine learning results.

With additional data and more input parameters, the four ML models provide better correlations of both the capacitance and power density than our previous work for pristine carbon electrodes.[15] It has been shown that the dependence of capacitance on the percentage of heteroatom doping can be approximately described in terms of a linear function.[35, 36] Therefore, ML models with more complicated regression algorithms are expected to yield a better performance with more data and additional information on the surface chemistry. Among the four ML models tested in this work, only RF results in a

correlation for the power density worse than that for the capacitance (equivalently, the energy density). Compared to that for the energy density, the distribution of experimental data for the power density is much more imbalanced because more than 85% of the data is agglomerated in the range between 0 and 2 kW/kg. As a result, RF would be strongly biased towards the low values of the power density and systematically underestimates the experimental values at high power density.[37]

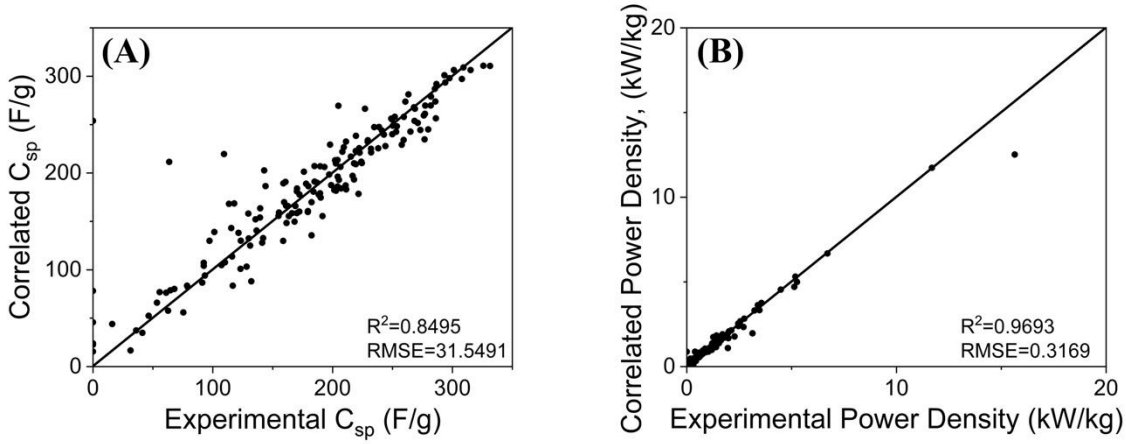

Figure 9.3. A comparison of the experimental and predicted values based on artificial neural network (ANN): (A) specific capacitance (C_{sp}, F/g) and (B) power density (kW/kg). In each panel, the diagonal line represents the perfect correlation between experimental and machine learning results.

According to cyclic voltammetry (CV) measurements, the capacitance depends not only on the structural parameters and surface compositions of carbon materials but also on the scan rate. As shown in Figure 9.4 for a few representative carbon electrodes, the ANN model is able to capture the dependence of the capacitance on the scan rate quantitatively. The excellent correlation indicates that ANN can be used to evaluate the supercapacitor performance at *in-operando* conditions. In general, the apparent capacitance declines as the scan rate increases and the retention depends on both the pore

configuration and the surface chemistry. Interestingly, Figure 9.4 shows that doping heteroatoms on the carbon surface does not always result in an increase of the capacitance. In the next section, we discuss the complicated relationship between capacitance and the structural features as well as the surface composition related to chemical doping.

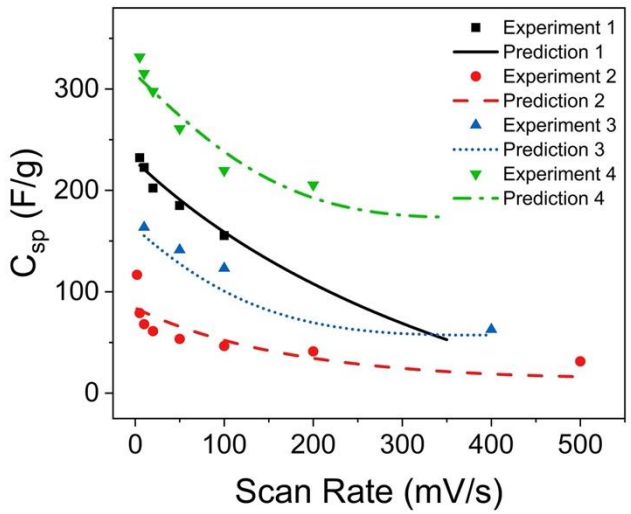

Figure 9.4. The specific capacitance versus the scan rate predicted by the artificial neural network (ANN). Here the black squares are experimental data for a pristine carbon electrode with micropore surface area SA_{micro}=636 m^2/g and mesopore surface area SA_{meso}=442 m^2/g, red circles are for a pristine carbon electrode with SA_{micro}=0 m^2/g and SA_{meso}=24 m^2/g, blue triangles are for a doped carbon electrode with SA_{micro}=219 m^2/g and SA_{meso}=840 m^2/g, and the surface composition of O=5.83 at%, N5=2.1518 at%, N6=1.6510 at%, NQ=1.3972 at%, green inverted triangles are for a doped carbon electrode with SA_{micro}=1227 m^2/g, SA_{meso}=1170 m^2/g, O=7.87 at%, N5=0.9447 at%, N6=1.1139 at%, NQ=0.6651 at%. The experimental data are from the literature.[19-21, 31]

9.3.2 Effect of Nitrogen- and Oxygen-Doping

To unveil how oxygen and nitrogen doping affects the supercapacitor performance, we first consider the correlation between the capacitance and single

heteroatom-doping with different local configurations. Figure 9.5 shows the variation of the capacitance with the surface areas of micropores (<2 nm) and mesopores (>2 nm and <50 nm) at different doping compositions. In all cases, the scan rate is fixed at 1 mV/s. Consistent with our previous work,[15] the capacitance of a pristine carbon does not always increase with the surface area even at a relatively low scan rate. The positive effect of increasing surface area diminishes due to both the reduction of the electrical conductivity of the electrode and the inaccessibility of micropore as well as the interference of the ionic charging behavior in neighboring pores.[38, 39] For an electrode with high micropore surface area, not only does the thickness of carbon flakes decrease, but the accessibility between micropores is limited as well. Introducing mesopores in pristine carbon would improve the connectivity between micropores and reduce the resistance in ion transport, thereby improving the overall capacitance.

Nitrogen-doping with different local configurations shows different effects on the capacitance and power density. At a given scan rate (1 mV/s), both pyrrolic nitrogen (N-5) and pyridinic nitrogen (N-6) shift the maximum capacitance to a higher micropore surface area compared to that for the corresponding pristine carbon. Nitrogen doping increases the maximum capacitance due to the improved surface wettability by the aqueous solution and due to the possible introduction of redox reactions and increased density of states.[40-42] At the same surface concentration of nitrogen atoms, the effects of N-5 and N-6 nitrogen doping are quite different. While N-6 doping leads to a significant increase of the maximum capacitance, the effect due to N-5 doping is modest. It has been reported that pyrrolic nitrogen plays a relatively insignificant role in

improving the density of states of the graphene surface in comparison to N-6 atoms, and thus makes little contribution to the quantum capacitance.[8] The difference between N-5 and N-6 doping may also be attributed to their different contributions to pseudocapacitance, i.e., capacitive effects due to redox reactions. In comparison to N-6, the binding energy between potassium cation and N-5 is probably too large for an effective redox reaction to take place during the charging/discharging cycles.[43] Although N-5 doping makes micropores more accessible to hydrated ions, it is not able to enhance the capacitance as much as N-6 doping because of the difference in surface activity.

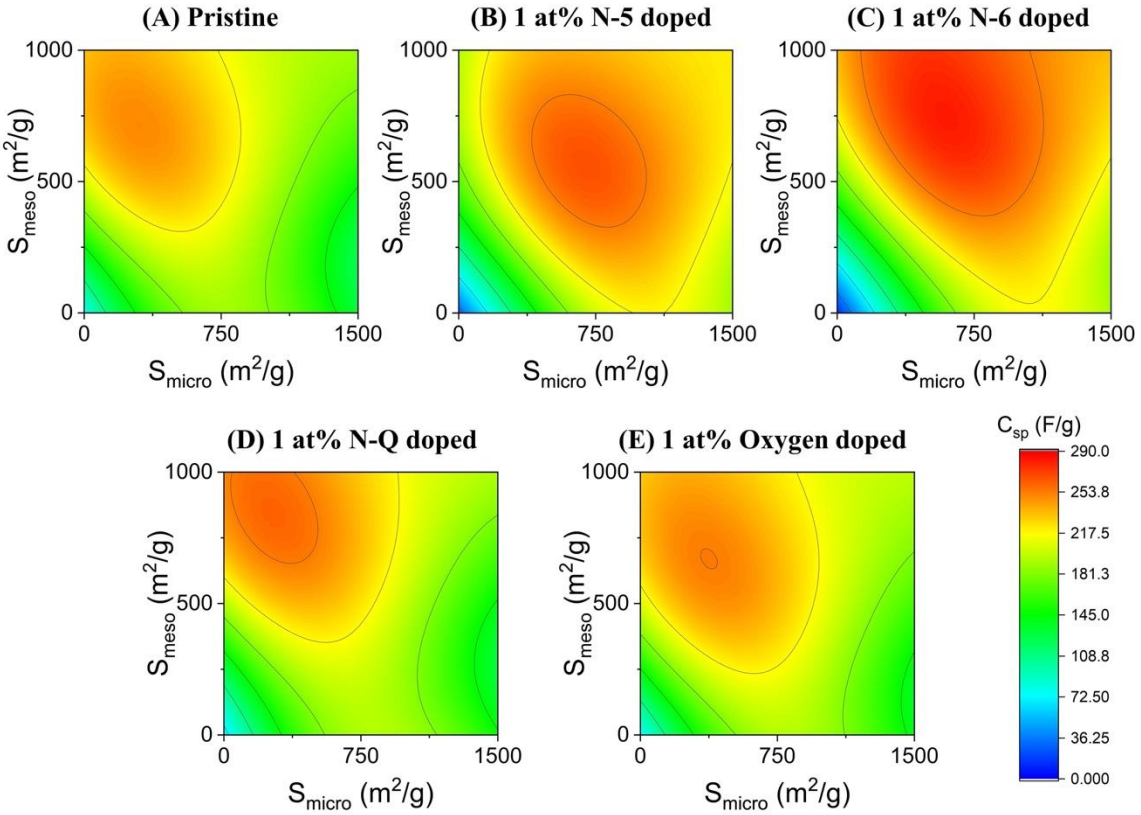

Figure 9.5. Capacitance versus the surface areas of micropores and mesopores predicted by the artificial neural network (ANN) for (A) pristine carbon and carbon doped by (B) 1 at% pyrrolic nitrogen, (C) 1 at% pyridinic nitrogen, (D) 1 at% quaternary nitrogen, and (E) 1 at% oxygen. In all cases, the scan rate is fixed at 1 mV/s.

Another way for nitrogen doping can be achieved by substituting carbon atoms at the electrode surface with quaternary nitrogen (N-Q). It has been shown that doping with quaternary nitrogen may lead to a much higher conductivity for carbon electrodes.[44] However, the general activation method results in defects that could cancel the beneficiary effects of N-Q doping.[44, 45] Figure 9.5D shows that the maximum capacitance increases only slightly when the pristine carbon is doped with 1 at% N-Q nitrogen.

Figure 9.5E shows that oxygen doping has only relatively minor effects on the capacitance of a carbon electrode. Whereas the chemical reaction between proton and quinone-oxygen in an acidic electrolyte solution has been well documented, the local configuration of oxygen atoms is unclear for doping the carbon electrode in a base electrolyte.[46] According to ANN, 1 at% oxygen doping would slightly increase the capacitance in comparison to that for the pristine carbon.

9.3.3 Supercapacitor Performance at Fast Charging/Discharging

Whereas the above discussion was mostly focused on how nitrogen- and oxygen-doping would influence the energy storage near the equilibrium condition (*viz*., at low scan rate), practical applications of supercapacitor involves rapid extraction of the energy in order to achieve high power. Understanding how heteroatom doping would help retain the capacitance at *in-operando* conditions (*viz*., fast charging/discharging or at a high scan rate in CV measurement) is important for the rational design of carbon electrodes. To address this question, we present in Figure 9.7 the retention rate at 300 mV/s versus the micropore and mesopore surface areas at the surface compositions the same as those shown in Figure 9.5. The retention rate is calculated with the capacitance at scan rate of 1 mV/s as the reference. The scan rate of 300 mV/s is chosen because it is most relevant to the assessment of supercapacitor performance for practical applications. The average power density for the pristine carbon electrode is higher than 3 W/g over the range of surface areas considered in this work (shown in Figure 9.6).

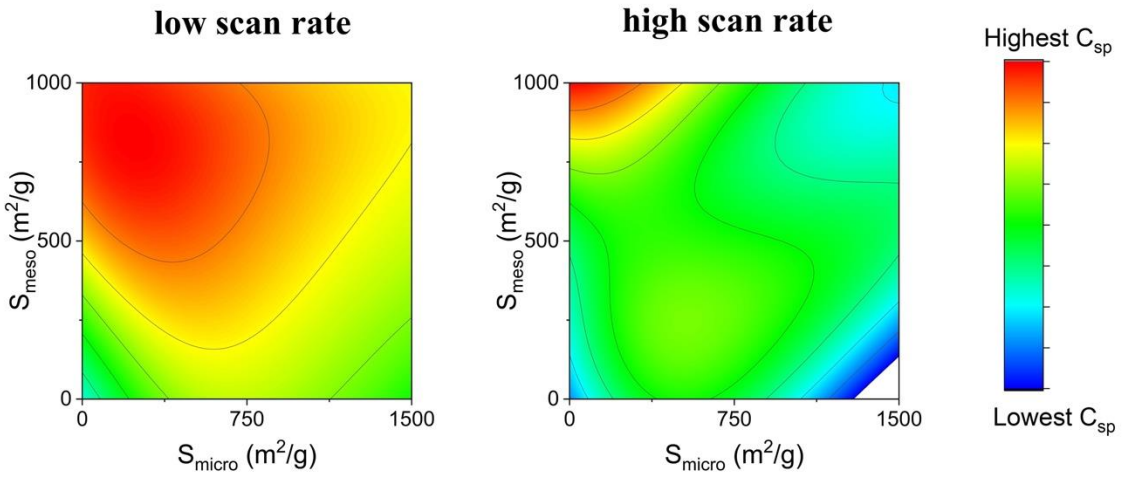

Figure 9.6. (A) Average power density of pristine carbon electrodes versus the scan rate predicted by *artificial neural network* (ANN). (B) Percentage of pristine carbon electrodes (with different surface areas of micropores and mesopores) that yields a power density exceeding 3 W/g at different scan rates.

Figure 9.7 presents the ANN prediction of the capacitance retention rates for the five representative carbon electrodes. Among the four different strategies of heteroatom doping considered in this work, only N-6 doping leads to a significant improvement of the retention rate. In this case, the high retention rate can be attributed to the contribution of N-6 doping to pseudocapacitance, which can be well maintained even far away from the equilibrium due to the fast surface reactions.[47]

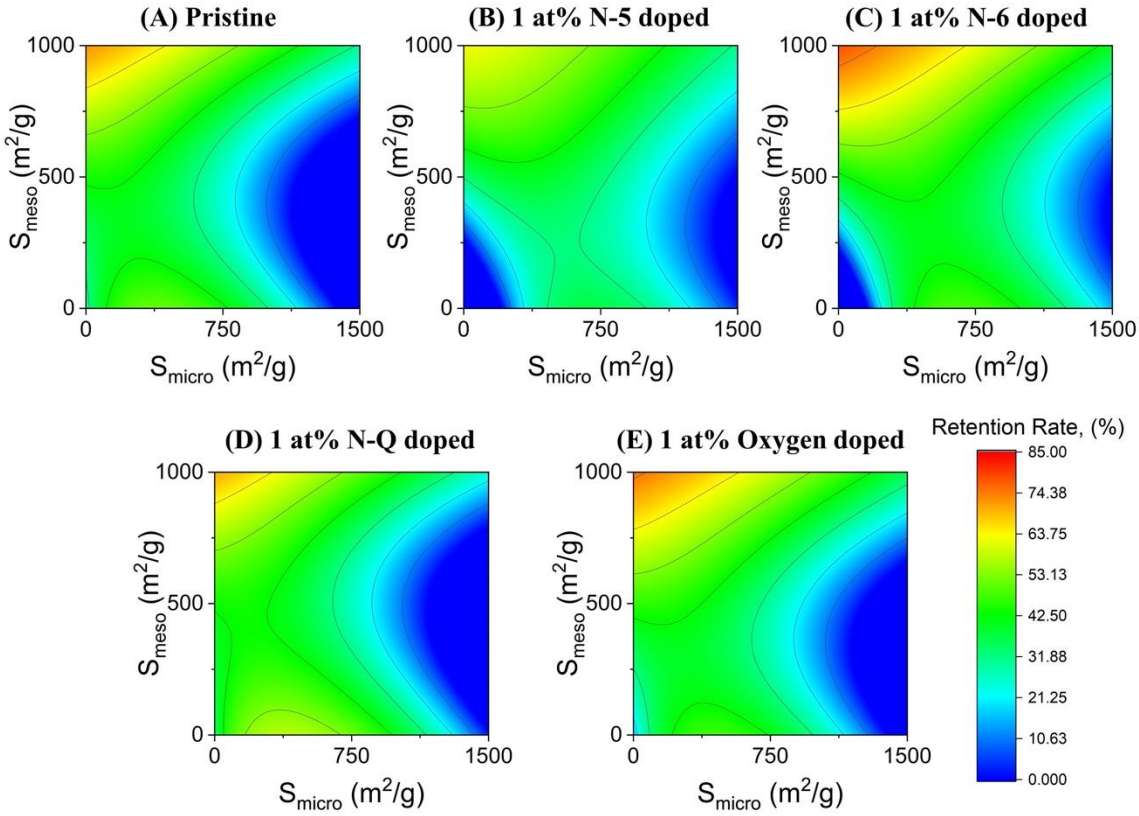

Figure 9.7. The capacitance retention rate at a scan rate of 300 mV/s versus the surface areas of micropores and mesopores predicted by the artificial neural network (ANN) for (A) pristine carbon and for the carbon doped by (B) 1 at% pyrrolic nitrogen, (C) 1 at% pyridinic nitrogen, (D) 1 at% quaternary nitrogen and (E) 1 at% oxygen.

Figure 9.7B shows a significant loss of the retention rate when the pristine carbon is doped with 1 at% N-5. In this case, a large binding energy between pyrrolic nitrogen and electrolyte ions prevents reversible charging/discharging.[43] On the other hand, 1 at% N-Q doping leads to a retention rate similar to that for the pristine carbon electrode. As aforementioned, the improved wettability and conductivity due to N-Q doping is counterbalanced by defects generated while introducing quaternary nitrogen atoms. Figure 9.7E shows that oxygen doping only slightly increases the retention rate. The

modest effect is because oxygen functional groups improve the surface wettability but also introduce additional resistance for ion transport.[35]

9.3.4 Carbon Electrodes with Optimized Features for Supercapacitor Energy Storage

With the quantitative correlations derived from the ANN model, we can now explore the best doping conditions for the carbon electrodes and the synergic effects arising from different local configurations of heteroatoms. Figure 9.8 shows the highest capacitance achieved by pristine and N/O co-doped carbon electrodes at 1 mV/s and 300 mV/s, respectively. Also shown in this figure are the dependence of the corresponding retention rates on the scan rate. For pristine carbon electrodes, the highest capacitance occurs when the porous carbon has the specific surface areas of mesopores and micropores at S_{micro}=300 m^2/g and S_{meso}=700 m^2/g and the maximum value (250.29 F/g) is achieved at a near equilibrium condition (scan rate of 1 mV/s). A larger surface area for mesopores (S_{meso}=1000 m^2/g) is required to achieve the highest capacitance far away from equilibrium (at the scan rate of 300 mV/s) because the larger pore is able to provide more ion accessible surface area during rapid electrosorption.

The ANN model indicates that, in comparison with those corresponding to an optimized pristine carbon, much larger micro- and mesopore surface areas (S_{micro}=1400 m^2/g and S_{meso}=1000 m^2/g) are required to achieve the highest capacitance for N/O co-doped carbon electrodes at both low and high scan rate. The shifting to higher surface areas can be explained by the fact that heteroatoms (N-6=3 at%, N-Q=6 at% and O=11.25 at%) doped on the carbon surface can significantly improve the wettability and

offer surface additional active sites for both quantum capacitance and pseudocapacitance. It is worth noting that the capacitance cannot keep increasing by increasing oxygen doping because oxygen functional groups may block the pore entrance, which is detrimental for ion transport at conditions far away from equilibrium.[35]

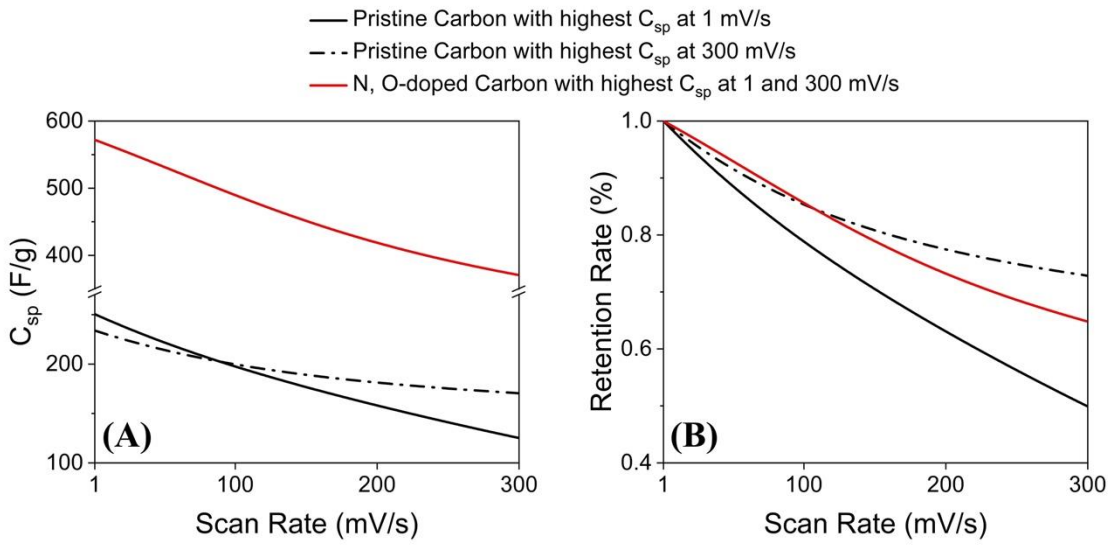

Figure 9.8. The capacitance versus scan rate predicted by the ANN model for pristine and nitrogen- and oxygen-doped carbon electrodes that yield the highest capacitance at 1 and 300 mV/s scan rates. (A) Carbon electrodes with the highest capacitance at 1 mV/s (black solid line, S_{micro}=300 m^2/g and S_{meso}=700 m^2/g) and at 300 mV/s (black dash-dotted line, S_{micro}=0 m^2/g and S_{meso}=1000 m^2/g), and the optimal electrode with heteroatom doping (red line, N-6=3 at%, N-Q=6 at% and O=11.25 at%). The heteroatom doping results in the same optimal electrode at 1 and 300 mV/s san rates. (B) The same as (A) but in terms of the retention rate.

According to Figure 9.8A, the capacitance retention rate for pristine carbon electrodes can be improved by reducing the surface area of micropores while increasing the mesopore area. With more mesopores but less micropores, the carbon electrode offers a higher capacitance at large scan rates because such a pore architecture can minimize the resistance of ion transport induced by micropores while maximizing the surface area for

electrosorption. For N/O co-doped carbon electrodes, the resistance is already reduced by doping quaternary nitrogen and oxygen. Although decreasing the micropore surface area would help improve the retention rate, a smaller surface area means a lower capacitance because less surface area is available for electrosorption and faradaic reaction.

9.3.5 Ragone Plot

The Ragone plot provides an intuitive benchmark for the comparison of different energy storage devices in terms of the energy density and power density. To locate the boundary in the Ragone plot for the supercapacitors with N/O co-doped carbon electrodes, we use the ANN model to calculate the *in-operando* capacitance and power density at all possible surface areas of micropores and mesopores and surface doping compositions. The energy density of a supercapacitor cell is calculated from eq (9.2) with a potential window of 1 V (since all the data used in this work were tested with this potential window).

Figure 9.9 shows predictions from the ANN model for the energy and power boundaries related to the performance of supercapacitors consisting of pristine and N/O co-doped carbon electrodes with 6M KOH aqueous electrolyte. The boundaries for other energy storage devices, such as batteries and conventional electrical capacitor are also presented in Figure 9.9 for comparison.[48] With the carbon surface doped with nitrogen- and oxygen- atoms, the supercapacitor region in the Ragone plot is significantly expanded in comparison to that for the pristine carbons. The ANN model predicts that the energy density of the best N/O co-doped carbon electrode can catch up with that for a commercial battery but with a much higher power density.

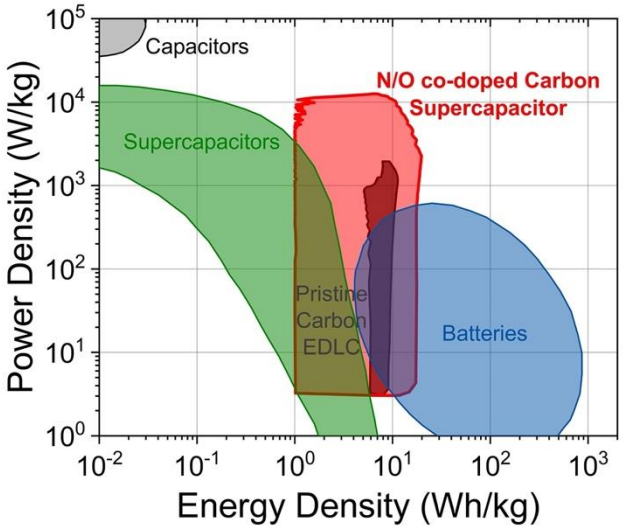

Figure 9.9. The Ragone plot for supercapacitors consisting of nitrogen/oxygen co-doped carbon electrodes with 6M KOH aqueous solution. The red region is predicted by the artificial neural network (ANN) model developed in this work. The grey, green, blue and black region correspond to the conventional electrical capacitors, batteries,[48] and pristine carbon electrical double layer capacitors predicted by an ANN model developed in our previous work.[15]

The structural features predicted by the ANN model may be realized by co-assembling melamine/formaldehyde with silica spheres as shown in Figure 9.10.[28] Among all systems considered in this work, carbon electrodes synthesized by the colloid-templated method achieves the closest doping configuration that maximizes the power and energy density. The unique choice of nitrogen precursor provides not only high nitrogen content but also adequate local doping configurations. To lower the doping percentage of pyrrolic nitrogen, one may reduce the amount of water used to dissolve melamine/formaldehyde and adjust activation temperature.[28] A higher mesopore surface area can be achieved by adding silica spheres of slightly smaller diameter or by using microwave-assisted hydrothermal method to dissolve melamine/formaldehyde into

the solution. The former procedure could increase the mesopore surface area by reducing the template size while the latter has shown to be successful in increasing the mesopore surface as this process helps create more mesopores.[20]

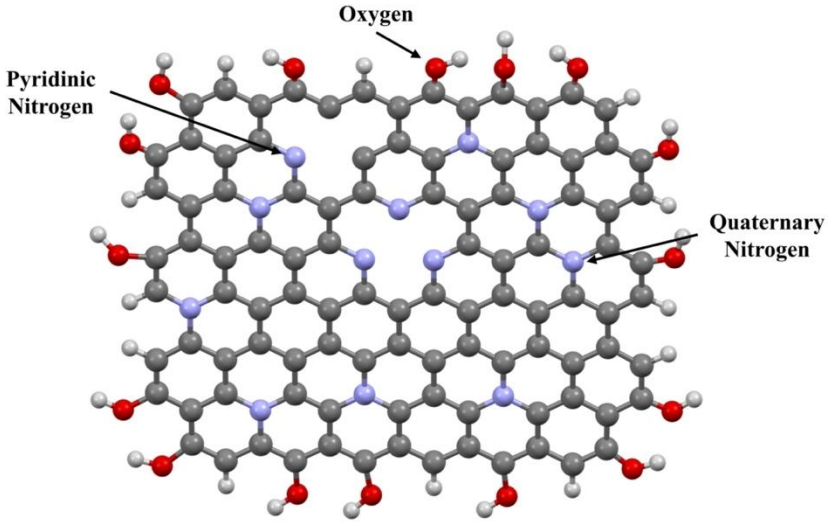

Figure 9.10. Schematic surface of the best N/O co-doped carbon electrode (S_{micro}=1400 m^2/g, S_{meso}=1000 m^2/g, N-6=3 at%, N-Q=6 at% and O=11.25 at%) predicted by the artificial neural network (ANN). Here grey, white, purple and red spheres represent carbon, hydrogen, nitrogen and oxygen atoms, respectively.

9.4 Conclusions

In this work, we have tested four different machine learning (ML) models to correlate the dependence of the supercapacitor performance on the structural features and surface compositions of N/O co-doped carbon electrodes. By testing with a large set of capacitance and power density data from the literature, we find that the artificial neural network (ANN) shows the best performance in terms of both energy and power density. The ANN model is thus further used to investigate how the surface composition and local configurations of heteroatoms, together with the structural features of porous carbon,

influence the supercapacitor performance in 6M KOH solution. Consistent with the predictions of physics-based models, doping with pyridinic nitrogen (N-6) can significantly enhance the capacitance of carbon electrodes. Such effect may be explained in term of additional contributions from the quantum capacitance and pseudocapacitance. Doping with pyrrolic nitrogen (N-5), however, would be detrimental to the capacitance, especially at conditions far away from equilibrium (*viz*., at high scan rate). The reduction in capacitance and retention rate can be attributed to too high binding energy between N-5 atoms and potassium ions. Both quaternary nitrogen (N-Q) and oxygen doping is able to improve the retention rate at fast discharging because the heteroatoms enhance the electrical conductivity and wettability of carbon electrodes.

While the machine learning models tested in this work could be improved if more data are available, the ANN model allows us to account for the synergetic effects of nitrogen and oxygen doping, and identify the best doping conditions and structural features of carbon electrodes with the highest energy density. The data-driven approach predicts that carbon electrodes doped with nitrogen and oxygen are able to achieve an energy density comparable to that corresponding to commercial batteries but with a much higher power density. Carbon electrodes with optimal features for energy storage may be synthesized by co-assembling melamine/formaldehyde with silica nanoparticles as the template or with microwave-assisted hydrothermal methods.

Bibliography

1. Berrueta, A., et al., *Supercapacitors: Electrical Characteristics, Modeling, Applications, and Future Trends.* Ieee Access, 2019. **7**: p. 50869-50896.

2. Simon, P. and Y. Gogotsi, *Capacitive energy storage in nanostructured carbon-electrolyte systems.* Acc Chem Res, 2013. **46**(5): p. 1094-103.

3. Stoller, M.D. and R.S. Ruoff, *Best practice methods for determining an electrode material's performance for ultracapacitors.* Energy & Environmental Science, 2010. **3**(9): p. 1294-1301.

4. Ghosh, A. and Y.H. Lee, *Carbon-based electrochemical capacitors.* ChemSusChem, 2012. **5**(3): p. 480-99.

5. Zhai, Y., et al., *Carbon materials for chemical capacitive energy storage.* Adv Mater, 2011. **23**(42): p. 4828-50.

6. Xia, J., et al., *Measurement of the quantum capacitance of graphene.* Nat Nanotechnol, 2009. **4**(8): p. 505-9.

7. Zhang, J., et al., *Preparation of activated carbon from waste Camellia oleifera shell for supercapacitor application.* Journal of Solid State Electrochemistry, 2012. **16**(6): p. 2179-2186.

8. Zhan, C., et al., *Enhancing graphene capacitance by nitrogen: effects of doping configuration and concentration.* Phys Chem Chem Phys, 2016. **18**(6): p. 4668-74.

9. Wang, L., et al., *Origin of theoretical pseudocapacitance of two-dimensional supercapacitor electrodes Ti3C2T2 (T = bare, O, S).* Journal of Materials Chemistry A, 2019. **7**(27): p. 16231-16238.

10. Wang, S., et al., *Insights into CO2/N2 Selectivity in Porous Carbons from Deep Learning.* ACS Materials Letters, 2019. **1**(5): p. 558-563.

11. Zhang, Z., et al., *Prediction of Carbon Dioxide Adsorption via Deep Learning.* Angew Chem Int Ed Engl, 2019. **58**(1): p. 259-263.

12. Su, H., et al., *Predicting the capacitance of carbon-based electric double layer capacitors by machine learning.* Nanoscale Advances, 2019. **1**(6): p. 2162-2166.

13. Zhu, S., et al., *Artificial neural network enabled capacitance prediction for carbon-based supercapacitors.* Materials Letters, 2018. **233**: p. 294-297.

14. Su, H., et al., *Predicting the Capacitance of Carbon-based Electric Double Layer Capacitors by Machine Learning.* Nanoscale Advances, 2019.

15. Zhou, M., et al., *Insights from machine learning of carbon electrodes for electric double layer capacitors.* Carbon, 2020. **157**: p. 147-152.

16. Wu, H., et al., *The effect of activation technology on the electrochemical performance of calcium carbide skeleton carbon.* Journal of Solid State Electrochemistry, 2012. **16**(9): p. 2941-2947.

17. Chen, X.Y., et al., *Nitrogen-doped porous carbon for supercapacitor with long-term electrochemical stability.* Journal of Power Sources, 2013. **230**: p. 50-58.

18. Wang, Y., et al., *Preparation of novel pigskin-derived carbon sheets and their low-temperature activation-induced high capacitive performance.* RSC Adv., 2014. **4**(85): p. 45318-45324.

19. Li, Y.-T., et al., *Hierarchical porous active carbon from fallen leaves by synergy of K2CO3 and their supercapacitor performance.* Journal of Power Sources, 2015. **299**: p. 519-528.

20. Tan, J., et al., *Nitrogen-doped porous carbon derived from citric acid and urea with outstanding supercapacitance performance.* Electrochimica Acta, 2015. **178**: p. 144-152.

21. Jiang, L., et al., *Construction of nitrogen-doped porous carbon buildings using interconnected ultra-small carbon nanosheets for ultra-high rate supercapacitors.* Journal of Materials Chemistry A, 2016. **4**(29): p. 11388-11396.

22. Hao, X., et al., *Bacterial-cellulose-derived interconnected meso-microporous carbon nanofiber networks as binder-free electrodes for high-performance supercapacitors.* Journal of Power Sources, 2017. **352**: p. 34-41.

23. Yang, W., et al., *Template-free synthesis of ultrathin porous carbon shell with excellent conductivity for high-rate supercapacitors.* Carbon, 2017. **111**: p. 419-427.

24. Zhang, D., et al., *Scalable synthesis of hierarchical macropore-rich activated carbon microspheres assembled by carbon nanoparticles for high rate performance supercapacitors.* Journal of Power Sources, 2017. **342**: p. 363-370.

25. Liu, W., et al., *Nitrogen-Doped Hierarchical Porous Carbon from Wheat Straw for Supercapacitors.* ACS Sustainable Chemistry & Engineering, 2018. **6**(9): p. 11595-11605.

26. Song, Z., et al., *Nitrogen-Enriched Hollow Porous Carbon Nanospheres with Tailored Morphology and Microstructure for All-Solid-State Symmetric Supercapacitors.* ACS Applied Energy Materials, 2018. **1**(8): p. 4293-4303.

27. Zhu, D., et al., *A general strategy to synthesize high-level N-doped porous carbons via Schiff-base chemistry for supercapacitors.* Journal of Materials Chemistry A, 2018. **6**(26): p. 12334-12343.

28. Song, Z., et al., *Synergistic design of a N, O co-doped honeycomb carbon electrode and an ionogel electrolyte enabling all-solid-state supercapacitors with an ultrahigh energy density.* Journal of Materials Chemistry A, 2019. **7**(2): p. 816-826.

29. Xue, D., et al., *Template-Free, Self-Doped Approach to Porous Carbon Spheres with High N/O Contents for High-Performance Supercapacitors.* ACS Sustainable Chemistry & Engineering, 2019. **7**(7): p. 7024-7034.

30. Ma, C., et al., *Sustainable recycling of waste polystyrene into hierarchical porous carbon nanosheets with potential applications in supercapacitors.* Nanotechnology, 2020. **31**(3): p. 035402.

31. Chen, X.Y., et al., *Nitrogen-Doped Porous Carbon Prepared from Urea Formaldehyde Resins by Template Carbonization Method for Supercapacitors.* Industrial & Engineering Chemistry Research, 2013. **52**(30): p. 10181-10188.

32. Mane, A.T. and V.B. Patil, *X-ray photoelectron spectroscopy of nanofillers and their polymer nanocomposites*, in *Spectroscopy of Polymer Nanocomposites*. 2016. p. 452-467.

33. Burden, F. and D. Winkler, *Bayesian regularization of neural networks.* Methods Mol Biol, 2008. **458**: p. 25-44.

34. Kumar, P., S.N. Merchant, and U.B. Desai, *Improving performance in pulse radar detection using Bayesian regularization for neural network training.* Digital Signal Processing, 2004. **14**(5): p. 438-448.

35. Seredych, M., et al., *Surface functional groups of carbons and the effects of their chemical character, density and accessibility to ions on electrochemical performance.* Carbon, 2008. **46**(11): p. 1475-1488.

36. Hulicova-Jurcakova, D., et al., *Combined Effect of Nitrogen- and Oxygen-Containing Functional Groups of Microporous Activated Carbon on its Electrochemical Performance in Supercapacitors.* Advanced Functional Materials, 2009. **19**(3): p. 438-447.

37. Haibo, H. and E.A. Garcia, *Learning from Imbalanced Data.* IEEE Transactions on Knowledge and Data Engineering, 2009. **21**(9): p. 1263-1284.

38. Kondrat, S., O.A. Vasilyev, and A.A. Kornyshev, *Feeling Your Neighbors across the Walls: How Interpore Ionic Interactions Affect Capacitive Energy Storage.* J Phys Chem Lett, 2019. **10**(16): p. 4523-4527.

39. Barbieri, O., et al., *Capacitance limits of high surface area activated carbons for double layer capacitors.* Carbon, 2005. **43**(6): p. 1303-1310.

40. Wang, H., T. Maiyalagan, and X. Wang, *Review on Recent Progress in Nitrogen-Doped Graphene: Synthesis, Characterization, and Its Potential Applications.* ACS Catalysis, 2012. **2**(5): p. 781-794.

41. Noori, A., et al., *Towards establishing standard performance metrics for batteries, supercapacitors and beyond.* Chem Soc Rev, 2019. **48**(5): p. 1272-1341.

42. Zhang, L.L., et al., *Nitrogen doping of graphene and its effect on quantum capacitance, and a new insight on the enhanced capacitance of N-doped carbon.* Energy & Environmental Science, 2012. **5**(11): p. 9618-9625.

43. Jeong, H.M., et al., *Nitrogen-doped graphene for high-performance ultracapacitors and the importance of nitrogen-doped sites at basal planes.* Nano Lett, 2011. **11**(6): p. 2472-7.

44. Lin, L., et al., *Nitrogen cluster doping for high-mobility/conductivity graphene films with millimeter-sized domains.* Sci Adv, 2019. **5**(8): p. eaaw8337.

45. Zhao, L., et al., *Visualizing individual nitrogen dopants in monolayer graphene.* Science, 2011. **333**(6045): p. 999-1003.

46. Andreas, H.A. and B.E. Conway, *Examination of the double-layer capacitance of an high specific-area C-cloth electrode as titrated from acidic to alkaline pHs.* Electrochimica Acta, 2006. **51**(28): p. 6510-6520.

47. Hulicova, D., M. Kodama, and H. Hatori, *Electrochemical Performance of Nitrogen-Enriched Carbons in Aqueous and Non-Aqueous Supercapacitors.* Chemistry of Materials, 2006. **18**(9): p. 2318-2326.

48. Mathis, T.S., et al., *Energy Storage Data Reporting in Perspective—Guidelines for Interpreting the Performance of Electrochemical Energy Storage Systems.* Advanced Energy Materials, 2019. **9**(39).

Chapter 10. Conclusions and Outlook

In this dissertation, both physics-based and data-driven models are used to accomplish the inverse design of nanoporous materials for gas separation and energy storage. One key contribution of this dissertation is the development of physics-based models to provide rapid and accurate prediction of gas sorption and transport properties in nanoporous materials. More specifically, accurate Helmholtz energy functionals for multicomponent gas mixtures are developed in classical density functional theory (cDFT) to provide accurate prediction of adsorption amount compared with grand canonical Monte Carlo (GCMC) simulation and experimental measurements. In addition, simplified string method is extended to find the minimum energy path (MEP) of rigid polyatomic molecules in nanoporous materials, which yields accurate prediction of diffusion coefficient compared with molecular dynamics (MD) simulation. In combination of graphic processing unit (GPU)-accelerated implementation, the developed computational theories and tools are orders of magnitudes faster than conventional molecular simulation methods. With efficient searching algorithm, the developed physics-based models empower, for the first time, the inverse design of nanoporous materials in membrane separation. Another key contribution of this dissertation is the demonstration of how data-driven models can help understand and inverse design nanoporous materials in gas separation and energy storage. For gas separation, data-driven models are used to recognize the patterns of nanoporous materials with excellent separation performance. For energy storage, data-driven models can well correlate the *in-operando* performance with the structural and chemical features of electrode materials. In collaboration with

experimental scientist, the predicted structural and chemical features of activated carbon electrodes with ultrahigh capacitance have also been verified experimentally.

Following the introduction and methodology chapters (Chapter 1 and Chapter 2), Chapter 3 and Chapter 4 discuss the development of cDFT for accurate prediction of multicomponent gas mixture adsorption in nanoporous materials and the computational framework and implementation for the massively parallel GPU-accelerated cDFT. The key findings can be summarized as follows:

1. For a number of gas mixtures in MOF-5 (without sub-pores inaccessible to gas molecules), the adsorption isotherms predicted by cDFT, using weighted density approximation for correlation effects and van der Waals one fluid theory, are in quantitative agreement with that calculated by grand canonical Monte Carlo (GCMC) simulation and experimental measurements, but massively parallel GPU-accelerated cDFT are two to three orders of magnitude faster.

2. Even at high pressure/loadings, cDFT still gives excellent prediction of adsorption isotherms and selectivity while the ideal adsorbed solution theory (IAST) fails to predict the separation selectivity qualitatively due to the missing adsorbate-adsorbate interaction and correlation effects.

3. The massively parallel GPU-accelerated implementation of cDFT can reduce the computational cost by more than two orders of magnitude for variety types of calculation related in gas adsorption including Langmuir-type adsorption isotherm, adsorption with vapor-liquid transition and high-throughput screening.

Chapter 5 extends the simplified string method with the massively parallel GPU-accelerated implementation to calculate the MEP of rigid polyatomic molecules in nanoporous materials. The key conclusions are:

1. The diffusion coefficients of guest molecules, ranging from nitrogen and ethane to benzene and xylene in nanoporous materials can be calculated from the MEP via the transition-state theory (TST) agree well with those calculated from MD simulation.

2. When the developed computational tool is used to high-throughput screen nanoporous materials for ethane/ethylene separation, more than 90% calculations can be done within 30 seconds which is orders of magnitude faster than conventional simulation method.

Chapter 6 and Chapter 7 showcase the integration of physics-based models developed in the dissertation and data-driven models to inverse design nanoporous materials for the separation of simple gas and rigid polyatomic gas molecules. The major findings are as follows:

1. Membrane separation can better balance the separation capacity and selectivity while adsorption separation often suffers from low separation selectivity at high separation capacity. The top nanoporous materials identified have much higher separation selectivity compared with state-of-art literature.

2. According to the high-throughput screening, for the membrane separation of D_2/H_2, nanoporous materials with the pore limiting diameter (PLD) ranging from 2 Å to 3 Å can achieve excellent separation selectivity of D_2/H_2 via kinetic quantum sieving

effects while for the membrane separation of C_2H_4/C_2H_6 nanoporous materials with PLD between 2.75 Å and 3.5 Å have the top 5% membrane selectivity.

3. With the data-driven models, the structural features of nanoporous materials with top 10% membrane separation can be identified to guide the inverse design of nanoporous materials. Whereas the search algorithm, more specifically genetic algorithm, along with developed physics-based model can directly design nanoporous materials with much higher membrane selectivity and permeability compared with all existing experimental MOF candidates.

Chapter 8 and Chapter 9 demonstrate data-driven models can be used to understand the *in-operando* performance of electrode materials and inverse design electrodes for the optimal performance. The key findings are as follows:

1. The quantitative correlations between the structural and chemical features of activated carbon electrodes and their *in-operando* performance in supercapacitor can be established via data-driven models.

2. The data-driven models can offer unique insights on how different structural and chemical features affect the overall *in-operando* performance of activated carbon electrodes in supercapacitor. While the maximum capacitance can be achieved by tuning micro- and mesopore surface area for pristine carbon electrodes, the effects of chemical doping, such as nitrogen- and oxygen-doping, depend on the local bonding configurations, doping percentage and available surface area.

3. The data-driven models also allow the direct inverse design of activated carbon electrode with optimal/target *in-operando* performance. The ultrahigh capacitance

with the structural and chemical features predicted by the data-driven models has also been verified in experiments.

In summary, this dissertation demonstrates how the inverse design of nanoporous materials for gas separation and energy storage can be accomplished by using physics-based and data-driven models. The theoretical framework and computational tools developed in this dissertation provide valuable computational infrastructures for molecular modeling and database construction of gas sorption and transport properties in nanoporous materials. The combination of physics-based and data-driven models in this dissertation opens up opportunities to customize nanoporous materials with the optimal/target performance for the gas separation and energy storage system of academic and industrial interest.

In order to leverage physics-based models to inverse design of nanoporous materials for a wider spectrum of applications, future development in physics-based models is still needed, for example, to describe the “phase transition” of water at relatively low saturation pressure in nanoporous materials for water harvesting nanoporous materials and flexibility effect of host materials and guest molecules for the inverse design of flexible nanoporous materials for more versatile gas separation and energy storage. For the future development of data-driven models, in addition to better molecular fingerprints to correlate the performance of nanoporous materials, better machine learning models need to be developed to efficiently project the design space for the search of new promising materials with target performance. While machine learning models have been attempted to replace conventional intermolecular potential in

molecular simulations, opportunities also exist to examine the integration of data-driven models and classical density functional theory (cDFT), more specifically whether machine learning models can be used to replace physics-based excess Helmholtz energy term in cDFT for the more accurate prediction of gas adsorption in nanoporous materials.

To sum up, I hope this dissertation demonstrated a clear theoretical and computational framework for the inverse design of nanoporous materials for gas separation and energy storage and could serve as the cornerstone for the future development of physics-based and data-driven models to investigate, understand and inverse design nanoporous materials.