ABSTRACT

DATA-DRIVEN 2D MATERIALS DISCOVERY FOR NEXT-GENERATION ELECTRONICS

by
Zeyu Zhang

The development of material discovery and design has lasted centuries in human history. After the concept of modern chemistry and material science was established, the strategy of material discovery relies on the experiments. Such a strategy becomes expensive and time-consuming with the increasing number of materials nowadays. Therefore, a novel strategy that is faster and more comprehensive is urgently needed. In this dissertation, an experiment-guided material discovery strategy is developed and explained using metal-organic frameworks (MOFs) as instances. The advent of π-stacked layered MOFs, which offer electrical conductivity on top of permanent porosity and high surface area, opened up new horizons for designing compact MOF-based devices such as battery electrodes, supercapacitors, and spintronics. Structural building blocks, including metal nodes and organic linkers in these electrically conductive (EC) MOFs, are recognized and taking permutations among the building blocks results in new systems with unprecedented and unexplored physical and chemical properties. With the ultimate goal of providing a platform for accelerated material design and discovery, here the foundation is laid for the creation of the first comprehensive database of EC MOFs with an experimentally guided approach. The first phase of this database, coined EC-MOF/Phase-I, is composed of 1,057 bulk and monolayer structures built by all possible combinations of experimentally reported organic linkers, functional groups, and metal nodes. A high-throughput (HT) workflow is constructed to implement density functional theory calculations with periodic boundary conditions to optimize the structures and calculate some of their most relevant properties. Because research and development

in the area of EC MOFs has long been suffering from the lack of appropriate initial crystal structures, all of the geometries and property data have been made available for the use of the community through an online platform that was developed during the course of this work. This database provides comprehensive physical and chemical data of EC-MOFs as well as the convenience of selecting appropriate materials for specific applications, thus accelerating the design and discovery of EC MOF-based compact devices. Machine learning (ML), a technique of learning patterns of numerical data and making predictions, can be utilized in material discovery. Taking advantages of the EC-MOF Database, ML is adopted to predict property data that needs expensive calculations according to the crystal structures only. The implementation of ML is much faster than the HT workflow when the number of structures increases constantly.

DATA-DRIVEN 2D MATERIALS DISCOVERY FOR NEXT-GENERATION ELECTRONICS

by
Zeyu Zhang

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemistry

Department of Chemistry and Environmental Science

August 2023

APPROVAL PAGE

DATA-DRIVEN 2D MATERIALS DISCOVERY FOR NEXT-GENERATION ELECTRONICS

Zeyu Zhang

Dr. Farnaz A. Shakib, Dissertation Advisor — Date
Assistant Professor of Chemistry and Environmental Science, NJIT

Dr. Hao Chen, Committee Member — Date
Professor of Chemistry and Environmental Science, NJIT

Dr. Mengyan Li, Committee Member — Date
Associate Professor of Chemistry and Environmental Science, NJIT

Dr. Yuanwei Zhang, Committee Member — Date
Assistant Professor of Chemistry and Environmental Science, NJIT

Dr. Joshua Young, Committee Member — Date
Assistant Professor of Chemical and Materials Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Zeyu Zhang

Degree: Doctor of Philosophy

Date: August 2023

Undergraduate and Graduate Education:

- Doctor of Philosophy in Chemistry,
  New Jersey Institute of Technology, Newark, NJ, 2023
- Bachelor of Science,
  Beijing University of Chemical Technology, 2017

Major: Chemistry

Presentations and Publications:

Z. Zhang, "Data-driven 2D Materials Discovery for Next-Generation Electronics," *American Chemical Society 2022 Middle Atlantic Regional Meeting*, New York City, NY, 2023

Z. Zhang, D. S. Valente, Y. Shi, D. K. Limbu, M. R. Momeni and F. A. Shakib, "In Silico High-Throughput Design and Prediction of Structural and Electronic Properties of Low-Dimensional Metal–Organic Frameworks," *ACS Applied Materials & Interfaces*, vol. 15(7), pp 9494-9507, 2023.

Y. Shi, M. R. Momeni, Y. J. Chen, D. K. Limbu, Z. Zhang and F. A. Shakib, "Water Induced Structural Transformations in Flexible Two-Dimensional Layered Conductive Metal–Organic Frameworks," *Chemistry of Materials*, vol. 34(17), pp 7730-7740, 2022.

Z. Zhang, "EC-MOFs Database, A New Experimentally Guided Computational Database for Electrically-Conductive Metal-Organic Frameworks," *American Chemical Society 2022 Northeast Regional Meeting*, Rochester, NY, 2022

Z. Zhang, D. Dell'Angelo, M. R. Momeni, Y. Shi and F. A. Shakib, "Metal–to–Semiconductor Transition in Two-Dimensional Layered Metal-Organic Frameworks: An Ab Initio Perspective," *ACS Applied Materials & Interfaces*, vol. 13(21), pp 25270-25279, 2021.

M. R. Momeni, Z. Zhang, D. Dell'Angelo and F. A. Shakib, "Tuning Electronic Properties of Conductive 2D Layered Metal-Organic Frameworks via Host-Guest Interactions: Dioxygen as An Electroactive Chemical Stimuli," *APL Materials*, vol. 9, pp 051109, 2021.

M. R. Momeni, Z. Zhang, D. Dell'Angelo and F. A. Shakib, "Gauging van der Waals Interactions in Aqueous Solutions of 2D MOFs: When Water Likes Organic Linkers More than Open-metal Sites," *Physical Chemistry Chemical Physics*, vol. 23, pp 3135-3143, 2021.

M. R. Momeni, Z. Zhang and F.A. Shakib, "Determining Role of Structural Flexibility on Catalytic Activity of Conductive 2D Layered Metal-Organic Frameworks," *Chemical Communications*, vol. 57, pp 315-318, 2021.

Z. Zhang, "Charge Transport in 2D Layered Metal-Organic Frameworks: Correlating Flexibility with Conductivity," *2021 Dana Knox Student Research Showcase*, New Jersey Institute of Technology, 2021

Z. Zhang, "A Study on Electronic Properties of 2D MOFs Affected by Structural Deformations," *2020 Virtual MRS Spring/Fall Meeting and Exhibit*, Virtual, 2020

A single spark can start a prairie fire.

Zedong Mao

ACKNOWLEDGMENTS

In this long journey to the Doctorate of Philosophy, I would not have been able to reach this far without the generous encouragement and help from the people I have met. Deepest appreciation to Dr. Jianbiao Zhang and Dr. Meifang Yan, who encouraged and led me to the beginning of this journey.

I would like to express my heartfelt gratitude to my advisor, Dr. Farnaz A. Shakib, for her invaluable guidance, unwavering support and unrestricted research directions throughout my journey. She is the best advisor to me that I have ever seen or heard. Acknowledgment to Dr. Mohammad R. Momeni for his intelligence and generous help.

Many thanks to Dr. Hao Chen, Dr. Mengyan Li, Dr. Yuanwei Zhang and Dr. Joshua Young, who are not only the defense committee members and but also the committee members in my qualifying exam and research proposal.

I would like to thank the College of Science and Liberal Arts for the Provost Scholarship and the Department of Chemistry and Environmental Science for the teaching assistant awards that support me and my research during the past five years.

Much appreciation to my friend and research companion Yuliang Shi, whom I've known since the first day in the university. I would like to thank Dr. David Dell'Angelo, Dr. Dil K. Limbu and Dylan S. Valente for their broad knowledge and collaborative spirit.

Meeting you is the greatest fortune in my life.

To Goddess Wang

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Chemistry, often referred to as the central science, has a rich and fascinating history that spans thousands of years. It is the branch of science that explores the composition, properties, and transformations of matter. From ancient civilizations' alchemical pursuits to the modern scientific advancements, chemistry has played a pivotal role in shaping our understanding of the world. The story of chemistry begins in ancient times, with the first known experiments conducted by early civilizations such as the Egyptians, Greeks, and Chinese. These early practitioners focused on practical applications, such as metalworking, pottery, and medicine. Alchemy, a precursor to modern chemistry, emerged during the Hellenistic period in Alexandria, Egypt. Alchemists sought to transmute base metals into gold and discover the philosopher's stone, believed to grant endless wealth. The Renaissance period marked a turning point for chemistry, as it transitioned from the mystical realm of alchemy to a more scientific discipline. In the 17th century, Robert Boyle's publication of "The Sceptical Chymist" challenged traditional alchemical beliefs and emphasized the importance of rigorous experimentation and observation [118]. Boyle's work laid the foundation for the modern scientific method, which would revolutionize the field of chemistry. The 18th century witnessed the emergence of quantitative chemistry. Antoine Lavoisier, often referred to as the "Father of Modern Chemistry," introduced precise measurements and the concept of chemical elements. Lavoisier's experiments on combustion and his formulation of the law of conservation of mass transformed chemistry into a quantitative science. His work also contributed to the development of the periodic table of elements. In the next century, John Dalton's atomic theory, which proposed that matter is composed of indivisible particles

called atoms, revolutionized our understanding of the structure of matter. Dmitri Mendeleev's development of the periodic table, organizing elements based on their atomic properties, provided a systematic framework for classifying and predicting the behavior of elements. Over the past one hundred years, chemistry underwent a series of transformative advancements. The field of organic chemistry flourished with the discovery of new synthetic methods and the elucidation of complex molecular structures. The development of quantum mechanics provided a deeper understanding of atomic and molecular interactions.

Today, modern chemistry continues to push boundaries, with advancements in nanotechnology, materials science, and environmental chemistry. Chemists are exploring sustainable energy sources, designing new drugs, and developing innovative materials with unique properties. The field has become increasingly interdisciplinary, collaborating with biology, physics, engineering and computer science to address complex scientific challenges. One prominent challenge in the fields of chemistry and material science is how to find the best compound for a given purpose? Discovery of new compounds generally relies on the knowledge, experience and intuition of scientists. However, most of the advances still rely on trial-and-error experimentation which requires significant cost and time [68]. Especially, with the rapid increasing number of new compounds, a comprehensive search among all possible candidates becomes an impossible task. In recent years, computational chemistry approaches have been significantly improved by the advance of new theories/methodologies and computing software/hardware. The accuracy and efficiency of computational methods turn them to comparable tools to experimental approaches and promise unprecedented breakthroughs in discovering new compounds. In the past decade, artificial intelligence (AI) has led a technological revolution in this field. AI is the science and engineering of making intelligent machines, especially intelligent computer programs that can address complicated tasks like language translation,

autopilot and website management. It has also shown a promising future in sciences where AI techniques are applied to give guidance to scientists by AI's unique views. In this dissertation, we will introduce a state-of-the-art pathway for discovering new materials which combines computational chemistry tools, advanced programming and AI techniques. Such an approach integrates the experimental findings and hypothetical systems to provide a comprehensive picture of the target compounds. We will utilize in approach in investigating the inherent properties and future potential applications of an intriguing class of materials, known as metal organic frameworks (MOFs), in next-generation electronics. In Chapter 1, we will cover the background of MOF materials, theory of computational chemistry and machine learning techniques. In Chapter 2, the structure-function correlation of two-dimensional (2D) MOFs is investigated, which provides some insights about the electrical conductive behavior transition due to structural flexibility. In Chapter 3, we will introduce our newly-developed electrically-conductive (EC)-MOF/Phase-I Database which is a platform that allows users to search the desired MOF, download the structures and get the property data of the material. In Chapter 4, machine learning (ML) techniques are implemented for materials discovery purposes based on this database. The predictions made by our models shows the effectiveness of our data and also great potentials of ML on saving the computational resource. Concluding remarks are provided in Chapter 5.

1.1 Metal Organic Frameworks

1.1.1 Traditional metal organic frameworks

The interests in metal organic frameworks (MOFs), also called porous coordination polymers at that time, starts around 1990 when Hoskins and Robson set the basis for the future of MOFs [60, 59]. The term MOF itself was popularized by the work of Yaghi et al. around 1995 where a layered Co-trimesate with reversible sorption

properties was synthesized [163, 162]. Later, the most studied MOFs, MOF-5 and HKUST-1 (shown in Figure 1.1), were reported by Li et al. [88] and Chui et al. [24] who emphasized their extraordinary properties including great stability and rich porosity. In modern definition, MOFs are a class of porous crystalline materials

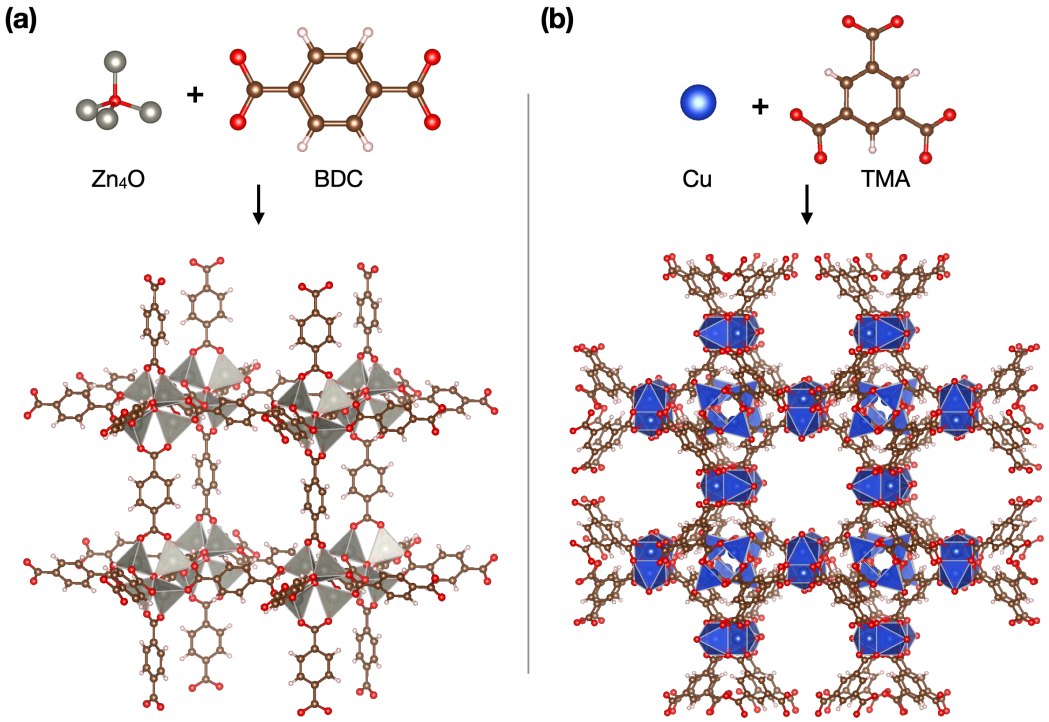

Figure 1.1 Composition and structures of MOF-5 (a) and HKUST-1 (b). The color coding is gray for Zinc, blue for Copper, red for Oxygen, Brown for Carbon and white for Hydrogen.

obtained through a process usually referred to as reticular synthesis [161]. Selected metal nodes and organic linkers, also called secondary building units (SBUs), are connected via strong coordination bonds to form ordered and permanently porous architectures [164]. As shown in Figure 1.1, MOF-5 was synthesized by diffusing triethylamine into a solution of zinc and 1,4-benzenedicarboxylate acid (H_2BDC) under the presence of dimethylformamide (DMF), chlorobenzene and hydrogen peroxide. Similarly, the reaction between copper(II) cations and benzene-1,3,5-tricarboxylic acid (TMA-H3) produces HKUST-1 MOF. All following reported MOFs

wer synthesized in similar manners. The modular nature of MOFs provided numerous opportunities to tailor their physical and chemical characteristics, which has led to more than 110,000 stable MOFs reported to date according to the Cambridge Structural Database (CSD) [46]. Their high surface area and permanent porosity have stimulated a spectrum of applications in gas adsorption/storage [136, 137], water treatment [120, 149], and catalysis [28, 102, 103, 105, 166], to name a few.

1.1.2 Electrically conductive metal organic frameworks

Most of traditional MOFs exhibit low electrical conductivity with large band gaps which limits their further utilization in electronics and optical devices [144]. The main approaches to improve the electrical conductivity of MOFs stand at the two charge transport pathways in MOFs: through-bond and through-space. More specifically, through-bond charge transfer requires a good overlap of bonding orbitals from the metal nodes and organic linkers and through-space charge transfer needs an appropriate spatial overlap between π conjugated systems. Through-bond pathway is mostly realized in the MOFs with 1D SBUs like MOF-74 and its analogues consisting of $(\text{-M-O-})_\infty$ chains [124]. The electrical conductivity of such MOFs is measured between 1.4×10^{-14} to 3.2×10^{-7} S/cm [159]. Considering the electronegativity of oxygen atoms, the electron gas is localized leading to a low conductivity. Sun et al. reported that substitution of $(\text{-M-O-})_\infty$ chains with $(\text{-M-S-})_\infty$ chains will improve the conductivity of the materials by increasing the metal-ligand covalency [139]. Through-space charge transfer focuses on the interaction between organic ligands instead of metal and ligand. The through-space pathway is mainly realized by the π-π stacking where the distance between organic ligands is critical to the conductivity. MOFs with La(III) cations and tetrathiafulvalene tetrabenzoate (TTFTB) linkers were reported as conductive materials by Xie et al [157]. By applying different experimental conditions, MOFs with three different topologies

were synthesized where their main difference is the distance between π-π stacking as shown in Figure 1.2 The distance at the stacking directions is 3.60 Å, 4.083 Å and

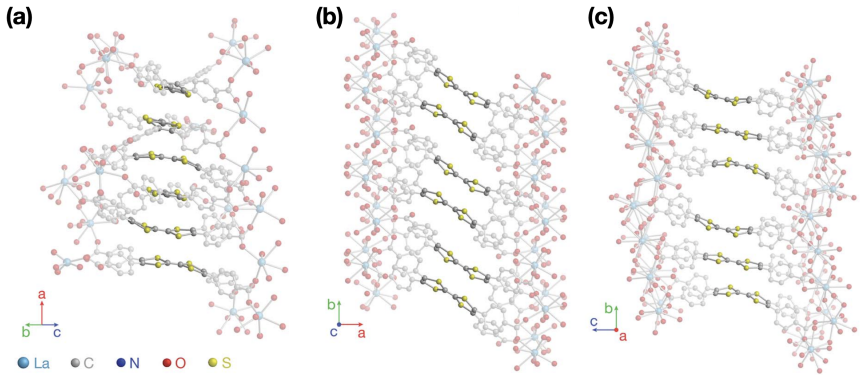

Figure 1.2 Crystal structure of (a) $La_4(HTTFTB)_4$, (b) La(HTTFTB) and (c) $La_4(HTTFTB)_3$. *Copyright 2019 Royal Society of Chemistry.*

7.072 Å for $La_4(HTTFTB)_4$, La(HTTFTB) and $La_4(HTTFTB)_3$, respectively. This increasing distance is accompanied by a decrease in electrical conductivity which is measured for them as 2.5×10^{-6}, 9×10^{-7} and 1×10^{-9} S/cm, respectively. In some special cases of MOFs, when pairing transition metals with chelating functional groups, charge transfer through extended conjugation is another possible pathway. MOFs with such structures are proposed to be metal organic analogues of graphene due to the similarity of extend π-d conjugation in the MOFs and the sp^2 hybridization in graphene. Generally, such MOFs possess two-dimensional (2D) structures and the extended conjugation occurs within the *ab* plane. To synthesize these 2D MOFs, π conjugation and redox-noninnocent chelating groups such as ortho-diols, -diamines and -dithiols are necessary in the organic linkers, as shown in Figure 1.3(a). Many organic ligands are partially oxidized during solvothermal growth, introducing charge carriers into the as-synthesized frameworks, which makes this pathway a dominant mechanism in 2D MOFs. The discovery 2D layered MOFs in 2012 introduced one important class of electrically conductive (EC) MOFs and opened a new research direction in this area [58]. 2D layered MOFs

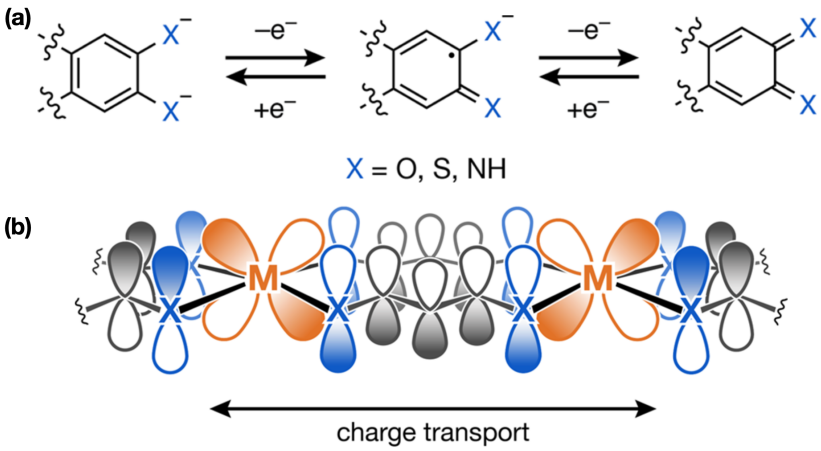

Figure 1.3 (a) Redox series of 2-, 1-, and 0 charge states for a deprotonated catecholoid fragment, a common motif in conductive MOFs with extended conjugation. (b) Charge Transfer via Extended Conjugation Pathway. *Copyright 2020 American Chemical Society.*

contain ortho-substituted π-conjugated organic linkers coordinated to (usually) early 3d transition metal nodes, forming extended π-conjugated 2D sheets. Weak van der Waals interactions allow stacking of these 2D sheets to form bulk crystalline materials with one-dimensional channels in the stacking direction. This architecture provides necessary paths for charge transport along both in-plane (extended conjugation) and out-of-plane (through-space) directions [169]. In 2012, the first 2D layerer MOF consisting of highly conjugated tricatecholate, 2,3,6,7,10,11-hexahydroxytriphenylene ($H_{12}C_{18}O_6$, HHTP) and Co(II) and Ni(II) cations was synthesized via solvothermal self-assembly, as shown in Figure 1.4. Inspired by this pioneering work, more and more 2D MOFs with conjugated systems have been synthesized. The number of such materials keeps increasing since they can provide the high surface area like traditional MOFs and electrical conductivity at the same time. These new 2D MOFs with excellent electrical conductivity and magnetic properties are introduced as viable candidates for field-effect transistors [155], supercapacitors [90], superconductors [63], spintronics [33] and cathode materials in different metal-ion batteries [143]. A brief

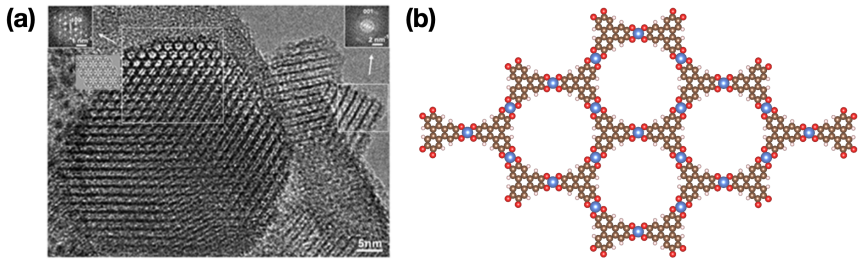

Figure 1.4 (a) High-magnification HR-TEM image of Ni-HHTP-1. *Copyright 2012 American Chemical Society.* (b) Built model for NiHHTP. Color code: Ni, blue; O, red; C, brown; H, white.

timeline of representative 2D MOFs with their reported electrical conductivities is show in Figure 1.5

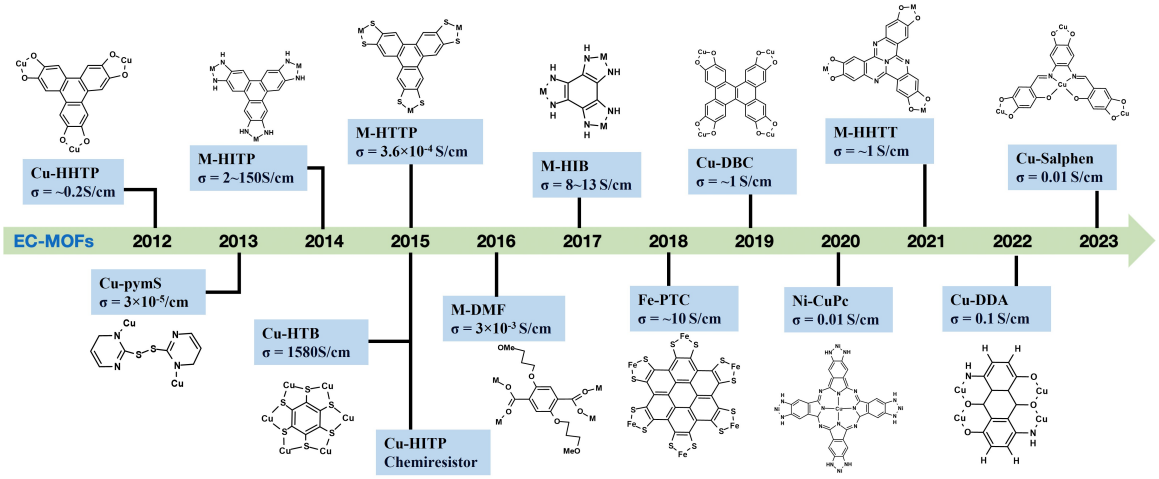

Figure 1.5 A brief timeline of representative 2D MOFs with their reported electrical conductivity.

1.2 Theoretical Aspects and Details

1.2.1 Schrödinger equation

In the late 17th century, Sir Isaac Newton formulated the famous laws of motion that relates the motion of a macroscopic object to the forces applied on it. These laws laid the foundation for classical mechanics and revolutionized our understanding of the physical world. They had become the basis for studying motion, force,

and interaction between objects in scientific research, engineering design, and technological advancements. However, in the 20th century, scientists found that the classical mechanics cannot describe the behavior of small objects like photons, electrons or light nuclei. To explain the behavior of such microscopic particles a new form of mechanics, named quantum mechanics, was developed by the efforts of generations of great scientists. In 1803, British polymath Thomas Young established the wave theory of lights based on the observation of double-slit experiments. Later, Max Planck proposed his hypothesis that energy is radiated and absorbed in discrete "quanta", which precisely describes the observation of black-body radiation. Inspired by this idea, Niels Bhor applied the concept of quantization of energy to electrons in hydrogen atoms and successfully predicted the spectral lines of hydrogen. Albert Einstein further developed this idea to show that the electromagnetic wave can be also treated as particles with discrete energy and frequency levels. Modern quantum mechanics was born in 1925 when Werner Heisenberg, Max Born, and Pascual Jordan invented the matrix mechanics [55]. And one year later, Erwin Schrödinger introduced the wave mechanics which is well-known as the time-dependent Schrödinger equation (TDSE) [130] whose general form is:

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = \hat{H}\Psi(\mathbf{r},t) \tag{1.1}$$

where i is the imaginary number and $\hbar$ is the reduced Planck constant. $\Psi(\mathbf{r},t)$ is the wavefunction that depends on position, i.e., three-dimensional vector $\mathbf{r}$, and time t. $\hat{H}$ is the Hamiltonian operator corresponding to the total energy of the system:

$$\hat{H} = -\frac{\hbar}{2m}\nabla^2 + V(\mathbf{r},t) \tag{1.2}$$

where the first term is kinetic energy and the second term is the potential energy. m is the mass of the particle and ∇ is the Laplace operator which can be expressed as

following in Cartesian coordinates:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{1.3}$$

Due to the complicated nature of TDSE, i.e., simultaneous dependence on position and time, many applications of quantum mechanics to chemistry rely on an approach called separation of variables to divide the total wavefunction to two functions that separately depend on position and time:

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) f(t). \tag{1.4}$$

This allows us to solve time-independent Schrödinger equation (TISE):

$$\hat{H}\psi(\mathbf{r}) = -\frac{\hbar}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.5}$$

to obtain the energy of the system E subject to time-independent wavefunction $\psi(\mathbf{r})$. Time-evolution part will be separately added if one is interested in the dynamics simulations. However, finding an exact solution of TISE for a many-electron system like a molecule is too complicated so that some approximations were created to simplify the problem. One significant approximation is Born–Oppenheimer approximation which says nuclei and electrons can be treated separately because nuclei are much heavier than electrons and in the time-scale of movements of electrons the positions of nuclei can be appropriately assumed fixed [15]. Thus, we only need to solve TISE for the electrons in a system. The TISE for a system with N number of electrons can be written as:

$$\hat{H}\psi(\mathbf{r}) = \left[-\frac{\hbar}{2m} \sum_{i=1}^{N} \nabla_i^2 + \sum_{i=1}^{N} V(\mathbf{r}_i) + \sum_{i=1}^{N} \sum_{j<i} U(\mathbf{r}_i, \mathbf{r}_j) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.6}$$

where E is the ground state energy of the system and terms in the brackets are kinetic energy of electrons, potential energy of each electron due to the fixed nuclei and the

interaction energy between different electrons. The wavefuntion $\psi(\mathbf{r})$ now is the wave function of N electrons which can be expressed as:

$$\psi(\mathbf{r}) = \psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) \tag{1.7}$$

1.2.2 Hartree-fock method

Although the nuclei is fixed in the system according to the Born–Oppenheimer approximation, solving TISE of electrons still needs more theoretical considerations to balance the accuracy and efficiency of the calculations. A pioneering methodology in this regards is called Hartree-Fock (HF) method. The basic assumption in the HF method is that the N electron wavefunction can be considered as a product of the individual wavefunctions of each electrons, the so-called Hartree Product

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \psi(\mathbf{r}_1)\psi(\mathbf{r}_2)\cdots\psi(\mathbf{r}_N)\cdot \tag{1.8}$$

Further, since each electron has an up or down spin, or spin α and β, respectively, the concept of spin orbitals were introduced. A space-spin coordinate $\mathbf{X} = \{\mathbf{r}, \omega\}$ is used to replace the position vector $\mathbf{r}$ and the notation of spatial wavefunctions $\psi(\mathbf{r})$ is also replaced by $\chi(\mathbf{X})$. The N electron wavefunction can be re-written using the space-spin coordinates:

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \chi_1(\mathbf{X}_1)\chi_2(\mathbf{X}_2)\cdots\chi_N(\mathbf{X}_N)\cdot \tag{1.9}$$

Finally, the HF method had to satisfy the antisymmetry principle, that is a wavefunction describing fermions (electrons) should be antisymmetric with respect to the interchange of any set of space-spin coordinates. This was achieved by representing the product of individual wavefunctions with a matrix form. i.e., Slater

determinant, as shown below:

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \frac{1}{\sqrt{N}} \begin{bmatrix} \chi_1(\mathbf{X}_1) & \chi_2(\mathbf{X}_1) & \cdots & \chi_N(\mathbf{X}_1) \\ \chi_1(\mathbf{X}_2) & \chi_2(\mathbf{X}_2) & \cdots & \chi_N(\mathbf{X}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{X}_N) & \chi_2(\mathbf{X}_N) & \cdots & \chi_N(\mathbf{X}_N)\cdot \end{bmatrix} \tag{1.10}$$

Notably, the movement of an electron is still considered independent of all other electrons except the Coulomb repulsion due to the average positions of all electrons. As a result, Hartree-Fock method is referred as an independent particle model or a mean field theory and it only partially accounts for electron correlation.

1.2.3 Density functional theory

The purpose of solving TISE is calculating the ground state energy of a system. But, one should note that an exact set of electron positions, $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N$, cannot be observed, the so-called Heisenberg uncertainty principle. The quantity that can be measured instead is the *probability* of finding N electrons at one specific set of position as following

$$p(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \psi^*(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) \tag{1.11}$$

The wavefunction of the N electron system can be further separated into the wavefunction of individual electron. Also, distinguishing each electron one by one is not relevant to our purpose, finding the ground-state energy. Hence, a summation over the probability of single electron wavefuntions gives the density of electrons:

$$n(\mathbf{r}) = 2\sum_{i=1}^{N} \psi^*(\mathbf{r})\psi(\mathbf{r})\cdot \tag{1.12}$$

The factor of 2 in Equation 1.12 is due to the spin states of electrons and Pauli exclusion principle. Hence, it is possible to use electron density instead of

wavefunction to describe a system. This basic assumption led to the development of density functional theory (DFT) based on two theorems proved by Kohn and Hohenberg in 1965 [76]. The two Hohenberg–Kohn theorems say (1) The ground-state energy from Schrödinger's equation is a unique functional of the electron density, (2) the electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger equation. In terms of a single electron wavefunction, the true functional described by the first theorem can be written as:

$$E\left[\{\psi_i\}\right] = E_{known}\left[\{\psi_i\}\right] + E_{XC}\left[\{\psi_i\}\right] \tag{1.13}$$

The E_{known} term can be expanded as:

$$E_{known}\left[\{\psi_i\}\right] = -\frac{\hbar^2}{m}\sum_{i=1}^{N}\int \psi_i^*\nabla^2\psi_i d^3r + \int V(\mathbf{r})n(\mathbf{r})d^3r + \frac{e^2}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{ion} \tag{1.14}$$

where the terms are the electron kinetic energies, the Coulomb interactions between the electrons and the nuclei, the Coulomb interactions between pairs of electrons, and the Coulomb interactions between pairs of nuclei. All other quantum effects that are not counted in the terms calculated in E_{known} is counted in the term E_{XC}, the exchange-correlation functional, compared to the true functional. In practice, the realization of DFT is based on the Kohn-Sham (KS) approach which is a variant of Hartree-Fock method [76]. The KS equation is the non-interacting Schrödinger equation of a fictitious system of non-interacting electrons that generate the same density as any given system of interacting electrons. Hence, the right electron density can be expressed in a way that involves solving a set of equations in which each equation only involves a single electron. The KS equation can be written as

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \tag{1.15}$$

where ϵ_i is the energy according to the single electron wavefunction $\psi_i(\mathbf{r})$. The potential noted as $V(\mathbf{r})$ is the interaction between the single electron and all nuclei in the system. The V_H, Hartree potential, describes the the Coulomb repulsion between the single electron and the total electron density defined by all electrons in the system

$$V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r'. \tag{1.16}$$

The V_{XC} defines exchange and correlation contributions to the single electron equations and can be also defined as derivative of the exchange-correlation energy

$$V_{XC}(\mathbf{r}) = \frac{\partial E_{XC}(\mathbf{r})}{\partial n(\mathbf{r})}. \tag{1.17}$$

To solve the KS equation, an initial, trial electron density $n_0(\mathbf{r})$ is needed. Using this density, a single electron wavefunction can be calculated according to the KS equation (1.15). Then, a new electron density $n_1(\mathbf{r})$ can be calculated as Equation 1.12. In the next step, we compare the two electron densities to see whether they are the same density. If same, the density is the ground state electron density and we can use it to calculate ground state energy. If not, another loop of calculating density and wavefunction will be implemented until the densities from two adjacent loops are the same. In a practical way, a convergence criterion will be set in the unit of energy, and if the energy difference between two adjacent loops falls below the pre-set criterion, we will consider it as the true ground state electron density. Because it is impossible to calculate the exactly true density in a DFT calculation.

1.2.4 Ab initio molecular dynamics

While DFT calculations provide accurate information regarding electron configuration and system energy, they do not adequately capture the dynamic nature of a molecular system. In particular, in materials with flexible structures, the inherent dynamical motions of the material can significantly influence their physical and chemical

properties. Therefore, employing molecular dynamics (MD) simulations becomes crucial in investigating the dynamic behaviors of such systems [65]. The MD method, widely employed a to investigate various systems, involves numerically solving the classical Newton equations of motion for a given system, considering appropriate boundary conditions and an initial state. MD simulations provide a means to compute both equilibrium thermodynamic properties and dynamical behavior of the system at finite temperature. The accuracy and reliability of MD simulations significantly depend on the specification of forces within the system. In practice, these forces are often determined using an empirical model known as a "force field." The force field approach has demonstrated remarkable success in accurately describing the behavior of a diverse range of systems. There are several different types of force fields, in the most common format, it invovles bonded potentials, angle potentials, torsion potentials, and nonbonded terms

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{nonbond}. \tag{1.18}$$

These terms capture the interactions and energies associated with bonded atoms, angles, torsions, as well as nonbonded interactions, including van der Waals and electrostatic forces [1]. In many cases, conventional force fields used in MD simulations do not explicitly incorporate electronic polarization effects and may have limited capability in accurately capturing chemical reactivity. Consequently, the methodology of *ab initio* molecular dynamics (AIMD) is frequently employed as an alternative approach. AIMD utilizes quantum mechanical methods, such as DFT, to calculate the forces and energies on atoms and molecules during the simulation. These calculations are based on the fundamental laws of quantum mechanics and consider the electronic structure and interactions explicitly as was explained in the previous sections. By incorporating electronic polarization effects and accurate treatment of chemical reactivity, AIMD provides a more rigorous description of the system's

behavior. AIMD simulations allow for the investigation of properties, reactions, and processes that depend crucially on electronic polarization and quantum mechanical effects. Although AIMD simulations are computationally more demanding compared to classical MD simulations, they offer valuable insights into the behavior of complex systems, particularly in cases where accurate treatment of electrons, their polarization and charge transfer is essential.

1.3 Data-Driven Technology for Material Discovery

Novel chemical compounds or materials with extraordinary properties are synthesized and applied in all aspects of our society. So far, scientists have created more than 10^8 materials for various application purposes since modern chemistry was established [9]. In the past century, scientists used to control reactions by imparting a desired property in a material using external stimuli. Nowadays, designing a material with desired property for a target application becomes more and more challenging because we are expecting a novel material to meet our needs in different aspects simultaneously. For example, for a novel battery material we expect it to show high thermal stability, high energy density, and high power density while being environmentally friendly and cheap. This means not one but many properties are need to align properly so that the material can be utilized in real applications. Different properties are annoyingly correlated which means we cannot treat them individually. Although great achievements have been made in designing materials, most of experiemnts along the way rely on the "trial-and-error" process which require vast investment of time and money [95]. The implementation of experiments depends on human knowledge, chemical intuition and low throughput trials [9]. Just like we cannot read every paper in one research area, there is a limit for human knowledge and recognition. Efficient and comprehensive search for novel materials require a more automated manner that

utilizes the power of computational chemistry, data science and artificial intelligence (AI) to be applied on a vast collection of materials, i.e., a database, at the same time.

1.3.1 Development of material databases

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. To get a board view of what we already have in the 10^8 materials, we need a container to collect the information of them which we call a material database. The earliest material database should be the Cambridge Structural Database (CSD) launched in 1965 [46] which provides a repository of experimental crystal structure information. Till now, CSD has become a well-developed database that contains over 1.2 million experimental structures. Such huge amount of data provides us the ability to screen the results, then find the best path for our goals. Inorganic Crsytal Structure Database (ICSD) was founded in 1970s at the Institute for Inorganic Chemistry of the University of Bonn, Germany [167]. The number of inorganic crystals now has reached over 281,000. Another well-known database, Material Project (MP), is initiated in 2013 and now there are 154,718 materials that are contributed from researchers all over the world [67]. Other than the databases that cover a broad range of materials, databases for specific classes of materials are also often reported. The Protein Data Bank provides structural data of biological macromolecules. Computational 2D Material Database (C2DB) includes mainly two-dimensional materials [10, 119]. Differing from the database from experimental results, there are Quantum MOF (QMOF) database and Open Quantum Materials Database (OQMD) that provides results based on quantum mechanical calculations [122, 126]. The development of different material databases greatly help and boost the material design and discovery by providing enormous data and information. The field of computational chemistry

and computational assisted materials discovery is rapidly growing by combining the advantages of material databases and tools from data science.

1.3.2 High-throughput computing

High-throughput (HT) methods have been applied in both experimental and computational research. The first systematically combinatorial synthesis was reported decades ago which shows the idea of HT methods [9]. HT processes conduct automation experiments in a feasible large-scale repetition to find the best material from an enormous number of possible candidates. With the development of material databases, a tool that can comprehensively screen all data points is needed due to the limitation of manual search as depicted in the comic below by Benayad et al. where scientists are looking for a needle in a haystack [9]. In computational

Figure 1.6 Classical experimental approach versus HT experimentation approach. *Copyright 2021 Wiley Online Library.*

chemistry, power of HT can greatly increase the efficiency of research by avoiding the manual calculations and extracting results because computational researchers always deal with hundreds or thousands of materials at the same time. Establishment and implementation of HT workflow will be explained in Chapter 3 in detail.

1.3.3 Machine learning techniques

Quantum mechanical calculations are a powerful tool with good accuracy, but they are still expensive in term of CPU time. For example, we need to use 4403 SUs to calculate the electrical conductivity of a system with 100 atoms. Translating this to a temporal context, if a single CPU is utilized, it would necessitate an estimated duration of 4403 hours (183 days) to complete the calculations. It is impossible and unnecessary to iterates such expensive calculations over all structures in a database. A viable alternative is using machine learning (ML) technique [141] to save time and computing resources. ML is a sub-field of AI that focuses on the development of algorithms and models which are capable of learning patterns and making predictions or decisions based on input data. In the field of materials science and discovery, ML has emerged as a powerful tool for analyzing large datasets and learning valuable insights to guide the design of new materials. A dataset which can be treated as a matrix consists of m rows of instances and in each row (instance), there are columns of data values. For a specific research purpose, we recognize one column of values as the target of each instance that we want ML model to predict, then we select n columns (features or descriptors) from the rest to form our training data which can be expressed as $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^m]^T$ where x is a instance. For each instance, $\mathbf{x}^i = [x_1^i, x_2^i, \cdots, x_n^i,]$. Each instance will have a target data y so that a $m \times 1$ vector, $\mathbf{y} = [y^1, y^2, \cdots, y^m]^T$, represents the target data of all instances. Furthermore, we need a hypothesis function $h(\mathbf{x}^i)$ to compute a value according to the input features of i-th instance. We have different forms of hypothesis in ML models and one of them, a linear function, has the following form:

$$h(\mathbf{x}^i) = \theta_0 + \theta_1 x_1^i + \theta_2 x_2^i + \cdots + \theta_n x_n^i \tag{1.19}$$

where θ_i is the parameter that will be optimized during training which can be integrated to a vector with $n+1$ dimension, $\theta = [\theta_0, \theta_1, \cdots, \theta_n]$. The error of i-th

instance is:

$$h(x^i) - y^i \tag{1.20}$$

For m instance, the total error of hypothesis $h(\mathbf{X})$ can be computed by a summation which we call it loss function:

$$J(\theta) = \sum_{i=1}^{m} f(h(\mathbf{x}^i) - y^i) \tag{1.21}$$

where function $f()$ is an extra function that converts the computed error into different forms applied by different ML models. The final goal is to minimize the loss function J so that the computed values (predictions) from the hypothesis get close to the target values. In most of ML models, predictions are made as the principles above and the main difference is the form of the hypothesis and loss function.

Tasks of ML can be simply categorized into two parts, supervised and unsupervised learning. Other learning method like semi-supervised learning, reinforced learning or active learning can be adopted depending on the usage. Supervised learning indicates ML models are trained by labeled datasets. This category of tasks can be further divided into two types according to the forms of the target. If we need the ML model to classify the instances in a dataset into specific number of classes, then it is a classification type. If we need the ML model to compute a value for each of the instance, it should be a regression type. Unsupervised learning, on the other hand, uses unlabeled datasets and focuses on separating instances into different classes according to their features without users supervision. As shown in Figure 1.7, a complete implementation of ML can be divided into three steps: (1) data pre-processing, (2) Training and testing the model and (3) make predictions. Before we train an ML model, data pre-processing is needed because different numerical scale, significant numbers and distribution between features can cause a misleading during minimizing the loss function. Also, too many number of features, that may contain

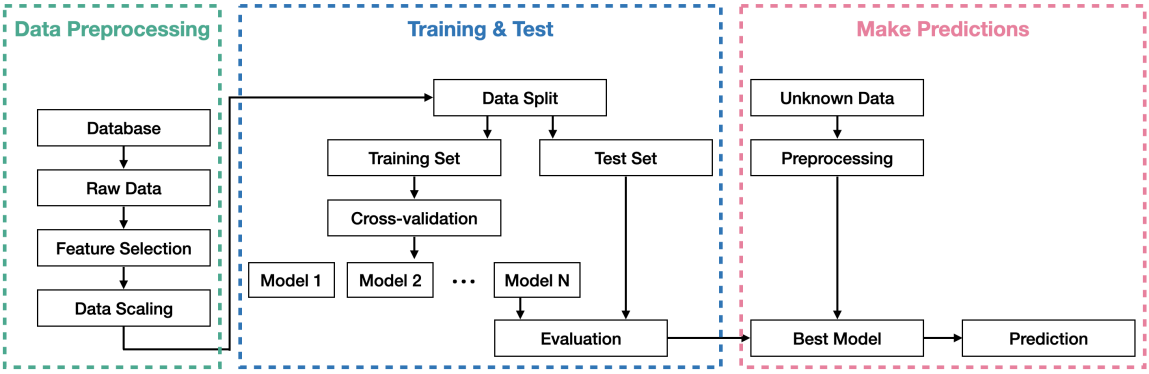

Figure 1.7 A machine learning workflow includes different steps of data preprocessing, training, testing and evaluating the model.

replicated features or highly similar features, can cause serious over-fitting problem so that a feature selection process is necessary. In the second step, the processed data will be split into training set and test set at a fixed ratio. Training set is used to train the model and find the best set of θ that minimize the loss function. To avoid the error induced by splitting, cross-validation is adopted to assess the model performance more objectively. During cross-validation, the training set is further evenly divided into k smaller sets and we call such situation as k-fold cross-validation. A 10-fold cross-validation is shown in Figure 1.8. For each one out of the ten small sets is used

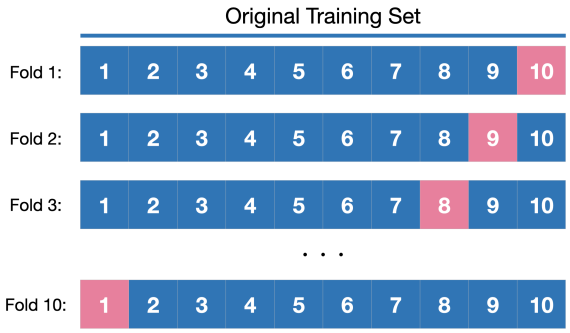

Figure 1.8 An illustration of a 10-fold cross-validation.

as validation set to assess the performance and the training will be repeated 10 times. The final performance is computing by taking the average of ten training loops. After the loss function is minimized according to the training set, the test set is used to

assess the performance of the model. When implementing ML in research, multiple ML models will be attempted and the best model will be selected according to the performance scored by the test set.

CHAPTER 2

METAL-TO-SEMICONDUCTOR TRANSITION IN TWO-DIMENSIONAL METAL-ORGANIC FRAMEWORKS

2.1 Discrepancy in the Literature on Conductive Behavior of 2D MOFs

Tuning electronic properties of 2D MOFs for any specific application can be a complicated challenge due to the flexibility of the layered structure held together via rather weak van der Waals interactions. We have recently characterized three categories of dynamical motions in 2D MOFs as well as deformation sites [104, 133] which can directly affect both intra- and inter-layer charge transport. The continuous strive for unravelling the electronic characteristics of 2D MOFs is most illustrated in the case of $Ni_3(HITP)_2$, HITP= 2,3,6,7,10,11-hexaiminotriphenylene (Figure 2.1), which was first synthesized and introduced as a semiconductor by Dincă and coworkers [131]. While all following theoretical studies categorized $Ni_3(HITP)_2$ as metallic [20, 38, 171], it was utilized to build the first 2D MOF-based field-effect transistor (FET) device emphasizing on its semiconducting character [155]. A subsequent theoretical study suggested that the discrepancy between theory and experiment is rooted in the presence of defects in experimental samples while theoretical studies were carried out on pristine models [37]. Specifically, it was estimated that manually displacing 2D layers with respect to each other would weaken the $\pi - \pi$ stacking and as a result an energy gap of 15-100 meV would appear along the interlayer direction [37]. This is not the end of story though. A very recent experimental study by Day et al. reported the synthesis of metallic $Ni_3(HITP)_2$ single rods [30] while still emphasizing on semiconducting nature of $Ni_3(HITP)_2$ pellets reported in 2014. [131] In short, $Ni_3(HITP)_2$ showcases the existence of a very close structural-electronic properties relationship in 2D MOFs. Another historical example is $Ni_3(HIB)_2$, HIB = hexaiminibenzene, which was synthesized as the first metallic 2D MOF in 2017 [35] while recent reports emphasize on its intrinsic semiconducting nature with a

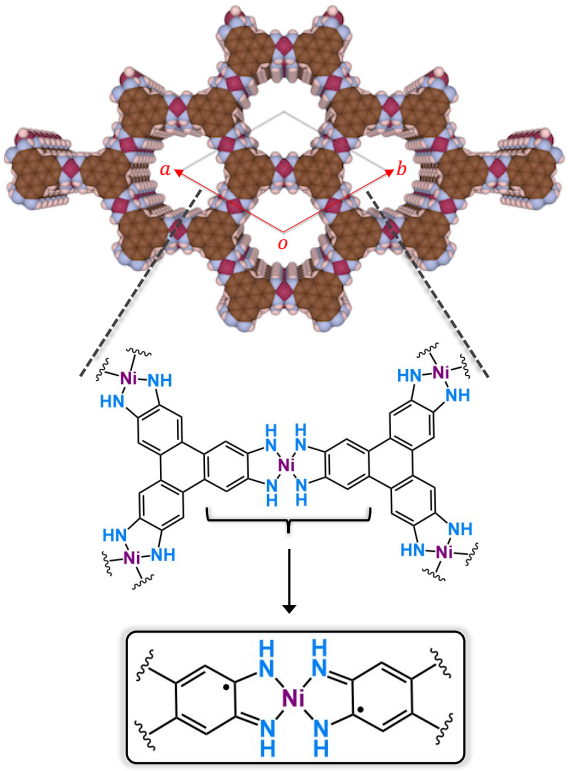

Figure 2.1 Layered structure of $Ni_3(HITP)_2$ (HITP = 2,3,6,7,10,11-hexaiminotriphenylene) 2D MOF with the unit cell and HITP linkers highlighted.

band gap of 0.49 eV. [57] The message of the short history of conductive 2D layered MOFs briefly stated above is not discrepancy but, a rich material genome that offers a spectrum of potential applications. The pressing challenge here is to have a road map for this class of materials whose electrical conductivity is largely influenced by their flexible and varying structure rather than composition. In pursuit of such road maps, here, we leverage a variety of computational toolbox to (1) provide an appropriate input crystal structure and density functional for electronic structure calculations; (2) characterize the electronic and structural features affecting charge transport pathways/mechanisms; and (3) unravel the correlation between temperature-induced flexibility of 2D MOFs and their intrinsic electrical conductivity. We show that the flexibility, originating from a variety of intrinsic dynamical motions and deformations, affect the intra- and inter-layer conductivity to different degrees. In this work, we

study $Ni_3(HITP)_2$ as an archetypal conductive 2D MOF but the produced physical insights by no means are restricted to this material. Computational details of this are presented in the next section. Section 2.3 provides the results and the associated discussions. Future outlooks are provided in Section 2.4 with our main findings summarized in Section 2.5.

2.2 Computational Details

2.2.1 Input crystal structure

Electronic band structure calculations of 2D MOFs are prone to significant variability due to the lack of single-crystal information since almost all experimental measurements have been carried out on polycrystalline pellets. [158] Computational studies have relied on building and minimizing a monolayer of the 2D MOF, then stacking similar layers and manually moving them along the stacking direction as well as the *ab* plane in order to approximately reproduce the unit cell features inferred from powder X-Ray diffraction measurements. [20, 21, 35, 37, 38, 131] Such approaches [35, 57] are susceptible to numerous errors as one can not guarantee accurate probing of the whole phase-space and visiting all the stationary points in order to locate the true local minima on the potential energy surfaces of the studied 2D MOFs. We propose to avoid any adcohism in electronic band structure calculations by starting from the only experimentally available single-crystal for 2D MOFs which is the hexagonal crystal structure of Co_3HHTP_2 [58] (HHTP = 2,3,6,7,10,11-hexahydroxytriphenylene). We built a crystal structure shown in Figure 2.2 comprised of two layers and a total of 150 atoms for $Ni_3(HITP)_2$ MOF by (1) removing the alternate $Co_3(HHTP)(H_2O)_{12}$ layers, (2) removing all solvent (water) molecules and hydration water from the $Co_3(HHTP)_2(H_2O)_6$ layers, and (3) replacing Co^{2+} with Ni^{2+} and hexahydroxytriphenylene linkers with hexaiminotriphenylene.

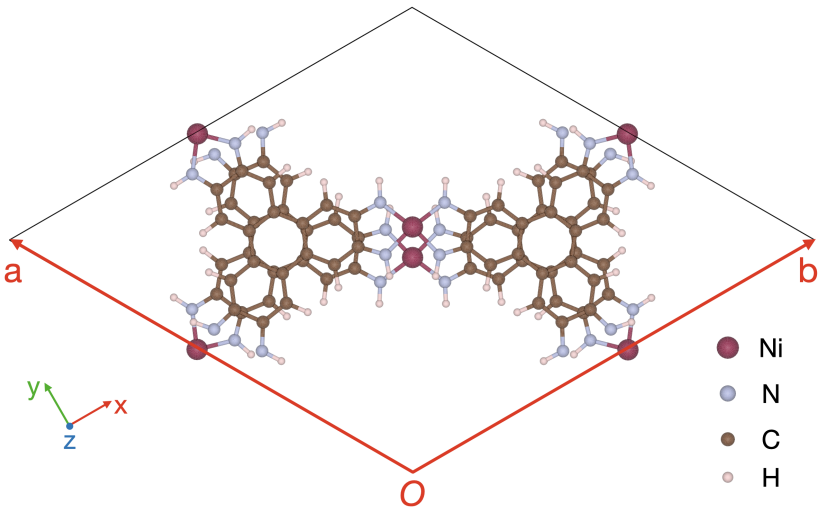

Figure 2.2 Optimized structure of $Ni_3(HITP)_2$ MOF by PBE-D3 functional.

Next, periodic electronic structure calculations were performed to minimize both cell vectors and atomic positions using the periodic boundary conditions. This brings us to the next computational dilemma in dealing with 2D MOFs which is the choice of an appropriate exchange-correlation functional. Since the layers in bulk 2D MOFs are stacked via rather weak van der Waals interactions, the correct description of these interactions is of utmost importance. Therefore, we conducted a benchmark study on minimized cell vectors and atomic positions obtained using different exchange-correlation density functionals starting with the widely employed Perdew-Burke-Ernzenhof (PBE) [114] functional within the generalized gradient approximation (GGA) formalism.

In addition, we tested the performance of the revised version of PBE (RPBE) [49], PBE for solids (PBEsol) [116] and the non-local van der Waals density functional (vdW-DF) [5, 86, 129] combined with a series of dispersion correction schemes in the form of PBE functional with Grimm's damped D2 and D3 dispersion corrections [44]. The double-zeta valence with polarization (DZVP-MOLOPT) basis set and core-electron pseudo potentials according to the Geodecker-Teter-Hutter (GTH)

Table 2.1 Computed Cell Vectors (a and c in Å) as well as Interlayer (d in Å) and Slipping Distances (d_s in Å)[1] in $Ni_3(HITP)_2$ 2D MOF Using Different Density Functionals and Dispersion Correction Schemes and the DZVP-MOLOPT Basis Set

XC functional	a (Å)	c (Å)	d (Å)	d_s (Å)
PBE	21.923	16.912	8.552	0.264
RPBE	22.460	17.016	9.407	0.570
PBEsol	22.272	18.283	9.002	0.543
vdW-DF	22.037	8.067	3.810	0.716
PBE-D2	21.774	6.493	3.223	1.472
PBE-D3	21.774	6.644	3.299	1.441
Sheberla et al.	21.75	6.66	3.33	∼1.8
Day et al.	NA	NA	3.3	NA

[1] $d_s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ where x_i and y_i are coordinates of Ni atoms in two different layers.

formulation [41, 51, 77] was used as implemented in CP2K version 7.1 [64]. The energy cutoff is set to 500 Ry and the convergence criterion of self-consistent field (SCF) computation is set to 1×10^{-5}. Table 1 lists the computed cell vectors, inter-layer distances (d), and slipping distances of two layers with respect to each other (d_s) in the ab plane obtained using all considered density functionals compared to the available experimental data. Pure GGA functionals, i.e., PBE, RPBE, and PBEsol, result in similar a vectors to the 21.75 Å value inferred from PXRD measurements [131]. However, all three functionals overestimate the computed c vectors by 10 Å ($\approx$ 17 Å compared to 6.66 Å from PXRD, Table 2.1) due to not accounting for dispersion corrections in these layered materials. The vdW-DF functional provides a better accuracy than the PBE series [73], resulting in a computed c vector of 8.067 Å and d of 3.810 Å (Table 2.1). Adding Grimm's damped D2 and D3 dispersion corrections [44] to the PBE functional produces even better agreement with experimentally

inferred results than vdW-DF reducing d to the striking range of 3.223 and 3.299 Å, respectively. Comparison of d_s values of ∼1.45 Å shows that these two functionals have the closest agreement with the experimentally inferred value of ∼1.8 Å[131]. Overall, based on these results, we choose PBE-D3 as the method of choice throughout this work.

2.2.2 Periodic ab initio molecular dynamics simulations

Accurate simulation of structural dynamics in flexible materials such as $\pi-$stacked 2D layered MOFs has many implications on their chemical stability and reactivity in both gas phase and aqueous solutions. Different electrical properties such as band gaps and band structures obtained using static 0 K optimized structures completely neglect the effects of external stimuli such as temperature on overall dynamical flexibility and electrical conductivity of the studied material. In order to provide a realistic estimation of the band gap and electrical properties of flexible layered 2D MOFs subject to temperature-induced structural dynamics we performed periodic ab initio molecular dynamics (AIMD) simulations starting from the $1\times1\times1$ optimized unit cells at the DFT level using our elected dispersion corrected PBE-D3 density functional. By comparing electrical properties of the resulted structures from AIMD simulations to the ones obtained from the 0 K static calculations one can disentangle the effect of incorporation of structural dynamics on the overall electrical properties of these highly flexible materials. For these simulations, the experimentally determined temperature and pressure of 293 K and 1 atm with a time step of 1 fs and a total production time of 8 ps were used. A lower cutoff of 300 Ry was used in all AIMD simulations to make them more tractable. The equations of motion were propagated in the isobaric-isothermal (NPT) ensemble to allow the simulation boxes to vary and hence the layers move along the c direction (breathing movement) and along the ab plane (slipping movement). The temperature was controlled by using the canonical sampling through

velocity rescaling (CSVR) thermostat [16]. Analysis of the trajectories reveals that the thermodynamic equilibrium is reached within the first 1 ps of the simulations as shown in Figure 2.3. All periodic AIMD simulations were performed using CP2K version 7.1 [64].

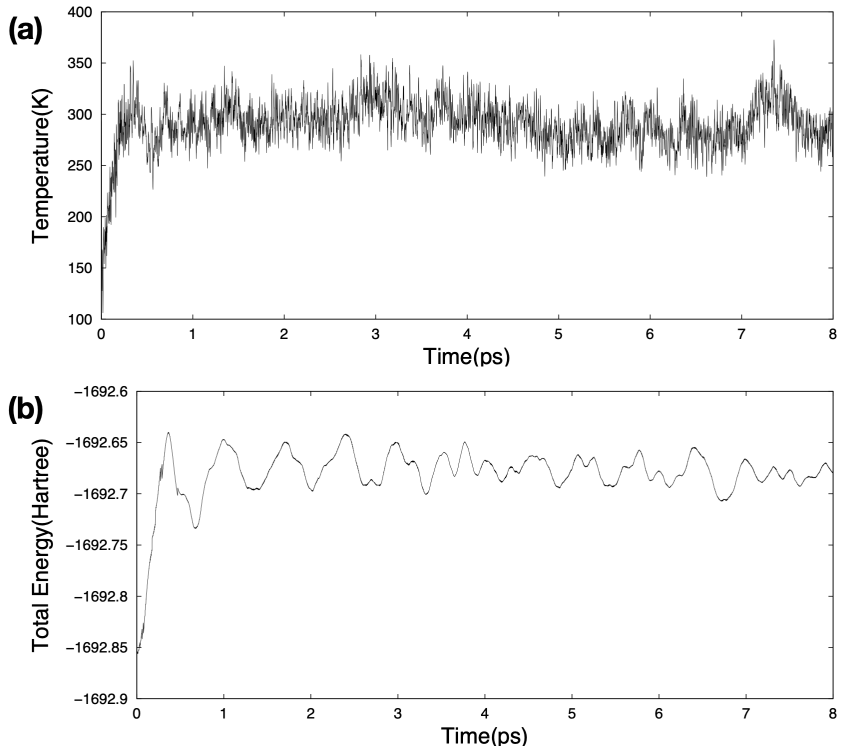

Figure 2.3 The temperature (a) and energy (b) equilibriums are reached after 1ps during the AIMD simulation.

2.2.3 Electronic band structure calculations

Periodic electronic structure calculations were performed at the PBE-D3 level as implemented in Vienna ab initio simulation package [80, 81, 78, 79]. Interactions between electrons and ions were described by Projector Augmented Wave (PAW) potentials [12, 82] with the energy cutoff of PAW potentials set to 500 eV. Hubbard U approach (DFT+U) is adopted to treat the d states of the Ni transition metals. In this study, only the difference between coulomb term U and exchange term J,

defined as $U_{eff} = U - J$, is taken into consideration with the value of U_{eff} set to 6.7 eV for Ni [98] and 10.0 eV for Zn [151]. Gaussian smearing was adopted in all calculations with a smearing width of 0.05 eV. The convergence criteria were 10^{-5} for SCF calculations and 10^{-6} for electronic property calculations. The k-point mesh in the Monkhorst-Pack scheme was set to 2×2×6 in the SCF part and twice denser in following calculations for the hexagonal unit cell. Electronic band structure calculations adopt the k-path in Figure 2.4. Spin polarized calculations were performed for all systems.

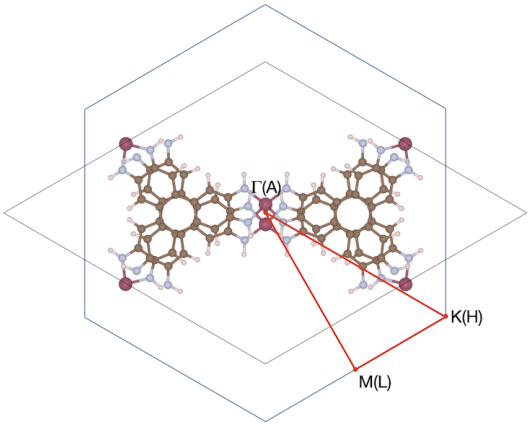

Figure 2.4 Phase space paths for $Ni_3(HITP)_2$ system. Same paths were used for $Zn_3(HITP)_2$.

2.3 Results and Discussion

To examine the relationship between flexibility and charge mobility in 2D MOFs, and to provide a guideline for fine-tuning MOF band gap openings, we compare the band structures as well as atomic-orbital projected density of states (pDOS) at 0 K vs. 293 K. First, it should be noted that the extension of layers in the ab plane vs. their stacking in the c direction leads to extended-conjugation vs. through-space charge transport (CT) pathways. Intra-layer extended-conjugation in 2D MOFs is the result of pairing transition metals with organic linkers containing

chelating functional groups, such as hexa-substituted triphenylenes and benzenes. Conjugation of the transition metal d orbitals with the extended $\pi - \pi$ system of the functional groups and organic core of the ligands has been known to be the main route of charge delocalization in 2D MOFs [131, 18, 62, 35, 32, 30]. However, it is not easy to disentangle extended-conjugation from through-space CT pathway between layers especially for triphenylenes linkers with a large π-conjugation that might favor the latter mechanism over the former. Although recent studies are paying more attention to through-space conductivity in 2D MOFs [135], the role of inter-layer $\pi - \pi$ interactions in charge delocalization of the bulk structures remains to be fully elucidated. Our analysis of $Ni_3(HITP)_2$ electronic band structures at 293 K reveals the rather high sensitivity of the electrical conductivity to thermally-induced structural deformations and displacements of layers. Since these dynamical motions can affect both intra- and inter-layer CT pathways, first we disentangle the contributions of extended $\pi - d$ conjugation and through-space $\pi - \pi$ interactions in charge delocalization subject to these geometrical features. We project these contributions on different building blocks of 2D MOFs, namely organic linkers and transition metal centers, to investigate how the character of the corresponding frontier orbitals correlate the flexibility of the framework to its electrical conductivity.

2.3.1 Metallic/semiconducting boundary: dependence on the nature of the frontier orbitals

The PBE-D3 optimized $Ni_3(HITP)_2$ at 0 K shows slipping of layers with respect to each other in the ab plane which is consistent with its previously characterized slipped-parallel stacking configuration [131, 20]. The layers are nearly planar allowing for an extended $\pi - d$ and $\pi - \pi$ conjugation along the layers. The C-N-Ni-N dihedral torsion of the two ligands around the central Ni atom in the ab plane is very close to $180 \sim 177 leading to nearly square planar SBUs, see Figure\ 2.5(a)$.

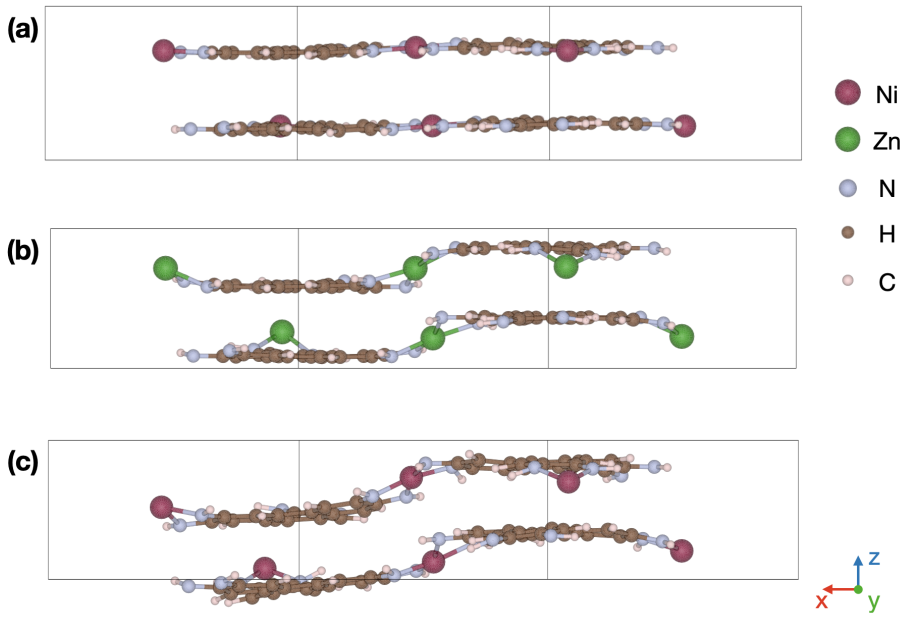

Figure 2.5 The optimized structure of (a) $Ni_3(HITP)_2$ and (b) $Zn_3(HITP)_2$ 2D MOF at 0 K. (c) Representative equilibrated $Ni_3(HITP)_2$ structures at 293K.

Figure 2.6a illustrates the band structures and projected density of states (pDOS) diagrams of $Ni_3(HITP)_2$. We note that although similar calculations have been already reported, e.g. in Ref. [38], however, the employed structures there were obtained by manually displacing the layers with respect to each other in the *ab* plane without minimizing the cell vectors and atomic positions afterward. We, on the contrary, obtained these results from a careful construction of an input crystal structure and minimizing both the atomic positions and cell vectors using our validated PBE-D3 functional, hence, avoiding any bias in the obtained results. As such, we provide a full explanation of our results which based on this robust methodology paint a different detailed picture of the band structure and pDOS diagrams of $Ni_3(HITP)_2$ compared to previous studies. Bands in MKΓ region of Figure 2.6a exhibit a relatively large dispersion away from the Fermi level yet, the conduction band minimum (CBM) is found to drop below the valence band maximum (VBM) for $k \to 0$ along the K-Γ line. This aspect is different from bands in Ref. [38]

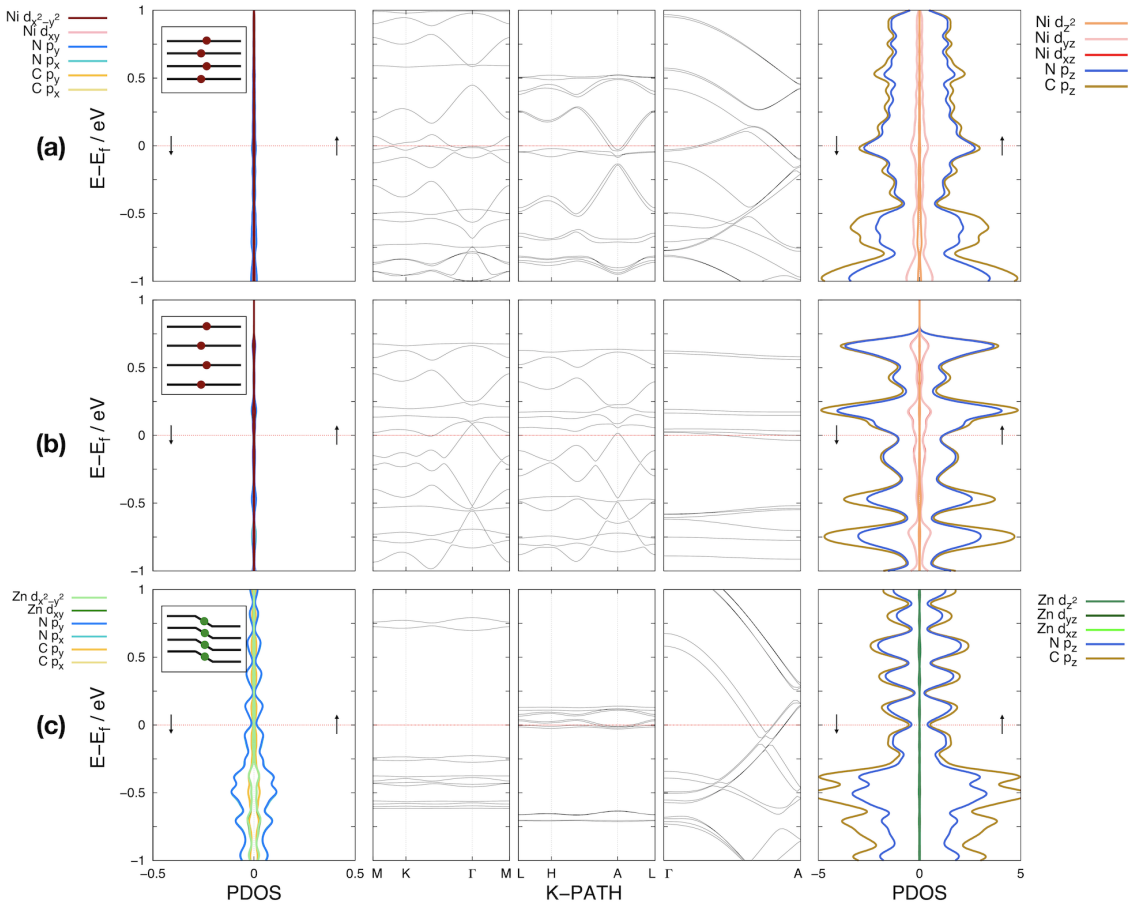

Figure 2.6 Band structure diagrams and pDOS profiles of (a) $Ni_3(HITP)_2$ 2D MOF at 0 K; (b) convergence to the $Ni_3(HITP)_2$ mono-layer by increasing the inter-layer distance to 5 Å; and (c) exclusion of the extended $\pi - d$ conjugation in $Zn_3(HITP)_2$ system at 0 K. The computed pDOS profiles are separated into two groups of orbitals with z component, far right panels, and orbitals with x and y character, far left panels. The insets in the far left panels are schematic representation of planar layers in $Ni_3(HITP)_2$ vs. stepped layers in $Zn_3(HITP)_2$. The arrows ↑ and ↓ represent spin-up and -down, respectively.

which show two conically-shaped valleys meeting at $\sim E_0$ close to the corner of the first Brillouin zone (BZ), similar to Dirac points observed in graphene bands [125, 31]. In contrast, our band diagram displays a semimetallic signature [72] for an indirect band gap opening, qualitatively equivalent to a semiconducting signature with the difference that VBM and CBM do not energetically overlap. In this regard, we point out that although the semimetallic/metallic boundary in these structures is difficult to define, from a semiconducting point of view the band gap can be easily tuned, for instance by slightly varying the inter-layer distance. Figure 2.6a also demonstrates the band structure of $Ni_3(HITP)_2$ at the LHA region where, again, we can see a semimetallic behaviour. Here, we notice another difference between this band diagram and the one reported in Ref. [38] as we find an intra-band splitting at the minimum of the conduction band. This effect probably is due to the spin degeneracy lift originated from spin–orbit coupling (SOC) under symmetry breaking along the LHA line (i.e., because of the slipping of layers in the *ab* plane), resulting in Rashba effect activation [173]. In fact, in the case of metal halide perovskites, it is shown that when the SOC is enhanced the Rashba effect may decrease the minimum of the conduction band resulting in a smaller band gap energy [70]. This topic goes beyond the scope of the present work but will be discussed in more details in our future studies. The far right panel of Figure 2.6a shows that the pDOS near the Fermi level is almost fully contributed by the p_z orbitals of the C and N atoms as well as the delocalized d orbitals of the Ni transition metals. A closer look at frontier orbitals shows that the two d_{yz} and d_{xz} orbitals of Ni which lie almost along Ni−N bonds, see the figure on top of Table 1, are the ones contributing to the Fermi level and create an intra-layer extended $\pi - d$ conjugation CT pathway. This is due to the close to perfect planar layers in $Ni_3(HITP)_2$ at 0 K which provides optimum overlap between the transition metal d_{yz} and d_{xz} orbitals and the p_z orbitals of the organic linkers.

Focusing on the out-of-plane band structure of $Ni_3(HITP)_2$ in Figure 2.6a, one can see large band dispersions within the $\Gamma-A$ region indicating strong $\pi-\pi$ interactions between layers as expected from the triphenylene linkers with their large π systems. As mentioned before, the role of through-space $\pi-\pi$ interactions on CT in 2D MOFs is not as widely acknowledged as extended $\pi-d$ conjugation. To disentangle these two competing CT pathways we analyzed the electronic structure of two extreme cases. Figure 2.6b illustrates the band structure and pDOS diagrams of the optimized $Ni_3(HITP)_2$ at 0 K when the inter-layer distance is increased from the original 3.299 Å to 5 Å in an attempt to suppress the $\pi-\pi$ interactions between the layers. Figure 2.6c exhibits the band structure and pDOS diagrams of an alternative framework where Ni^{2+} atoms are replaced with Zn^{2+} creating $Zn_3(HITP)_2$ 2D MOF. Here, in contrary to the d^8 electronic configuration of Ni^{2+} in square planar coordination environment the filled d^{10} configuration of Zn^{2+} causes the SBU to act as a nodal plane separating the redox active triphenylene linkers and thereby disrupting the extended $\pi-d$ conjugation and intra-layer CT pathway. Analysis of the electronic structure subject to increasing the inter-layer distance along the c direction has been previously carried out on a crystal structure where the layers were manually displaced in the ab plane [38]. It was shown that both valence and conduction band minima locations may change by increasing the inter-layer distance, suggesting a possible strategy to tune the semiconducting properties in $Ni_3(HITP)_2$, for instance by flattening valence and conduction bands. Our calculated band diagram at 5.0 Å, Figure 2.6b still shows band dispersions within the $\Gamma-A$ region indicating that van der Waals interactions between layers has a much longer range than what presumed before. It can be said with certainty that unlike graphene, the bulk structure of 2D MOFs with hexa-substituted triphenylene linkers possess relatively good conductivity in both the intra- and the inter-layer directions. Nevertheless, flattening of the valence and conduction bands in the $\Gamma-A$ region is promising a narrow band-gap

semiconductor should the inter-layer distance be increased beyond 5.0 Å where the structure will eventually converge to the mono-layer level. Further analysis of the band diagrams shown in Figure 2.6b reveals that increasing the inter-layer distance not only affects through-space but also extended-conjugation CT pathway. The band structure along the LHA region shows a semiconducting signature where there is a gap opening between the top of the valence band and the bottom of conduction band at the point A. Therefore, our findings show that a structural change in the stacking direction, let it be by deformations, slipping of layers or change of the inter-layer distance, affects the charge mobility through-space and through extended-conjugation pathways, suggesting a way to adjust or even switch between the intra-layer vs. inter-layer conductive anisotropy in 2D MOFs. We will show in the next section that the intrinsic dynamical motions in 2D MOFs turn these structural changes to an inherent characteristics of these materials which can greatly affect their overall electrical conductivity.

Comparison between optimized structures of $Ni_3(HITP)_2$ and $Zn_3(HITP)_2$ shows a stark difference in stacking of the layers. While these layers are almost planar in $Ni_3(HITP)_2$ they adopt a stepped form in $Zn_3(HITP)_2$ due to the absence of $\pi - d$ conjugation, see schematic representations in the insets of Figure 2.6c and optimized structures in Figure 2.5(b). As shown in Figure 2.6c, one can see a band-gap opening as big as $\sim$1 eV in the intra-layer MKΓ region, clearly signaling a semiconducting nature for this material in the *ab* plane. Interestingly, the lack of the intra-layer $\pi - d$ conjugation is accompanied by stronger inter-layer $\pi - \pi$ interactions which causes 0.5 Å shrinking in the optimized *c* vector compared to the analogues $Ni_3(HITP)_2$ system (i.e., from 6.644 Å to 6.143Å). This confirms the interconnection/competition between the two CT pathways where weakening of one may result in strengthening of the other. Out-of-plane band diagram shows that deep valence bands cross the Fermi level in the $\Gamma-$A region demonstrating lower effective masses of charge carriers than

$Ni_3(HITP)_2$ and consequently higher charge mobility in the c direction induced solely by through-space $\pi - \pi$ interactions between organic linkers. This, again, emphasizes on the important role of $\pi - \pi$ interactions on the bulk electrical conductivity of 2D MOFs.

2.3.2 Localized charge states and isotropic behaviour analysis at finite temperature

As we have shown recently [133, 104], $\pi-$stacked layered 2D MOFs are flexible materials with a wide range of dynamical motions especially in the presence of an external stimuli such as temperature, pressure, electric and/or magnetic field. Even the optimized $Ni_3(HITP)_2$ at 0 K shows layers that are slipped in the ab plane ($\sim$1.44 Å) and a small twist of ligands around the central Ni site ($\sim$177 Here, AIMD simulations were carried out at 293 K temperature and 1 atm pressure on $Ni_3(HITP)_2$ in the NPT ensemble (constant pressure/constant temperature), allowing the simulation box to vary, for a total production time of 8 ps. The layers in the equilibrated system show a stepped geometry unlike the planar geometry of the 0 K optimized structure, see Figure 2.5(c), with a more vivid range of dynamical motions. Figure 2.7 demonstrates schematic representation of these motions as well as their statistical analysis in Cartesian coordinates during the entire 8 ps AIMD simulations. As evident in Figure 2.7a the in-plane slipping motion of the layers, represented by the position of the Ni atoms in each layer, is a continuous movement and can occur up to an extra 1 Å displacement in the x direction and up to 1.5 Å along the y direction. Another motion revealed in the equilibrated system is the breathing of the layers in the stacking direction as depicted in Figure 2.7b. As pointed out before, changes in the inter-layer distance has a great impact on both the in-plane and out-of-plane electrical conductivity and hence one can envision the importance of including temperature and pressure in the simulations in order to paint a more realistic picture of CT pathways and mechanisms in these highly flexible materials.

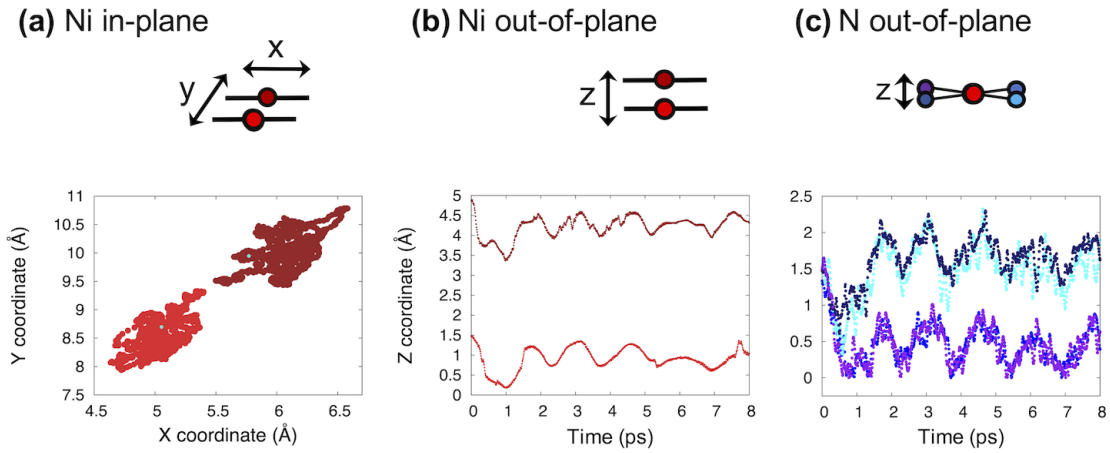

Figure 2.7 In-plane vs. out-of-plane motions of the secondary building units (SBUs) in $Ni_3(HITP)_2$ during an entire 8 ps AIMD simulation. (a) Motion along the x and y axes for two Ni atoms belonging to the first (red) and second (brown) layer. Initial positions are shown as green points. (b) Out-of-plane motion of the Ni atoms in the first (bottom) and second layer (top). (c) Out-of-plane motion of N atoms around the central Ni metal sites.

Figure 2.7c depicts the out-of-plane motions of the N atoms around the central Ni centers, indicative of a stepped geometry for the layers as four amino groups are divided to two groups of two in the z direction. Moreover, the never-ending motion of the N atoms during the AIMD simulation as shown in Figure 2.7c reflects a continuous SBU deformation at 293 K, resulting in a constant change of coordination environment from square planer to nearly tetrahedral. The deformations along each $Ni_3(HITP)_2$ layer was also confirmed by taking the average over the collective atomic motions during the duration of the performed dynamics. Figure 2.8a shows the profiles obtained from taking the average over the out-of-plane displacements. We can see how the difference between the average over one organic linker unit (i.e., 18 C atoms) and one layer (i.e., 36 C atoms) increases during the simulation. This difference by contrast is quite small if we compare one SBU to four SBUs (i.e., the total number of SBUs in one layer). In Figure 2.8b we project these averages on the molecular plane in xy direction. Again, one can see larger deformations for

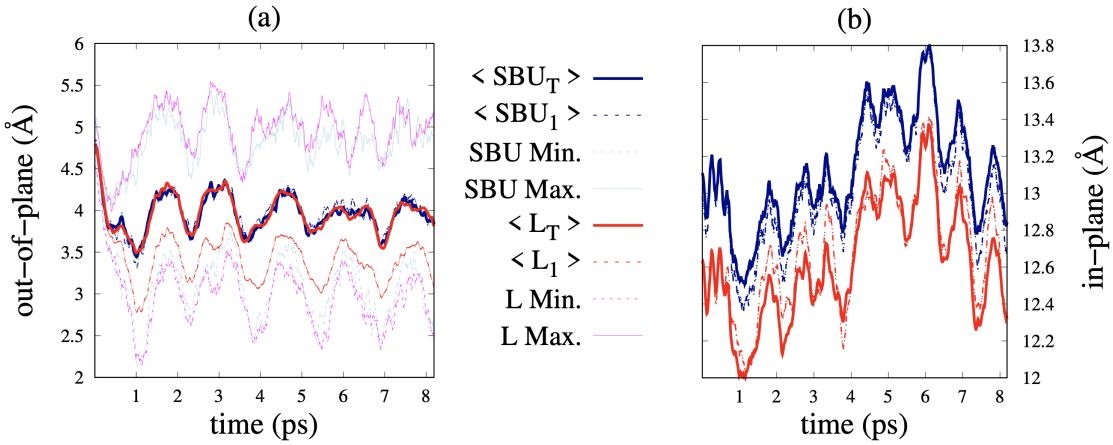

Figure 2.8 Inorganic (SBU) vs. organic (L) collective (a) out-of-plane and (b) in-plane motion in $Ni_3(HITP)_2$ during an entire 8 ps AIMD simulation. Subscript T and l stand for average over one layer and one unit, respectively. Profiles of the maximum and minimum values are also shown for the out-of-plane displacements.

the organic units compared to SBUs. Figure 2.9(a) illustrates the band structure and pDOS diagrams of the AIMD equilibrated $Ni_3(HITP)_2$ at 293 K. There are striking differences between these diagrams and those of the optimized structure at 0 K. We see intra-layer band openings in the range of 50 to 200 meV, which are close to the range of 20-120 meV previously reported [37] for $Ni_3(HITP)_2$ through introducing various forms of defects into the structure. Similar to intrinsic (undoped) semiconductors, the Fermi level lies at the center of the band gap and a narrow band gap is observed due to redox matching between the SBU and the organic linkers. Interestingly, the out-of-plane band structure along the $\Gamma-A$ region also shows the signature of semiconducting behaviour in the $\pi - \pi$ stacking direction compared to the metallic behaviour in the 0 K structure. We emphasize that this semiconducting nature is the result of different classes of thermally-induced interconnected dynamical motions which change the nature of frontier orbitals. Comparison between pDOSs with z component between 293 K and 0 K, i.e., far right panels of Figures 2.9 and 2.6, respectively, shows that the metal d_{xz} and d_{yz} orbitals are pushed toward lower energies. On the other hand, the stepped geometry in the 293 K structure allows for

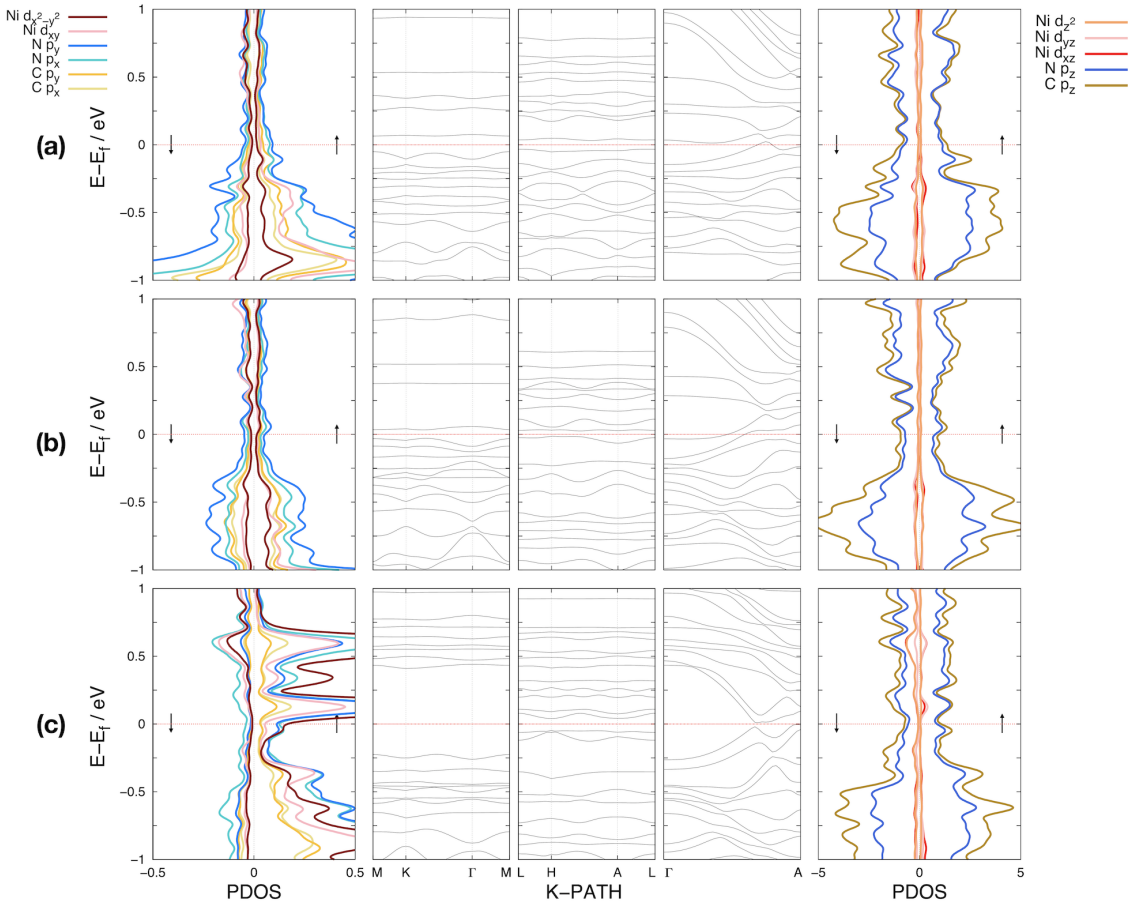

Figure 2.9 Band structure diagrams and pDOS profiles of the $Ni_3(HITP)_2$ 2D MOF at 293 K and 1 atm. Top: at the 8ps during equilibrium; middle: at the 1ps during equilibrium; bottom: at the 3ps during equilibrium.

contribution of the metal d_{xy} and $d_{x^2-y^2}$ orbitals to the Fermi level albeit to a smaller degree, compare far left panels of Figures 2.9 and 2.6. Clearly smaller contribution of the Ni d orbitals to the Fermi level disrupts the $\pi - d$ extended-conjugation pathway in the stepped geometry compared to the close to perfect square planar geometry in the 0 K system. This leads to intra-layer band-gap openings and flat bands with high effective band masses. The same effect can be seen in the LHA region with a relatively smaller band gap opening and a competition between direct and indirect band gaps. The flat bands could potentially result in low charge mobility (μ) as μ is inversely proportional to the effective mass (m^*) which in turn is related to the curvature of the electronic band structure in the reciprocal space, ($\mu = \tau e/m^*$) with e and τ being the elementary charge and the scattering time, respectively. However, as in the case of any other intrinsic semiconductor, thermally or optically excited electrons can contribute to the electrical conductivity. In the absence of a photonic excitation, intrinsic semiconduction takes place at temperatures above 0 K as sufficient thermal energy is required to transfer electrons from the valence band to the conduction band. With small band gaps, such as in the case of the $Ni_3(HITP)_2$ 2D MOF, thermal energy is sufficient to create significant concentrations of electron and hole carriers. In case of a large band-gap though, defect and/or doping strategies will be required in order to increase the carrier concentrations and hence increase the conductance. It is worthy of mentioning that although flat bands of equilibrated $Ni_3(HITP)_2$ 2D MOF signals high effective mass and lower delocalization, one may argue that it can decrease the recombination rate of electrons and holes and lead to an unexpected enhanced charge mobility. We note that the out-of-plane bands are still relatively dispersed through Fermi level which again demonstrate the importance of the through-space CT pathway in 2D MOFs.

Finally, we draw attention to the fact that the flexible nature of 2D MOFs is a dynamic feature. All three groups of dynamical motions characterized in this

section are constantly changing the structural features and electronic structure of the material. Any different combination of these motions can lead to a varied band structure diagram. To demonstrate this, we extracted snapshots of the $Ni_3(HITP)_2$ from the AIMD trajectory at 1 ps and 3 ps and calculated their electronic band structures, see Figure 2.9(b) and (c). Although both of these snapshots show an intrinsic semiconductor but the intra-layer band gaps can decrease lower than 10 meV or increase by more than 700 meV. In the case of the 3 ps snapshot, even the inter-layer band structure shows signs of a very narrow band-gap opening. Therefore, due to the interconnection of several structural motions in inducing band-gap openings we find it more instructive to provide a qualitative description of the correlation between flexibility and conductivity instead of a quantitative one. Interested reader can apply the the results of the previous section to decipher the band diagram features, for example band-gap openings along the out-of-plane direction can be a combination of increasing inter-layer distance and slipping of layers at the same time which reduces the perfect through-space $\pi-\pi$ interaction between layers. One should also remember that the relevance of computed band structures here, or based on any other similar dynamics simulations in future, to the real transport properties of a 2D MOF should be considered in the light of the presence of a high density of charge-scattering sites, specifically grain boundaries, in the synthesized polycrystalline pellets.

2.4 Outlook

It is well-known that electrical conductivity of 2D MOFs is dictated by the density and mobility of charge carriers, i.e., electrons and holes. Here, we established that intrinsic flexibility of the $\pi-$stacked layered frameworks should be considered as another important factor dictating the electrical conductivity. Our analysis clearly shows that transition from metallic to semiconducting occurs in the case of $Ni_3(HITP)_2$, as an archetypal 2D layered MOF, should the thermally-induced deformations are

allowed to happen. Furthermore, the effect of flexibility on the electrical conductivity is an unceasing phenomenon. Overall, our analysis reveals that the structural deformations and flexibility of the framework dictate the in-plane vs. out-of-plane orbital contributions at the Fermi level. The extent of these structural deformations in turn is sensitive to the nature of the building units (i.e., linker and SBU). For example, we have shown elsewhere that the high affinity of Co^{2+} centers to oxo-substituted triphenylene linkers create inter-layer connections to the neighboring layers within the 2D framework [133]. As long as one considers the matching between the SBUs and linkers (tessellation), the nature of orbital contributions at the Fermi level, and the structural deformations or flexibility of the framework, they can qualitatively predict the intraband mobility in a specific region of the band spectrum. Eventually, the conductive properties of the material can be tuned by balancing the intra-/inter-layer carrier transport anisotropy at a given temperature and/or pressure. However, a more accurate fine tuning of electrical properties requires a quantitative analysis of the intraband transfer integrals vs. the reorganization energy involved in the charge transport. Carriers move either through delocalized bands (through band mechanism) or through trapped polaronic states (hopping mechanism), depending on the strength of the electron-phonon coupling. Let us assume for instance that we wish to design a 2D MOF with semiconductive absorption properties. Figure 2.10 shows a schematic representation of the $E - k$ diagram of indirect and direct band gaps. Provided a specific tessellation of the 2D MOF topology according to the node and linker design principle [60] and based on the expression $(SBU)^m (L)^n$ [19] we can localize charge states on the lattice through modifying intra-/inter-layer CT pathways via slight framework deformations. One can realize this objective by varying external temperature, pressure and electric/magnetic fields. The induced polaronic states would flatten the curvature of the electronic band structure, allowing the emergence of a narrow band gap. In the resulting indirect band-gap, electron-hole recombinations

based on the radiative event is unlikely as it requires the photon to obey total momentum conservation according to In order to supply a set of design principles, our

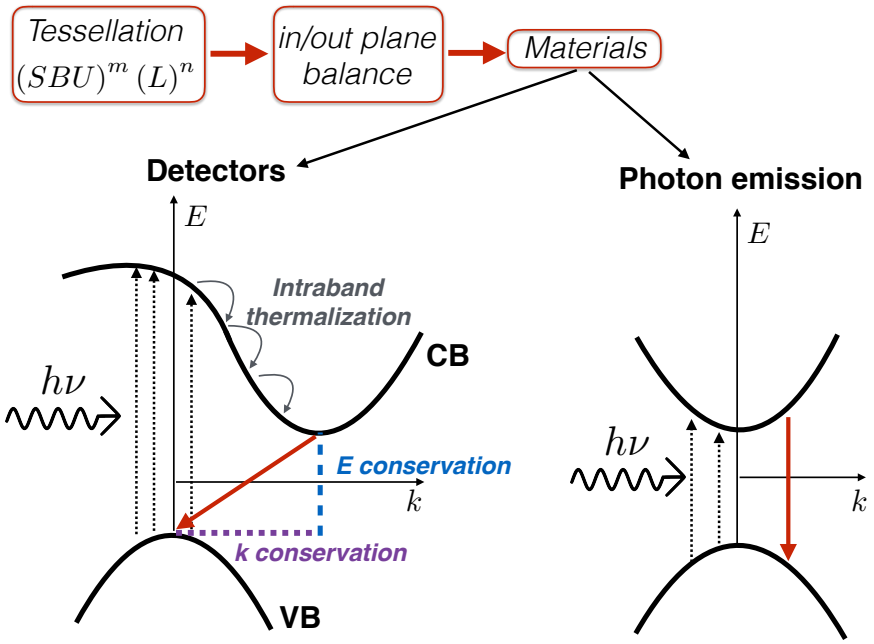

Figure 2.10 $E - k$ diagram representations to display the structure-properties relationship in 2D MOFs. Given a $(SBU)^m (L)^n$ composition to construct the MOF tessellation [19], the semiconductive properties of the material can be fine-tuned by adjusting the in-plane vs. out-of-plane orbital contributions.

current research is focusing on quantitative control over the reorganization energy, which might represent the ultimate road map to predict semiconductive MOFs with tailored structures and transport properties.

2.5 Concluding Remarks

2D layered MOFs, comprised of paired transition metals with hexa-substituted triphenylene or benzene linkers, are a new class of materials that offer breakthrough applications due to electrical conductivity in addition to exceptional high surface area and permanent porosity of conventional MOFs. The progress in their implementation in devices is hampered by the lack of understanding of the intrinsic structure and

dynamics of the layered architecture. Here, we show for the fist time that *static* electronic structure calculations result in a close to a perfect crystal structure and square planar coordination environments fail to reflect the realistic experimental conditions. Through comparing electronic band structures calculated at 0 K vs. the MD equilibrated systems at 293 K we show for the first time that incorporation of intrinsic thermally-induced structural deformations in 2D MOFs is crucial in realization of a qualitatively correct picture of their electrical conductivity. Particularly, a semiconducting nature was found for Ni_3HITP_2 as a result of the incorporation of different dynamical motions such as deformation of SBUs from square planar geometry, change of inter-layer distances and slipping of the layers compared to each other. We provided insights into the electrical conductivity via disentangling extended $\pi - d$ conjugation from $\pi - \pi$ through-space CT pathways where it was shown for the first time the $\pi - \pi$ interactions between layers to be if not more crucial but more important as extended-conjugation in dictating the electrical conductivity of the bulk material. Overall, our work highlights the importance of including temperature-induced lattice vibrations in simulation of flexible 2D MOFs in order to obtain a more realistic picture of their electrical properties.

CHAPTER 3

IN-SILICO HIGH-THROUGHPUT DESIGN AND PREDICTION OF STRUCT-URAL AND ELECTRONIC PROPERTIES OF LOW-DIMENSIONAL METAL-ORGANIC FRAMEWORKS

3.1 Background and Motivations

Various building blocks for EC-MOFs are reported in the literature and, as mentioned in the previous chapters, MOFs in general can be rationally designed by choosing different combinations of organic linkers and metal node building blocks. However, considering the vast and virtually infinite chemical space of MOFs, it is extremely labor intensive and time consuming to synthesize all different combinations of building blocks to find the best materials for any desired application. A more efficient and systematic way is to create a comprehensive database of different classes of MOFs and then screen them for desired applications using accelerated high-throughput screening (HTS) techniques [123]. Chung et al. [25] created a computation-ready, experimental (CoRE) MOF database with over 5,000 MOFs based on CSD in 2014. Various other datasets have evolved from the CoRE MOF database, such as CoRE MOF 2014+DDEC [108] where partial atomic charges were determined for 50% of the reported MOFs using density functional theory (DFT) calculations as well as CoRE MOF 2014-DFT-optimized [107] where DFT geometric relaxation was performed for 879 structures. CoRE MOF database itself was updated in 2019 with the total number of MOFs being increased to 14,000 [26]. On the other hand, the Cambridge Crystallographic Data Centre (CCDC) created a MOF subset based on existing CSD which contains the largest number of experimentally synthesized MOFs to date [101]. A total number of 69,666 MOFs were gathered in this subset after screening the original CSD based on seven different criteria and removing solvent molecules from the MOF pores. A more recent work by Rosen et al. [122] in 2021 introduced a new database called Quantum MOF (QMOF) database containing 15,713 MOFs

that were successfully optimized and analyzed by HTS periodic DFT workflows. Not only experimental MOFs are collected in different databases, a hypothetical MOF database is also recently built by Wilmer et al. [153] where 137,953 new MOFs were created from a combination of 102 different building blocks. More than 300 candidates were selected with excellent methane storage capacity. More recently, Gharagheizi et al. [40] created a collection of ∼20000 1D MOFs curated from CSD nondisordered MOF subset which include ∼2000 electrically conductive structures. All these carefully curated databases allow selection of materials with desired properties and performance for specific applications by fast screening of hundreds and/or thousands of structures. However, they all exclude or at best partially include π-stacked layered EC-MOFs since they are a very new class of materials. Considering the wide potentials of EC-MOFs, it is crucial to first build a comprehensive database for them which will then allow various HTS techniques to be routinely applied in order to accelerate materials design and discovery. Here, we report the first installation of our experimentally-guided, computationally-ready database of EC-MOFs, coined EC-MOF/Phase-I, containing 1,057 bulk and mono-layer structures. This database is available to the community use via the developed online platform during the course of this study at https://ec-mof.njit.edu. All the structures in this database follow a comprehensive naming rule as shown generally and with an example in Figure 3.1. This naming rule, which will be explained in the next section, gives enough information to the user about the nature of metal nodes, functional groups, organic linkers, connectivity between building blocks and type of unit cells. In this way, the users can easily have access to the desired structure in the database through the choice of each of these components where they can build and download their structure. Furthermore, the users not only have access to the crystal structures but also geometric data and electronic properties obtained using our in-house HTS workflow. We will provide the details of the applied procedures and developed

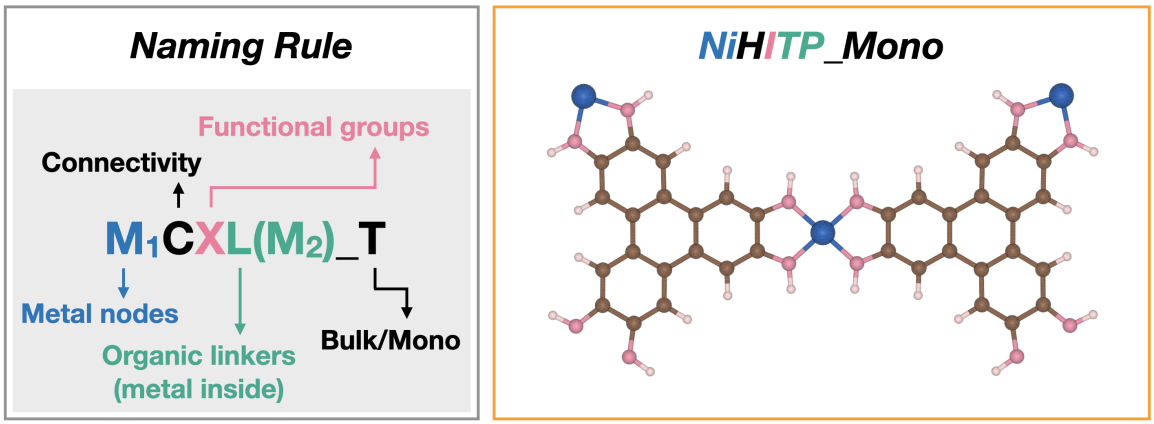

Figure 3.1 Naming rules used for the structures included in EC-MOF/Phase-I database (left) and an example of a built NiHITP crystal structure (right).

software for building the database in the next section. In section 3.3, we will present the results of our HTS investigation of the carefully curated 1,057 bulk and mono-layer systems. Section 3.4 outlines future directions with concluding remarks given in section 3.5.

3.2 Computational Developments

3.2.1 EC-MOFs from literature

In this work, we first performed a thorough literature survey and summarized all reports on EC-MOFs that have been either synthesized and/or theoretically investigated. With the intention of developing a structure creation tool for automatic generation of initial crystal structures of EC-MOFs for our database, we initially focused on identifying structural features that induce the highest electrical conductivity. Accordingly, we restricted the first version of our database to π-stacked EC-MOFs with planar layers and extended π-conjugation through organic linkers and metal nodes with 2+ oxidation state allowing for effective $d-\pi$ conjugation. Notably, Hofmann-type MOFs are excluded from the first version of our database because depending on the nature of the coordinative bonds they can form 3D structures where π-stacking can occur along the in-plane rather than out-of-plane direction.

Moreover, presence of Pt^{2+} nodes can disturb in-plane electrical conductivity in some of these MOFs [128, 127]. Other classes of MOFs were excluded on the basis of the presence of twists in the layers [93] which can disrupt extended in-plane conjugation, presence of bi- and tri-nuclear metal nodes [91] which can disrupt effective in-plane $d-\pi$ conjugation, and more than 4 Å interlayer distances [38] that reduce out-of-plane $\pi-\pi$ interactions drastically. The subset of EC-MOFs gathered in our database follow the color-coded naming rule depicted in Figure 3.1 based on three structural building components including organic linkers, metal nodes, and functional groups. Figure 3.2 demonstrates all the structural components used in building EC-MOF/Phase-I database. The eight metal nodes including Mn, Fe, Co, Ni, Cu, Zn,

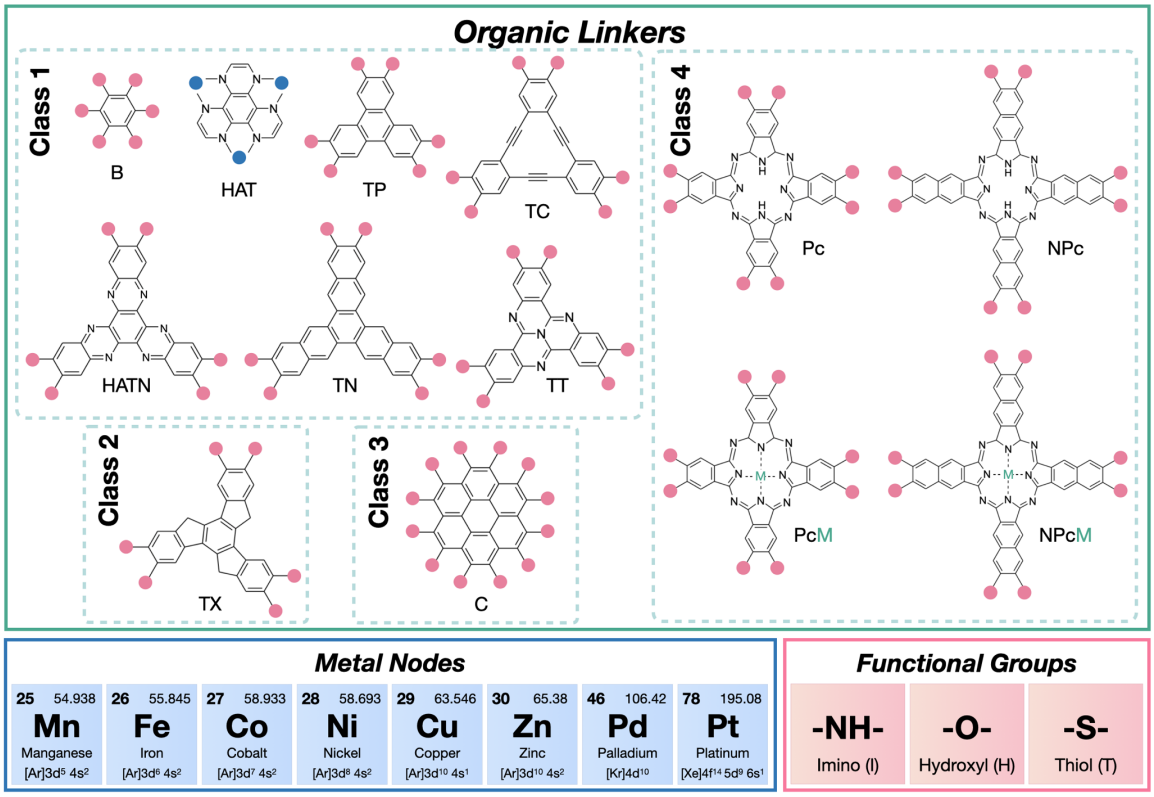

Figure 3.2 Three subsets of structural building components including organic linkers (classes 1-4), metal nodes, and functional groups (green, blue and pink boxes, respectively). Connecting sites of organic linkers to the metal and functional groups are highlighted by blue and pink dots, respectively.

Pd and Pt, are identified by M_1 in our naming rule while X denotes three functional

groups as H for hydroxyl, I for imino, and T for thiol that connect to the metal nodes. The linker subset of structural building components, denoted as L in our naming rule, consists of 11 organic linkers including benzene (B), triphenylene (TP), trinaphthylene (TN), truxene (TX), coronene (C), tribenzocyclyne (TC), tetraazanaphthotetraphene (TT), hexaazatriphenylene (HAT), hexaazatrinaphthalene (HATN), phthalocyanine (Pc) and naphthalocyanine (NPc). Pc and NPc linkers can accommodate one more transition metal atom inside. Transition metals including Fe, Co, Ni, Cu, Zn and Pd are observed in the synthesized Pc and NPc-based MOFs [165, 109, 100, 145, 53, 84], therefore, PcM and NPcM with 6 different interior metal centers are considered. The connections between metal nodes and organic linkers happen through 3, 4 or 6 bidentate sites. Accordingly, 6, 8 or 12 functional groups are placed around one linker. This is identified with C (as in Connectivity) in our naming rule which can be replaced with H for hexa, O for octa and P for per. Furthermore, M_2 is the metal atom inside the organic linkers, if existed. At last, T indicates the type of unit cell which has two options, bulk or mono-layer. Naturally, some of the resulted structures are originated from reported works but most are hypothetical. Hence, our building strategy screens all combinations of the building blocks but keeps a tight connection between experimental and hypothetical structures.

3.2.2 Crystal structure producer

Based on the criteria established in the previous section, we identified four different classes of EC-MOFs according to the shape and connectivity of the organic linkers as demonstrated in Figure 3.3. An in-house structure building tool, coined Crystal Structure Producer (CrySP), is developed to create periodic structures of EC-MOFs that fit into these four geometric classes without any minimization using generic force fields. The optimization of all structures was carried out at the DFT level which will be explained in the next section. The CrySP algorithm starts by rotating

the organic linker to the desired position, then functional groups are placed around the organic linker according to the connectivity of the linker. Next, metal nodes are added according to the bond length between metals and functional groups which is specified in advance. After necessary transformations, the structure is rotated to fit into the specified unit cell. Structures in our Database, coined EC-MOF/Phase-I, can be classified into 3 different lattices, including honeycomb (hcb), hexagonal (hxl) and square (sql) lattices.[92] We further divide hxl structures into 2 classes due to the

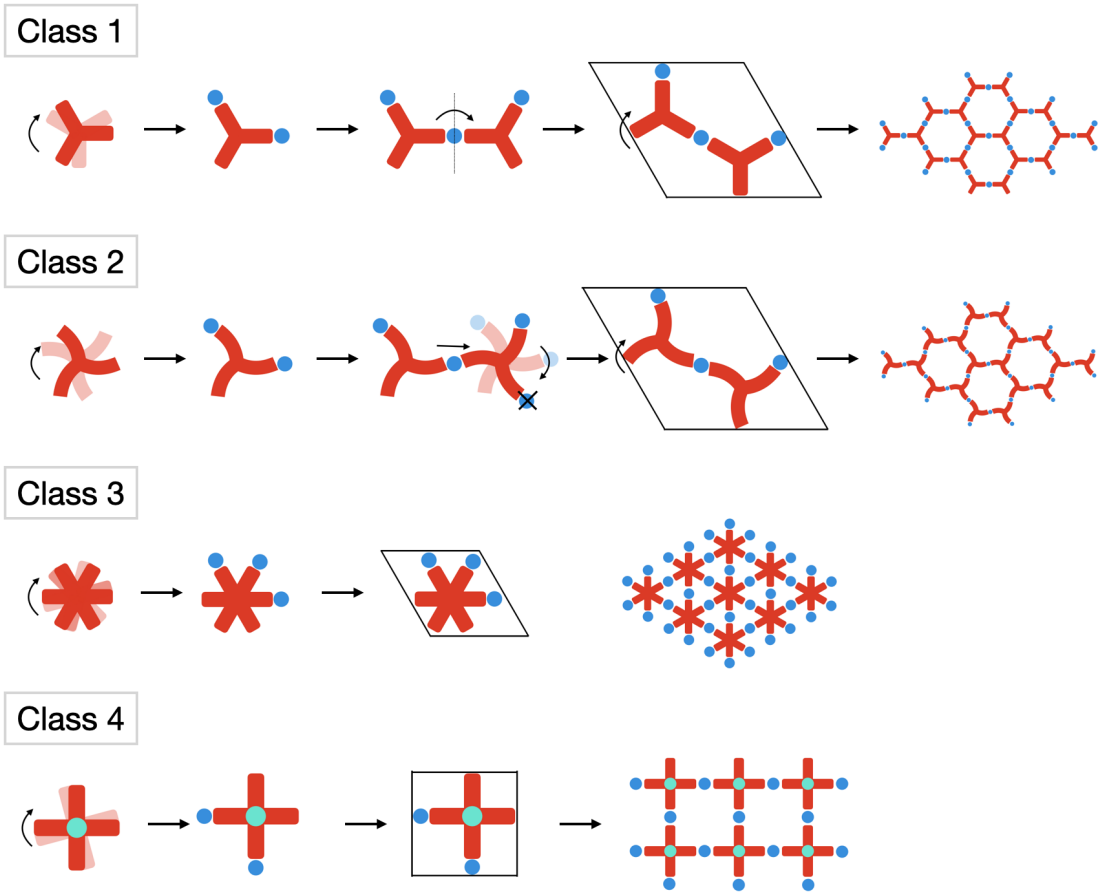

Figure 3.3 Building logic of CrySP for four different classes of MOFs created in this work. Classification of organic linkers is given in Figure 3.2. In this figure blue represents metal nodes, red organic linkers and green metal atoms located inside the linkers.

unique building procedure needed for TX-based MOFs. The details of the building procedures of the four classes may vary as schematically shown in Figure 3.3. In Class

1, CrySP calls up the XYZ coordinates of the desired organic linker and places metal nodes around it in appropriate positions and distances. Then, the entire structure is reflected along the y axis. The last steps are moving and rotating the structure to fit into the specified unit cell. In Class 2, the reflection step is replaced by moving and rotating operations due to the special symmetry of the linker. In Class 3 and 4, once the organic linker is placed at the center of the unit cell, functional groups and metals are added around the organic linker without any transformational operations. At the same time, CrySP calculates cell parameters for each class of materials by taking the positions of metal nodes as reference points. Finally, CrySP creates the structures in the desired formats including XYZ, Crystallographic Information file (CIF) or POSCAR. The resulted structures at this stage are all mono-layers. To optimize the mono-layer structures, a vacuum space of 20 Å is added to the c direction of the unit cells. Bulk structures, containing two layers in the unit cell, are created with a fixed inter-layer distance of 3.25 Å along the out-of-plane direction while considering different crystal packings known as AA or AB stackings [75]. Accordingly, a total of 1,072 structures are created by CrySP and gathered in the first version of our EC-MOF Phase-I database.

3.2.3 Details of high-throughput screening workflows and periodic electronic structure calculations

To maximize the advantages of EC-MOF/Phase-I, we apply high-throughput screening (HTS) techniques to explore different properties of these materials (see Figure 3.4). Multiple steps of periodic density functional theory (DFT) calculations are performed in our HTS workflows, which are all carried out using the Vienna *ab* initio simulation package (VASP) version 5.4.4 [80, 81, 79, 78]. Interactions between electrons and ions are described by projector-augmented wave (PAW) potentials [13, 83] with a cut-off energy of 500 eV. Spin-polarized calculations are performed for all systems resulting in high-spin states being the most common for all structures rather than low-spin states.

Perdew-Burke-Ernzenhof (PBE) functional with Grimme's damped D3 dispersion correction within the generalized gradient approximation (GGA) formalism [115, 45] was employed in this work which provides accurate results for our EC-MOF systems, as discussed in greater details in our recent work [169]. First stage of our HTS workflow is to geometrically relax all structures obtained from CrySP to find the ground state energy configurations for both the ions and electrons. The criteria for optimization convergence is set to less than 10^{-4} eV for electronic energy and the magnitude of the largest force acting on the atoms is set to less than 0.02 eV/Å. Three degrees of freedom are allowed to change, atomic positions, cell volume and cell shape for all bulk systems. Cell volume is fixed during optimization of the

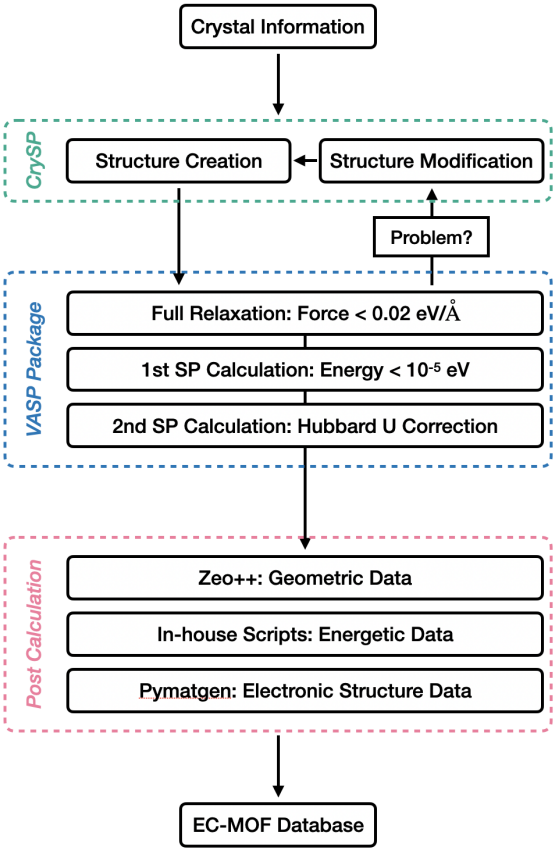

Figure 3.4 High-throughput screening workflow employed in this work including different steps of geometric relaxations and single-point energy calculations for all structures included in the EC-MOF Phase-I database.

mono-layers because we use a slab model with c = 20 Å to prevent interaction

between layers from different periodic cells. If the optimization encounters structure and setting-related issues or has any difficulty to reach convergence, an automatic debugging process is executed to collect error messages which automatically updates the input files for re-optimizations. Once the optimization is completed, the 1^{st} round of single-point (SP) calculations is carried out to check if there is any structural or setting-related issues and to provide a good initial electron density for the following calculations. The convergence criteria for electronic energy is set to 1×10^{-5} eV. The Brillouin zones are sampled using $2 \times 2 \times 6$ or $2 \times 2 \times 4$ k-point mesh for bulk systems depending on different ratios of cell parameters and $3 \times 3 \times 1$ k-point mesh for all mono-layer systems. Gaussian smearing method, with a smearing width of 0.05 eV is adopted in this step to provide accurate electronic energy and density information. The 2^{nd} round of SP calculations reads the electron density calculated in the 1^{st} round, which increases efficiency. The smearing method is changed to the tetrahedron method with Blöchl corrections which provides more accurate results for the calculated band gaps and density of states (DOS) [134]. DFT in its GGA formalism is known for underestimating band gaps. Hence, Hubbard U approach (DFT+U), which semi-empirically optimizes the Coulomb interaction potential (U), is adopted in this step to give a better description of electronic structures. Only d and f electrons were treated by this approach with the employed semi-empirical U parameters for each metal reported in the Table 3.1.

Table 3.1 Hubbard U Parameters Adopted in This Work

Element	Mn	Fe	Co	Ni	Cu	Zn	Pd	Pt
U (eV)	5.5	6.5	5.3	6.7	10.4	10	4	2.9

To validate the accuracy of our calculated band gaps, the computed values obtained at the PBE-D3 level with the U correction were benchmarked against the HLE17 meta-GGA functional as implemented in the Minnesota VASP Functional

Module codes and is already shown to provide band gap values for different materials which are in good agreement with experimental results [142, 23]. For this benchmark, we have chosen semiconductors with the highest band gap values among both bulk and mono-layer systems. The results in Figure 3.5 show that DFT+U qualitatively reproduces the same trend as HLE17 data. Accordingly, successful completion of both SP steps provided necessary information for us to extract the final property data from our EC-MOF/Phase-I database.

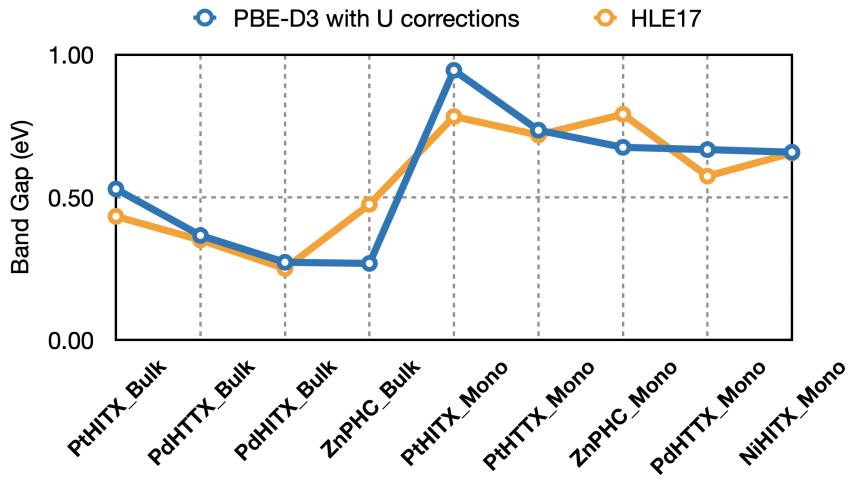

Figure 3.5 Calculated band gap values for selected semiconductors from EC-MOF/Phase-I database. The results from DFT+U and HLE17 functionals are shown in blue and orange, respectively.

3.2.4 Post-calculation analysis

Our HTS calculations provide us with different properties including geometric, energetic and electronic data for all structures generated and contained in our EC-MOF/Phase-I database. Periodic crystal structures are extracted as CIF files after the geometry optimization stage. Analysis of the Voronoi network, as implemented in Zeo++, [140] is used to determine different geometric data including largest cavity diameter (LCD), accessible void-volume fraction and accessible surface area, using a Helium probe with the radius of 1.4 Å, for all bulk EC-MOFs. From SP calculations, the total energy and electronic structures of the systems are obtained.

Absolute formation energy which is an important indicator of the stability of different structures, is calculated as [71]:

$$E_f = E_{tot} - \frac{1}{N}\sum_{i=1}^{N} x_i \mu_i \tag{3.1}$$

where E_{tot} is the calculated energy for the entire bulk material, N is the total number of atoms with x_i and μ_i being the number and chemical potential of element i in the structure. Calculations of each element in their most stable configuration are implemented to find the energy values per atom which is needed in the equation above. Inter-layer binding energies were calculated as following [61]:

$$E_b = (E_M - \frac{1}{n}E_B)/A \tag{3.2}$$

where E_M is the calculated energy of the mono-layer, n is the number of layers in the bulk structure, E_B is the calculated energy of the bulk structure and A is the surface area of the mono-layer. All the needed energetic data are extracted and calculated by our in-house codes. Electronic band gaps are obtained from VASP using Python Materials Genomics (Pymatgen).[66] Band center of the d orbitals are extracted using Pymatgen and the following equation [50]:

$$C_d = \frac{\int_{-\infty}^{+\infty} \rho E dE}{\int_{-\infty}^{+\infty} \rho dE} \tag{3.3}$$

where C_d is the position of the d-band center, ρ is the density of the d orbitals, E is the energy of the d orbitals relative to the Fermi level and ρdE is the number of states contributed by d orbitals.

3.3 Results and Discussion

Our database contains 1,063 EC-MOF crystal structures which are fully relaxed using the periodic boundary conditions at the DFT level. A wide range of data including

structural information, formation energies and electronic properties are extracted and included in our EC-MOF/Phase-I database that can be accessed via https://ec-mof.njit.edu. Such data are useful for other researchers who need to screen and target attractive properties without additional efforts. Here, we discuss these calculated properties in more details.

3.3.1 Structural data analysis

After full geometric relaxation, a total number of 535 bulk EC-MOFs are successfully converged to the force criteria outlined above except the ZnHTTX_Bulk system. In the case of the mono-layer structures, eight MPIC systems failed to maintain their connectivity and topology due to strong steric hindrances. Hence, they were not included neither in the online EC-MOF/Phase-I database nor here. This results in 1,063 structures in EC-MOF/Phase-I database. Another notable point about mono-layer systems is that the optimized structures of HAT-based MOFs are slightly distorted due to the steric hindrance between two neighboring organic linkers. Similar phenomena were observed by Lyu et al. [96] where mono-layer HAT-based MOFs could only be synthesized by an on-surface reaction approach. Here, we first present structural data extracted from 535 optimized bulk systems. Figure 3.6a shows frequency of structures as a function of largest cavity diameter (LCD) which is defined as the diameter of a sphere that can fit into the largest pore of the materials [140]. LCD of 2D MOFs in our EC-MOF database ranges from 0 to 25.2 Å depending on the size of the organic linkers. Structures with the highest LCD values consist of TC, HATN and TN-based MOFs with the largest linker sizes of all. Structures with small LCD values are mostly C-based MOFs. The high number of connectivities of one C linker to the metal nodes limits the spatial space leading to an LCD value of 0 Å. It should be noted that the computed 0 value for LCD does not mean that the system is not porous but the diameter of the Helium probe is larger than the diameter

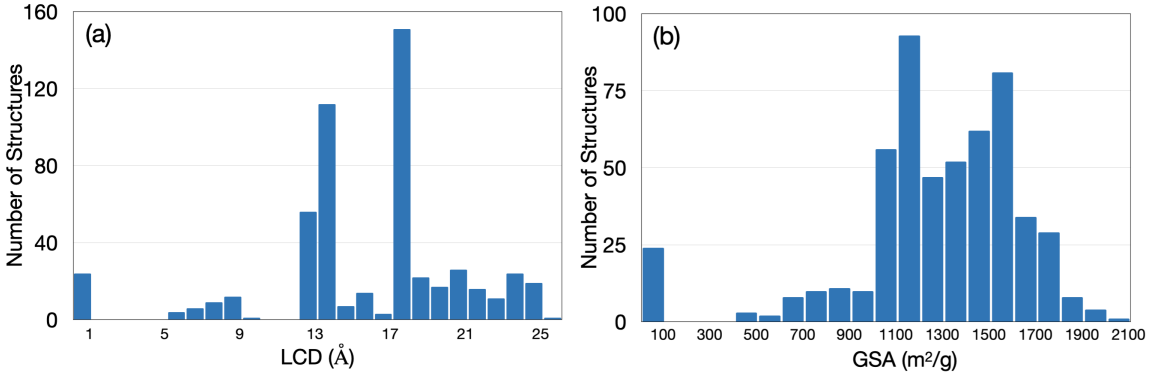

Figure 3.6 Number of structures from our EC-MOF database as a function of the calculated (a) largest cavity diameters (LCD) and (b) gravimetric surface areas (GSA).

of the pore in these structures. The calculated LCD of the rest of the structures is qualitatively ranked according to the size of their organic linkers. Pc and PcM-based MOFs, where linkers have similar sizes, constitute a large body of MOFs with LCD of ∼13 Å, Figure 3.6a. Similarly, NPc and NPcM-based MOFs constitute the peak at ∼17 Å in Figure 3.6a. Figure 3.6b shows frequency of structures as a function of gravimetric surface area (GSA). In the range above 1700 m^2/g, TC, HATN, TN and TX linkers along with several tetragonal linkers take the lead as the MOFs with the largest GSA. Due to the unique structures of the TC linkers, TC-based MOFs not only have the conventional 1D channels but also another 1D channel with a smaller radius inside the linkers themselves. Therefore, TC-based MOFs possess the highest GSA with an average value of 1711 m^2/g over all combinations of metals and functional groups, as reported in Table 3.2.

Similar results were reported in the work by Park et al. [117], where CuHHTC was synthesized and featured as a special 2D MOF with an enhanced surface area. In their experimental work, an unprecedentedly high GSA of up to 1196 m^2/g was measured that can be compared to CuHHTC in our EC-MOF database with a GSA of 1737 m^2/g. GSA of TT-based 2D MOFs is also reported by Dincă et al. [34] where the GSA of the synthesized CuHHTT is around 1360±20 m^2/g compared to a

Table 3.2 Average Largest Cavity Diameter and Gravimetric Surface Area Values with Respect to Each Organic Linker

Linker	LCD(Å)	GSA(m^2/g)
TC	21.359	1711
HATN	23.800	1639
TN	23.908	1589
TX	20.444	1547
TT	19.737	1502
NPc	17.637	1438
NPcM	17.659	1434
TP	15.381	1201
Pc	12.962	1120
PcM	12.990	1105
B	7.866	663
HAT	5.911	313
C	0	0

value of 1549 m^2/g calculated in this work. According to our average calculated GSA data, TT-based 2D MOFs rank in the 5^{th} place as shown in Table 3.2. Structures found in the lowest range of GSA, i.e., below 500 m^2/g, are mainly C and HAT-based MOFs due to the specific connection of the organic linkers and metal nodes in these MOFs and the size of probe used in Zeo++. The practical implication of these small cavities will be adsorptive separation of smaller sized guest molecules for applications as molecular sieves. When comparing the GSA data in our EC-MOF database with those from experimentally measured surface areas, our calculated GSA values are always higher than what is measured experimentally by Brunauer–Emmett–Teller (BET) equations. The discrepancy can be mainly related to (1) the size of crystal particles compared to prefect bulk crystals in our EC-MOF database, (2) presence of defects in synthesized materials, and (3) pores that are not completely evacuated

from solvent molecules. Our calculated GSA data therefore provide the maximum surface area that each EC-MOF can possibly reach regardless of the synthesis and activation procedure. Notably, the GSA data in our EC-MOF database qualitatively show similar trends with available experimental results. To reveal the relation between spatial space and accessible surface area, we converted GSA to volumetric surface area (VSA) and plotted them against the void fraction of 13 different organic linkers as shown in Figure 3.7. Data points for each linker are highlighted by distinct colors and

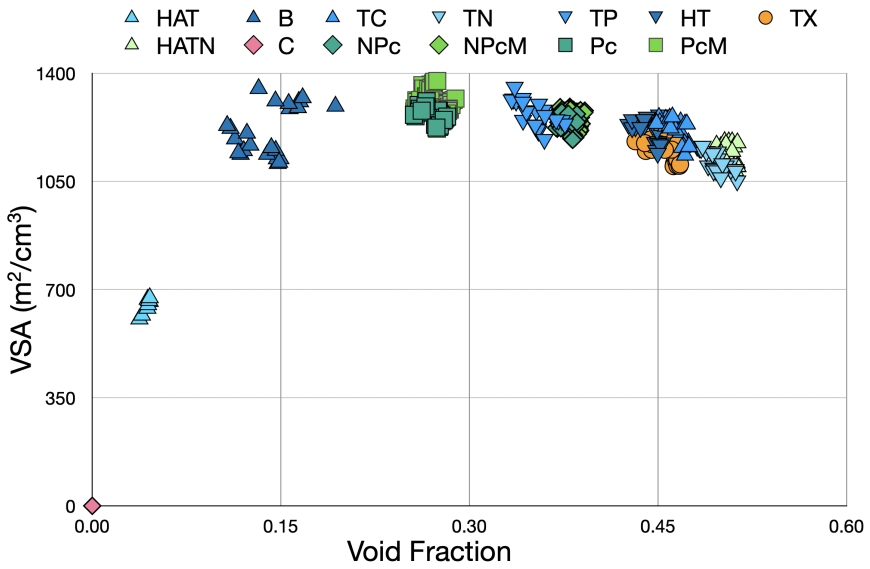

Figure 3.7 Calculated volumetric surface areas (VSA) as a function of void fraction according to different organic linkers.

shapes. In agreement with our GSA data, MOFs containing TN, HATN, TC, TX and TT linkers possess high void fractions ranging from 0.43 to 0.51 due to the size of the linkers. The data points located between 0.1 and 0.4 void fraction also have high VSA values. However, the difference between VSA values of MOFs with different linkers in this region is less noticeable compared to the difference in their GSA values, as shown in Table 3.2, due to the inclusion of the density of MOFs in converting GSA to VSA. The void fraction of MOFs with C and HAT linkers are as low as 0.07. Except these two classes, the VSA of the rest of structures are all higher than 1000 m^2/cm^3 while their void fraction ranges all the way from 0.1 to 0.5. Hence, regardless of accessible

surface areas, the appropriate MOFs for adsorption/separation applications should be chosen based on the void fraction as well. For energy/gas storage purposes, on the other hand, gravimetric/volumetric capacities of the active compounds are more relevant [42, 3]. However, applications of these MOFs in adsorption should be further investigated by analyzing the host–guest interactions. For instance, maximizing the performance of MOFs in hydrogen storage requires maximizing VSA while preserving a large void fraction [42]. On the other hand, number of adsorption sites should be the primary concern in the case of chemical adsorptions [87]. Here, as an effective descriptor for the host–guest interactions, the d-band center model is integrated into our EC-MOF database which is further discussed in the following section.

3.3.2 Thermodynamic data analysis

Formation energy (E_f) is one of the most principal parameters to consider when determining whether a hypothetical material will be synthesizable.[119, 99] Following our high-throughput workflow, calculation of E_f for 1,063 structures were implemented using Equation 3.1 with the results presented in Figure 3.8. Two bulk and four mono-layer structures failed to reach the energy convergence during this stage, which decreased the total number of structures in our EC-MOF/Phase-I database to 1,057. Both PcM and NPcM-based MOFs which share the same structural building blocks but with different metal centers inside the organic linkers are combined into one data point. This is because the different choices of metal centers were found to have a minor effect on the calculated E_fs, i.e., less than 0.09 eV/atom. Data points with a blue color in Figure 3.8 mean negative E_fs and show that the corresponding structures are more likely to be synthesizable. The red colored data points on the contrary correspond to positive E_fs and hence structures that are thermodynamically unstable. Accordingly, 96.06% of the bulk and 92.75% of the mono-layer structures have negative E_fs (see Figure 3.8). To find a trend in the

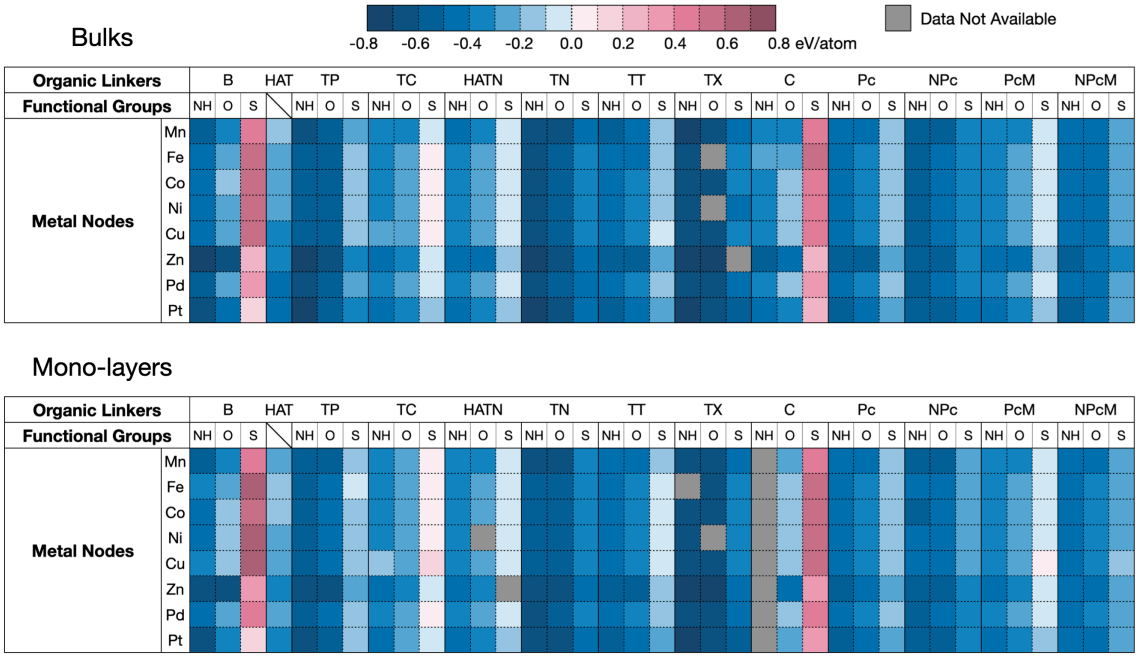

Figure 3.8 Calculated formation energies (E_f) of bulk (up) and mono-layer (down) structures contained in our EC-MOF/Phase-I database.

computed E_fs one should pay attention to the different bonding/interaction motifs that exist in the built MOFs. Overall, in all EC-MOFs there are three main types of bonds/interactions including covalent bonding within the organic motif, coordinative bonding between metals and functional groups of the linkers and comparatively weaker van der Waals interactions between layers. As an example, B, TP and TN linkers could be placed in an incremental sequence considering that each linker is comprised of the previous one plus three more benzene rings. Having more benzene rings in the structure increases the number of covalent bonds which is stronger than the other two interactions/bond types. Hence, the calculated E_fs of TN-based MOFs are more negative than their TP-based counterparts which in turn are more stable than the B-based MOFs, all due to the higher number of covalent bonds. A similar trend can also be observed in the cases of Pc vs. NPc or PcM vs. NPcM where the latter ones have four more benzene rings than the former ones. Figure 3.8 also illustrates the E_fs with respect to three functional groups within each linker

family. Its very clear that in each linker family, EC-MOFs with -NH functional groups always possess the most negative formation energies whereas the ones with -S groups possess the least negative or in some cases even positive values. The stability of the coordinative bonds between transition metals and different functional groups is determined by the compatibility between the two interacting units. Nitrogen atoms have lower electronegativity than oxygen leading to a stronger bonding to metals. Compared to sulfur atoms, sizes of nitrogen and oxygen atoms are more similar to carbon atoms, which induces a better overlap between atomic orbitals in the extended π-conjugated layers. Considering different transition metals, among eight employed metal nodes, Mn^{2+}, Zn^{2+} and Pt^{2+}-containing MOFs tend to have more negative E_fs regardless of organic linker and/or functional group types. Half-filled and fully-filled electronic configurations of Mn^{2+} and Zn^{2+}, respectively, are more stable in the same row of the periodic table when forming 2+ ions. Other than the standard way of creation of MOFs from different building blocks, physical and chemical properties of the EC-MOF based layered materials can be further modulated by tuning their inter-layer interactions by reducing the number of stacked layers all the way down to a mono-layer. Consequently, mono-layer materials exfoliated from their corresponding bulk systems may possess unconventional properties in the fields of gas adsorption [94], optics [174] and electrocatalysis [69]. EC-MOF mono-layers can be synthesized through the top-down or bottom-up approaches. Chemical exfoliation methods such as intercalation and electrochemical exfoliation are among the most common techniques in the top-down strategy [174]. On the contrary, the bottom-up approach is a more efficient and convenient approach to build mono-layer materials by choosing ideal precursors, modulators and surfactants [69]. To shed light on the possibility of reaching at a mono-layer EC-MOF, we have calculated the inter-layer binding energies (E_b) according to Equation 3.2 for all structures gathered in our EC-MOF/Phase-I database.

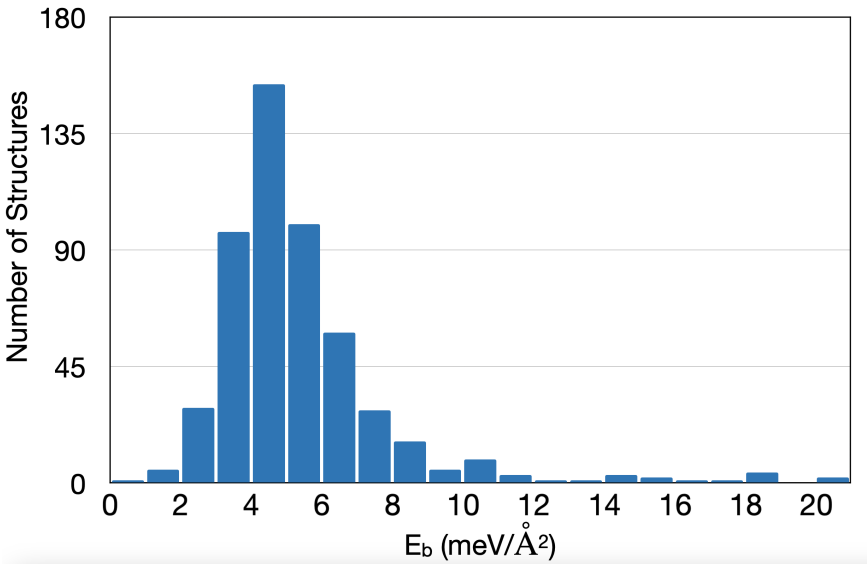

Figure 3.9 Distribution of structures according to their calculated inter-layer binding energies.

Figure 3.9 shows the histogram for the distribution of all calculated E_b values with a prominent range of 0 to 20 meV/Å^2 and a bin width of 1 meV/Å^2. The E_b of the majority of our EC-MOFs lie in the range of 2-9 meV/Å^2. This can be compared to the interlayer binding and exfoliation energies for a large number of layered compounds, including graphite and MoS_2, which are around 20 meV/Å^2 and are considered before for successful exfoliations [11]. It should be mentioned that the low E_b values of the EC-MOFs in comparison to other layered materials relates to the high porosity of our structures. This indicates the promising potential of most of these EC-MOFs for thin-film fabrication which is of utmost importance in compact device implementations. It is also worthwhile noting that, according to Equation (2), calculated E_b values are inversely proportional to the surface area of the mono-layer. Hence, within the same family of porous materials one will naturally obtain smaller E_b values for the systems with larger pore sizes, given similar inter-layer van der Waals interactions. For example, EC-MOFs with calculated E_b values larger than 13 meV/Å^2 are mainly HAT and C-based MOFs with possibly strong $\pi - \pi$ interactions

per unit area. To be noted, two structures, MnHAT and PtHHTP show E_b values higher than 20 meV/Å^2, 23.17 and 21.57 meV/Å^2, respectively.

3.3.3 Electronic property data analysis

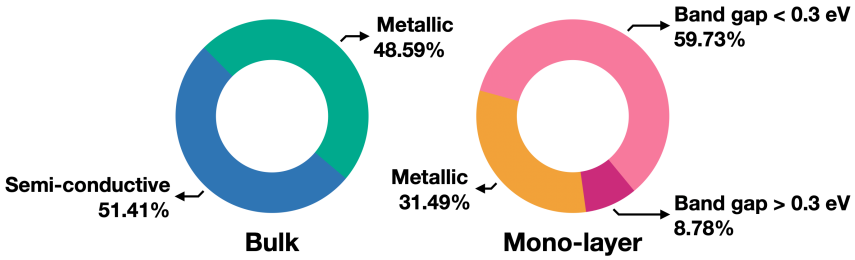

Figure 3.10 Distribution of the metallic and semiconductor systems in bulk (left) and mono-layer (right) materials. Mono-layer semiconductors are further divided into two regions according to their band gap values.

The electrically conductive behavior of π-stacked layered MOFs is in contrast to other conventional MOFs that are mostly classified as insulators with highly localized electrons [85]. To provide more insights on the electrically conductive behavior of the structures gathered in our EC-MOF database, we have calculated their fundamental band gaps as explained in section 3.2. As a result, 48.59% of the bulk EC-MOFs are calculated to be metallic compared to 31.49% for mono-layers, Figure 3.10. This can be explained by the absence of the inter-layer charge transport pathway along the stacking direction for the mono-layer structures. We also calculated the ratio of semiconductors within each class of EC-MOFs with different organic linkers and found that for most structures, the ratio of semiconductors in mono-layers are indeed higher than the ratios in the corresponding bulk systems. Figure 3.11 shows the nature of conductive behaviors, and band gap values if any, of all structures in the EC-MOF/Phase-I database. 424 structures are found to be metallic with 259 of them being bulk and the rest of 165 being mono-layer systems. For the semiconductors, the maximum band gap value of 0.95 eV is reached by PtHITX_Mono. In Figure 3.11(a), we can find that most bulk materials are metallic or narrow gap semiconductors where

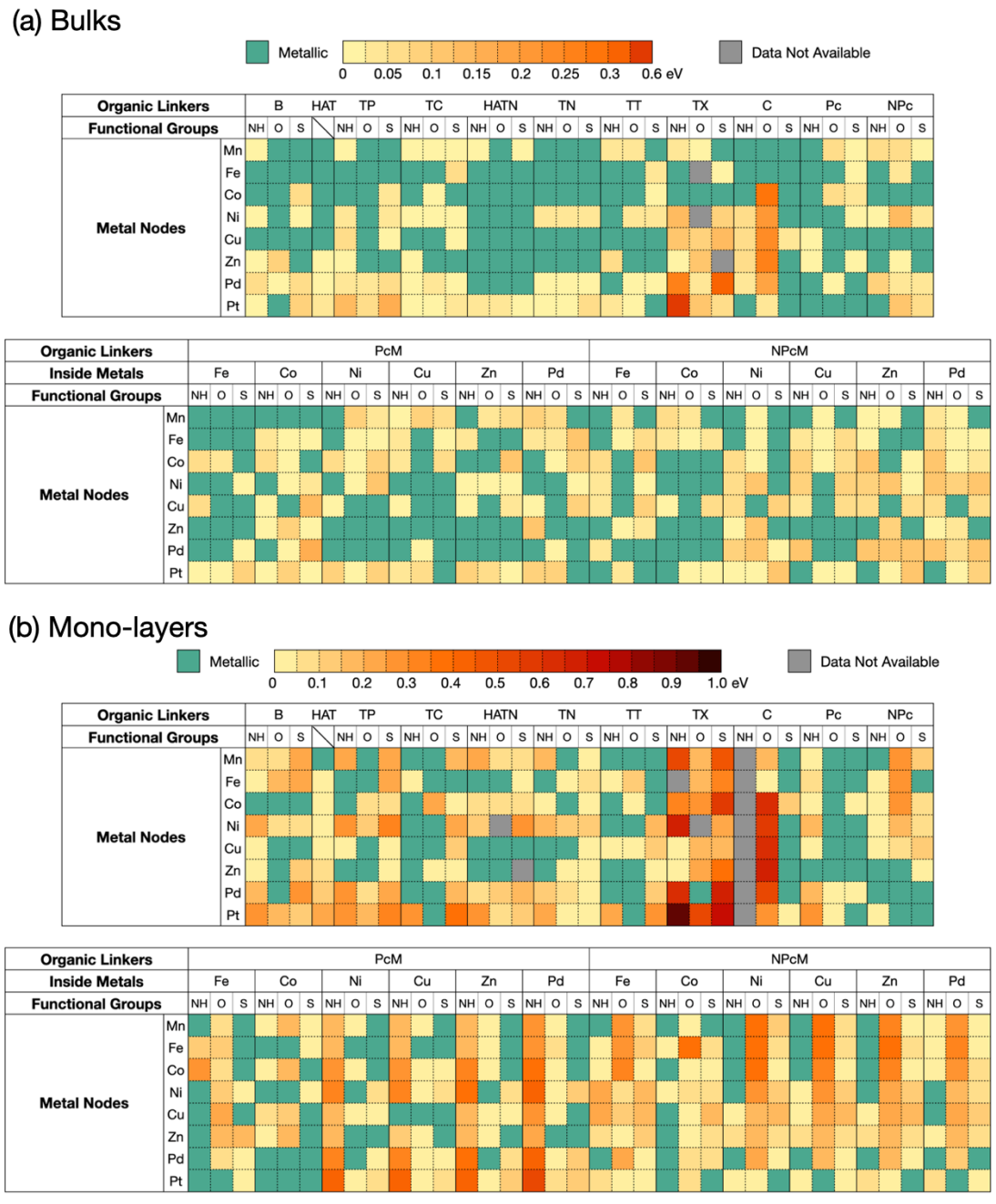

Figure 3.11 Calculated band gaps of (a) bulk and (b) mono-layer systems.

the band gap can be less than 0.15 eV. A maximum band gap of 0.53 eV is recorded for PtHITX_Bulk. In comparison, only one tenth (∼2000) of structures in curated 1D MOFs collection reported by Gharagheizi et al. [40] are metallic while another one tenth show a band gap below 1 eV. Furthermore, 546 1D MOFs are insulators (band gap $>$ 4 eV) while the band gap of most structures are in the range of 1-4 eV. It should be again emphasized that, apart from dimensionality that overall affects the properties of MOFs, this difference is related to the fact that we intentionally built our EC-MOF/Phase I database with structural features that induce a high degree of electrical conductivity in the resulted MOFs. According to our previous work [169], a transition of conductive behavior between metallic and semiconductor can be easily induced by temperature, pressure or solvent because of the intrinsic flexibility of these materials. Therefore, such conductive behaviors indicate the great advantage of EC-MOFs compared to conventional MOFs. Indeed, some of them have already been proven to be promising candidates for applications in batteries [17], photodetectors [7] and voltammetric detection[74]. In Figure 3.11(b), not only the number of semiconductors increases in mono-layers but the band gap values also increase with the highest band gaps being observed for TX linkers. According to Zhao et al. [172] who reported the successful synthesis of the CuHHTX MOF in 2020, the conductivity of CuHHTX shows an Arrhenius-type dependence on the temperature. The dependence is a typical sign of a semiconductor material which is in agreement with our calculated data. The large band gaps computed for the TX-based MOFs can be explained by the disconnection of the π-conjugation in the benzene rings of the TX linker. Therefore, one can hypothesize that the inter-layer conduction should play a more prominent role in these systems. PHC-based MOFs are another class that show wide band gaps. They have the highest number of connectivity, 12 hydroxyl groups per linker. As a result, the very electronegative oxygen atoms tend to localize the charge in the C linker leading to a wider band gap in these systems compared to other

systems. Another interesting trend is found in the PcM and NPcM-based mono-layers with an inside Ni, Cu, Zn and Pd metals. These mono-layers with imino functional groups always possess high band gap values because of the presence of the inside metals compared to the corresponding Pc and NPc-based mono-layers. However, no trend was found for the Fe and Co as inside metals. Therefore there could be some long range interactions at play between different metals and functional groups that needs to be further investigated.

3.3.4 Adsorption performance analysis

As stated above, the large surface areas and excellent conductivity of EC-MOFs makes them outstanding candidate materials for electronics and energy storage materials. They have been experimentally synthesized for electrode materials in sodium-ion [112], lithium-sulfur [17] and zinc [106] batteries. Therefore, it is of high interest to provide insights not only on the conductivity of EC-MOFs but also their capability of adsorbing different guest molecules relevant to these applications. Such capability is normally specified in terms of adsorption energy (E_{ads}) which is calculated as:

$$E_{ads} = E_{Guest-Host} - E_{Guest} - E_{Host}. \tag{3.4}$$

Here, $E_{Guest-Host}$ is the total energy of the guest adsorbed systems, E_{Guest} is the energy of the isolated guest molecules in their stable phases and E_{Host} is the energy of the host adsorbent materials. To obtain E_{ads} values, accurate DFT relaxations of the guest-host systems are required which are rather expensive considering the number of the studied materials. Also, for each different guest molecule, a separate relaxation of the entire guest-host system is required because there is no universal model for such interactions. Creation of EC-MOF database provides the foundation for screening hundreds of low-dimensional materials, utilizing appropriate descriptors, for adsorption purposes without expensive DFT relaxations of the guest-host systems.

For example, in the case of transition metals as adsorption sites, d-band center of metals can provide some preliminary insights into guest-host interactions without carrying out expensive calculations [50]. The position of d-band center is shown to be inversely related to the computed E_{ads} in many studies on metal alloys and also 2D MOFs [168, 89]. More specifically, as the energy of the d-band center increases with respect to the Fermi level, the stronger the interactions will become between the guests and the host materials [110]. The distribution of the d-band centers of bulk and mono-layer structures in EC-MOF Database, except PcM and NPcM linkers-based MOFs, is demonstrated in Figure 3.12.

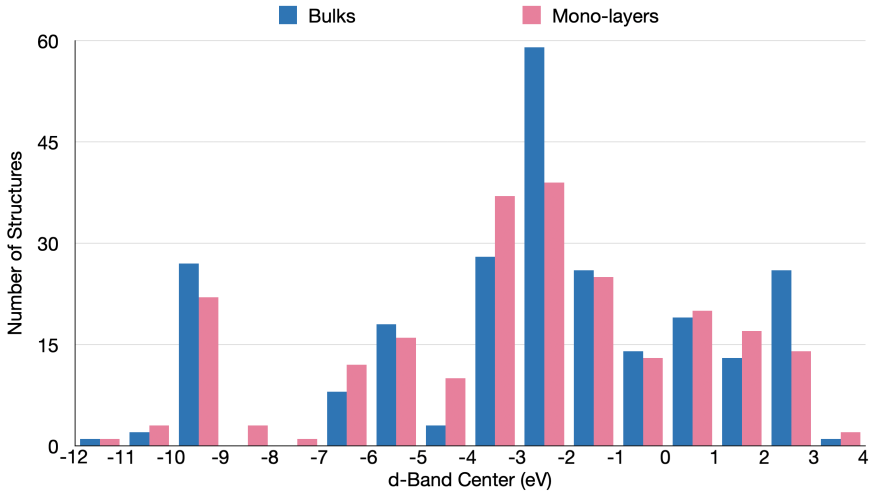

Figure 3.12 Distribution of the d-band center in Bulks and Mono-layers

Here, we present an exemplary case of using d-band center model to estimate the applicability of selected structures from EC-MOF database as cathode materials in Lithium-sulfur batteries (LSBs). The Li_2S molecule as the final product of the discharge process in LSBs is adapted here as the guest molecule. Two classes of well-studied B and TP-based MOFs, with imine functional groups, are chosen from our EC-MOF database to serve as the host materials. Comparing to the sp^3 hybridization of the hydroxyl and the thiol groups, the imine groups with the sp^2 hybridization have less effect on adsorption process because all three p orbitals are forming σ bonds or

conjugating with π framework of the MOF structures. The fully relaxed structures of the MHIB-Li_2S and MHITP-Li_2S systems (M = Mn, Fe, Co, and Ni) are shown in Figure 3.13.

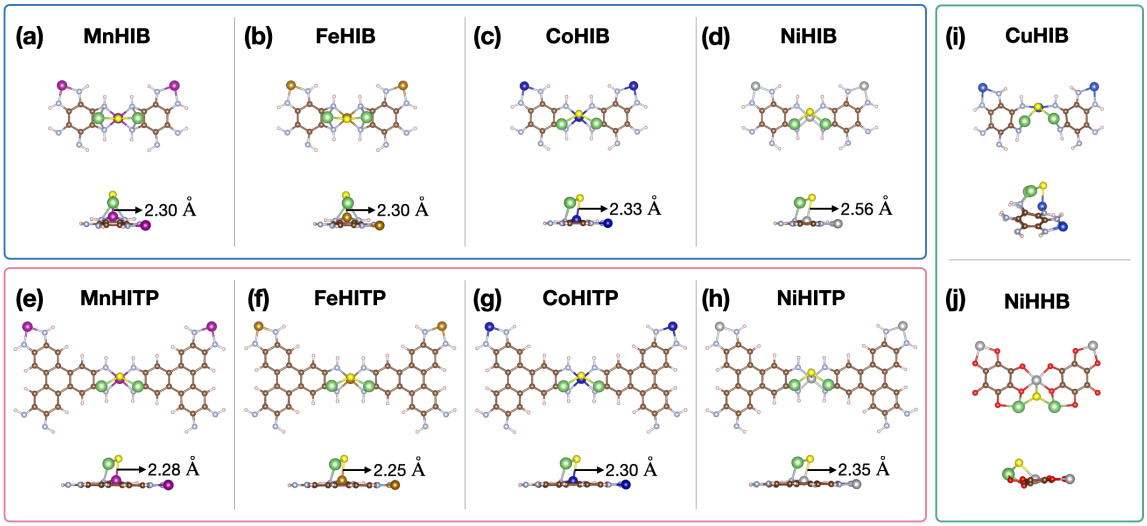

Figure 3.13 Fully relaxed MHIB(a–d) and MHITP(e–h) MOFs (M = Mn, Fe, Co, and Ni) with Li_2S as the guest molecule. Representative models in different adsorbing behaviors (i and j). In this figure the colors yellow, light blue, red, brown, green and cream represent sulfur, nitrogen, oxygen, carbon, lithium and hydrogen atoms, respectively, while metal atoms are distinguishable between different panels.

All systems containing Mn^{+2}, Fe^{+2}, Co^{+2} and Ni^{+2} metal centers show a similar adsorbing behavior where the Li_2S molecule is adsorbed on the top of the metal nodes. Therefore their adsorption energies should in principle follow the trend for the d-band center model. The calculated adsorption energies as obtained from Equation 3.4 are plotted vs the corresponding d-band center data in Figure 3.14, where we can see a clear inverse correlation as expected. The calculated length of the Metal-S bonds also indicates the same correlation to the adsorption energies. While we use this example as a proof of concept, care should be taken in choosing a proper descriptor for any specific screening purpose. For example, in the case of the other 4 metal nodes (i.e., M = Cu, Zn, Pd, and Pt), interaction between functional groups of the MOF and Li atoms of the Li_2S was found to be the dominant adsorption interaction, Figure 3.14

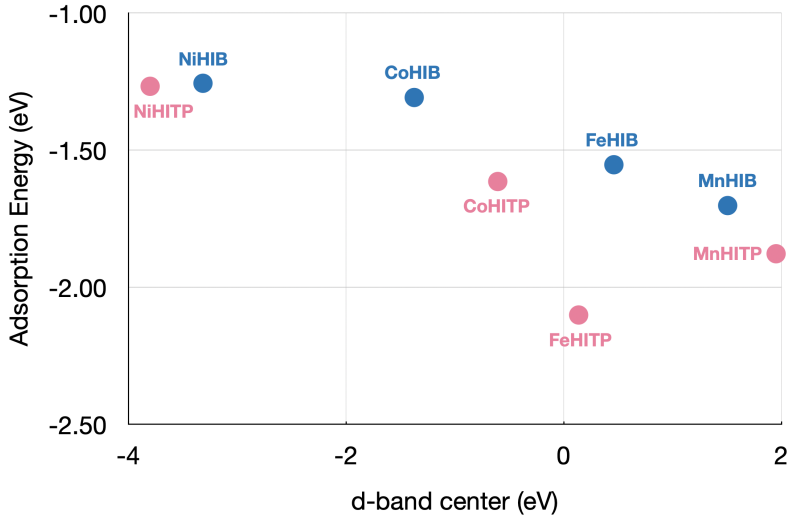

Figure 3.14 Computed adsorption energies of Li_2S molecules with selected 2D MOF mono-layers show an inverse relation with the corresponding calculated d-band centers.

(i). Thus, the calculated E_{ads}s were found to not correlate with the computed trend for the d-band center model. Similarly, NiHHB shows strong contribution of hydroxyl groups for Li_2S. On the other hand, d-band center has been shown to be a very appropriate descriptor to study adsorption of adsorbates like OH*, O* and OOH* species in oxygen reduction reaction [39, 175] where E_{ads}s are well correlated with the d-band center as long as the metal centers act as the primary adsorption sites.

3.4 Future Work

Our EC-MOF/Phase-I database provides an easily accessible computationally ready database for π-stacked layered electrically conductive MOFs with diverse information on their structural, electronic and adsorption properties. We plan to update this database as research in this field grows and expands. It is necessary to update this database in two directions, diversity of the MOFs included and types of properties calculated for them. As stated previously, we have elected to restrict the first version of the database according to the structural features that induce the highest electrical conductivity. However, as mentioned before, this is an active area of

research and new organic linkers are introduced almost everyday [48]. Also, MOFs with various topologies will be included in the future versions of the database, e.g., three-dimensional MOFs [138, 29]. Based on our building strategy, permutations of the building blocks will give a comprehensive understanding of these classes of MOFs as well. On the other hand, other relevant properties need to be calculated, such as performance on adsorption/storage of common gas molecules [27], partial atomic charges normally used to interpret trends while modeling chemical reactions [121] and density of states/band structures which reveal detailed charge transport pathways. Machine learning techniques have also shown great promise in materials science research for prediction of formation energies [43], adsorption energies [111, 148], band gap values [122] and designing new materials [97]. The created crystal structures and their calculated property data gathered in our EC-MOF/Phase-I database provide an ideal data set for applying various machine learning techniques in order to explore their potentials for different applications.

3.5 Concluding Remarks

In this work, we introduced for the first time an exclusive database for electrically conductive MOFs containing computationally-ready structures and their property data. A total number of 1,072 structures are created by taking permutations among the subsets of different structural building blocks using our in-house package, Crystal Structure Producer (CrySP). Multiple stages of calculations are applied to the database by applying a high-throughput screening workflow to optimize the structures and calculate their different property data. 1,063 out of 1,072 structures were successfully optimized at the DFT level and 1,061 of them successfully completed all stages of calculations whose properties including largest cavity diameter, gravimetric/volumetric surface area, void fraction, formation energy, inter-layer binding energy, electronic band gap and d-band center added into the database.

Obtained trends for different classes of MOFs were discussed in great details and different materials were classified and analyzed according to their structural or electronic properties providing comprehensive and important information on different families of EC-MOFs. Also, a limited scale screening of adsorption performance of Li_2S molecule is implemented to indicate the possibility of using the d-band center model as an efficient energy descriptor for adsorption energies. Finally, an exclusive graphical user interface is developed and released for this database where all curated structures at the DFT level can be visualized and downloaded with their relevant calculated properties tabulated.

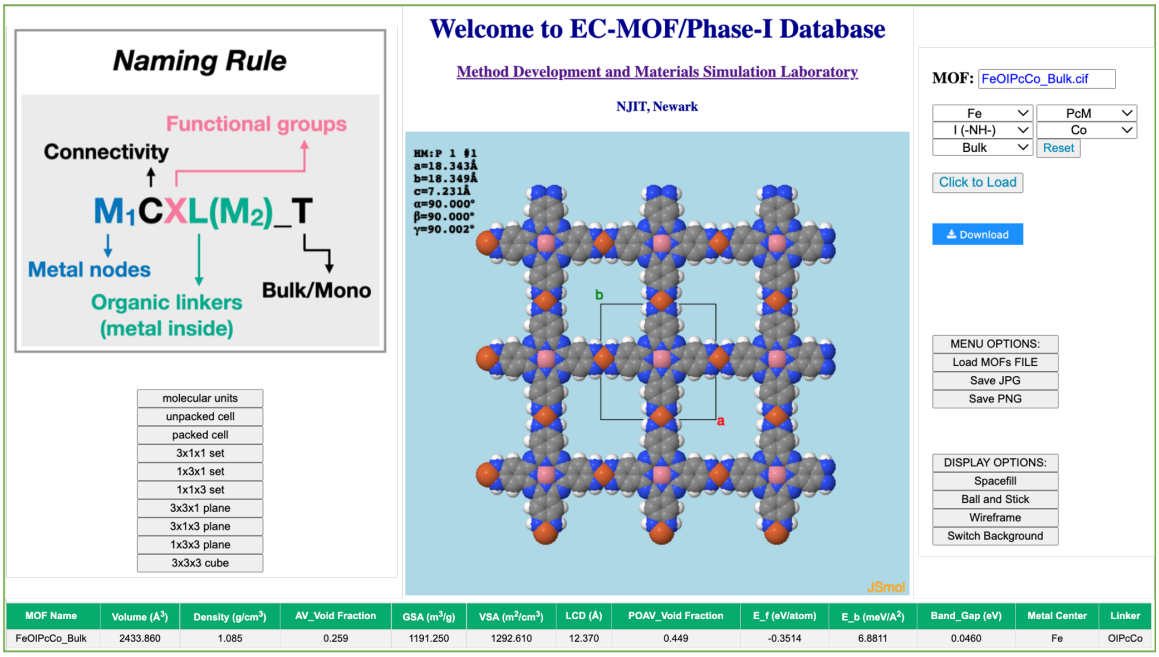

Figure 3.15 Screenshot of the web-based graphical user interface developed for the EC-MOF/Phase-I database.

CHAPTER 4
PREDICTING PROPERTY DATA IN EC-MOF/PHASE-I DATABASE USING MACHINE LEARNING

4.1 Motivation of the Database for Electrically Conductive MOFs

High-throughput (HT) screening is a useful tool to manage a big pool of data by implementing computational calculations at different stages and setting a series of thresholds for each stage until a certain number of candidates satisfy all the thresholds. As the quantity of data and the size of structures increases, such computational efforts could be inevitably expensive, especially if accurate ab initio calculations are desired. HT screening of EC-MOFs based on quantum-mechanical calculations, even if only at the highest stages, becomes daunting due to their chemical diversity and large size of unit cells [22]. On the other hand, while Density Functional Theory (DFT) calculations are widely used in HT screening, common functionals always underestimate the band gaps of semiconductors. As a result, hybrid functionals like HSE[56]or HLE[23] are required for accurate electronic structure calculations which need even more computational resources. Recently, machine learning (ML) techniques have become increasingly popular in scientific research and industrial production [36, 132]. In the field of computational chemistry, ML is mostly used to produce as close results as possible to ab initio calculations with much fewer computational costs [146]. It is very efficient in cases where the number of structures in a database is too high to implement HT computing or the computational cost of specific calculations is astronomical like evaluating electrical conductivity with a dense k point. To train a ML model for specific purpose, it is critical to find the appropriate features which sufficiently and properly describe the studied systems so that the target property can be reasonably deduced [22]. ML techniques have already been used to predict the gas storage and separation abilities of MOFs [147, 4, 156]; however, there are limited reports on its usage to predict the

electrical conductive behavior of EC-MOFs. This is mainly due to the extended size of their unit cells which causes difficulty in finding highly representative features. Therefore, to establish the foundation of applying ML techniques on EC-MOFs, the first step is a proper feature creation and engineering.

In this section, we create features and representations to predict the intrinsic properties of the structures in EC-MOF Database including metallicity, band gap values (E_g) and formation energy(E_f) using ML models. When building the database, calculating of such properties require expensive single-point (SP) calculations at DFT level. However, considering the increasing number and size of MOF structures, the HT computing calculations become computational demanding. The properties of bulk structures are highly effected by the interaction between layers like slipping behaviors which induces unnecessary variation. Hence, we only use mono-layers of EC MOFs in this study to focus on relating the intrinsic properties to the chemical compositions and geometry of mono-layers. Firstly, we apply generic statistical reduction methods (GSRM) to create representations of the unit cells based on elemental properties as shown in Table 4.1 [150]. GSRM can generate representations regardless of

Table 4.1 Elemental Information Used in Generic Statistical Reduction Methods to Create GSRM Features

Atomic number	Electronegativity
1st ionization energy (kj/mol)	Covalent radius (pm)
Van der waals radius (pm)	Melting point (K)
Row in periodic table	Column in periodic table
Number of unpaired electrons	Number of valence electrons
s orbital valence electrons	p orbital valence electrons
d orbital valence electrons	Number of unfilled states
Unfilled s states	Unfilled p states
Unfilled d states	

the size or the elemental diversity of unit cells. A total number of 17 elemental

properties is selected and 5 statistic quantities (standard mean, geometric mean, standard deviation, maximum value and minimum value) are calculated for each of the properties [54], as following:

$$\mu = \sum_{i=1}^{n} x_i f_i \tag{4.1}$$

$$a = (\prod_{i=1}^{n} x_i f_i)^{\frac{1}{n}} \tag{4.2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i f_i - \mu)^2}{n}} \tag{4.3}$$

$$Max = max\{x_1, x_2, \cdots, x_n\} \tag{4.4}$$

$$Min = min\{x_1, x_2, \cdots, x_n\} \tag{4.5}$$

$$f_i = \frac{N_i}{\sum_{i=1}^{n} N_i} \tag{4.6}$$

where x_i is the elemental information of $i-th$ element in the MOF composition, f_i is the atomic fraction of $i-th$ elements. Although it is shown that GSRM features perform well in predicting metallicity for inorganic compounds with small unit cells [54], they are not sufficient for EC MOFs which contain more than 100 atoms per unit cell. Secondly, we complement it by extracting the geometric data which already exists in our EC-MOF Database. Additionally we create new geometric and chemical features related to the electrical conductive properties which are not a part of the published database. A full list of 17 database features can be found in the Table 4.2.

As a result, the training set employed here consists of 102 columns of features. These

Table 4.2 Features from EC-MOF Database

Lattice constant a (Å)	Lattice constant b (Å)
Lattice constant c (Å)	Lattice angle α (degree)
Lattice angle β (degree)	Lattice angle γ (degree)
Volumn (cm^3)	Total number of atoms
MF bond ratio	CN bond ratio
CO bond ratio	CS bond ratio
CC bond ratio	PP bond ratio
Total number of bonds	MM distance (Å)
Total number of d electrons	

features are used to train various ML models for predicting target properties including metallicity, class 0 for metal and class 1 for semiconductor, band gap values (E_g) and formation energy (E_f). Finally, we further validate our ML models by creating a hypothetical structure that is similar but does not exist in our database. The organic linker of hypothetical MOFs is a derivative of 1,4,5,8,9,12-hexaazatriphenylene (HAT). We will compare the predicted values from the trained ML models to the quantum mechanical results calculated at DFT level.

4.2 Implementations of ML and Discussion

An overall workflow is illustrated in Figure 4.1. The data set is a matrix of 524 mono-layer structures times 102 features for each. Notably, 102 columns of features are not in the same order of magnitude, e.g. a feature like the row in the periodic table is single digit but another feature like the melting point could be up to three digits. Thus, features should be properly scaled into a reasonable range to eliminate any bias induced by different units or magnitudes. A data set of the target properties also consists of 524 rows according to the number of mono-layer structures and three

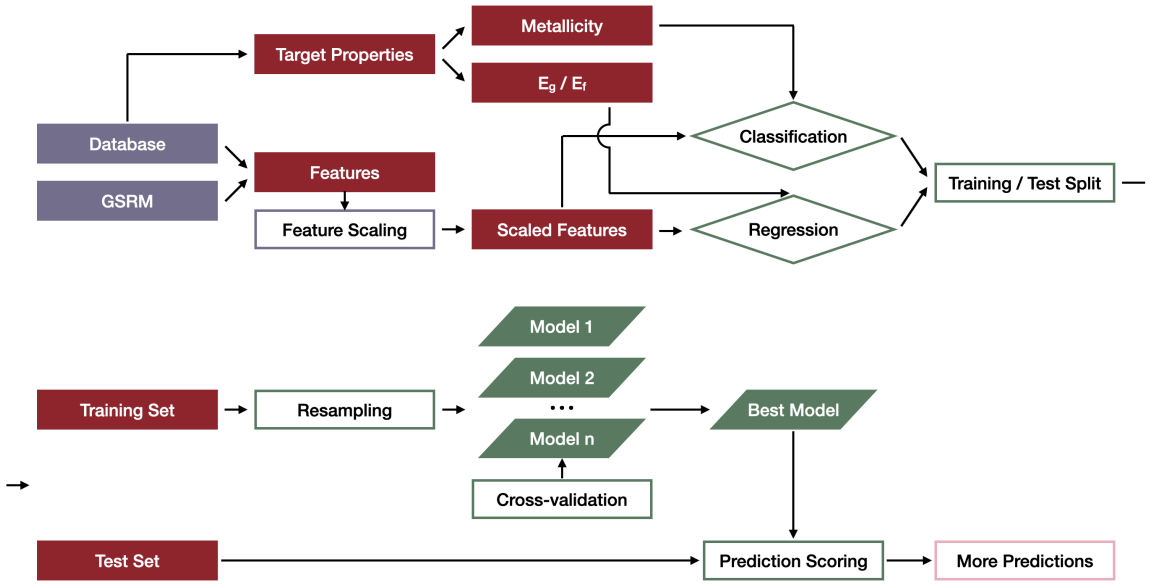

Figure 4.1 A complete workflow of ML implementation in our study. Data source is labeled by purple solid rectangles, data flows are labeled by red solid rectangles, ML functions are labeled by green hollow rectangles, ML models tested are labeled by green solid parallelograms and types of ML tasks are in green hollow diamonds.

target properties, metallicity (0 or 1), E_g and E_f. Figure 4.1 demonstrates two types of predictions, i.e., classification for metallicity and regression for band gap and formation energy values. Due to the limited number of data, the ratio of training/test set is set to 90%/10% to get a better training performance. In each case of training, various ML models are tested in order to select the best model for the specific task. A 10-fold cross-validation (CV) test is adopted in all models for the training set. The model with best performance is re-trained using the whole data set and will be used for making predictions related to hypothetical structures. Implementation of ML is carried out using scikit-learn package (version 1.2.2)[113]. The random state in all cases is set to 1 for reproducible results.

4.2.1 Classification of metallicity

Supervised classification are employed to predict the metallicity of the mono-layer EC MOFs. To avoid the unbalanced weight of different features, a pre-process of

Table 4.3 Optimal Settings for Tested Classifiers in Scikit-learn Package

Classifiers	Settings
RC	$alpha = 0.00001$
LR	$solver =' liblinear',\ C = 2$
SGD	$loss =' squared\_hinge',\ alpha = 0.001$
SVC	$nu = 0.3$
KNC	$Defaults$
GPC	$Defaults$
NN	$activation =' identity',\ solver =' lbfgs',\ alpha = 1$
DTC	$min\_samples\_leaf = 2$
ETC	$n\_estimators = 110,\ criterion =' log\_loss',\ min\_samples\_split = 10$
GBC	$n\_estimators = 120,\ learning\_rate = 1.0,\ max\_depth = 5$
BC	$n\_estimators = 25$
RFC	$n\_estimators = 110,\ criterion =' log\_loss'$
SC	$C = 2$

scaling the data is implemented as dividing the data by the maximum value of the column ($x/|Max|$) so that all data drops into a reasonable range, $[-1, 1]$ without changing the sparsity. By observing the distribution of the *whole data set*, there are 165 metallic MOFs and 359 seimconductors, indicating a slightly imbalanced distribution. An over-sampling strategy, adaptive synthetic sampling, is adopted to balance the *training set* and avoid any bias from the imbalanced distribution [52]. This is done using the imbalanced-learn package [47]. The over-sampling strategy results in a near-even ratio of 337 metallic and 319 semiconductive structures.

Twelve classification models are tested including Ridge Classification (RC), Logistic Regression (LR), Stochastic Gradient Descent classifier (SGD), Support Vector Machine classifier (SVC), k-Nearest Neighbors classifier (KNC), Gaussian Process Classification (GPC), Neural Network classifier (NNC), Decision Trees classifier (DTC), Extra Tree classifier (ETC), Gradient Boosted Decision Trees

classifier (GBC), Bagging classifier(BC) and Random Forest classifier (RFC). A stacking classifier (SC) that combines the results from multiple classifiers to reduce their biases, based on stacking generalization method[154], is also adopted and a LR classifier is used as the final classifier. For every model, optimal setting, listed in the Table 4.3, is chosen based on the mean training accuracy of the 10-fold CV. To test

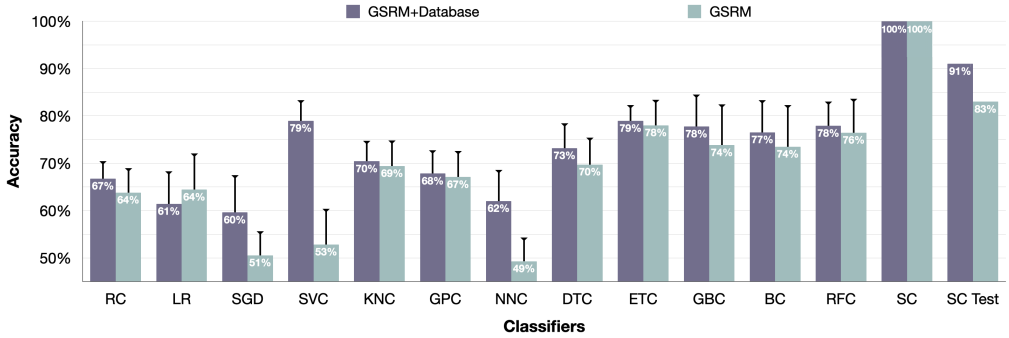

Figure 4.2 Accuracy (a) and standard deviation (b) of tested ML Classification models using GSRM+Database features (dark) and GSRM features only (light). Performance of SC on test set is attached on the right.

our assumption about the insufficiency of GSRM features in this study, all models are trained by (1) the whole 102 columns of features (GSRM+Database features) and (2) the GSRM features only. Figure 4.2 exhibits the accuracy of all chosen models which is higher or close to 50%. The error bars on the top of each accuracy bar are computed according to the standard deviation (SD) among 10 folds of CV process. The accuracy of SC in training (SC) and testing (SC Test) is also shown on the right. Overall, the training accuracy using GSRM+database features is always higher than the one using GSRM features, which shows the additional features from EC-MOF Database are necessary in our ML process. Also, the error bars of SD based on GSRM+Database features are generally lower than the GSRM features. Accordingly, including database features imparts a more stable performance to the CV process. Such stability is highly preferable in small data sets because the performance is less likely to be affected by the random split of training and test sets. Seven out of 12 classifiers reach an accuracy higher than 70%. The highest accuracy of single classifier

is reached by SVC and ETC, around 79%. Detailed evaluations of the classifications are shown in the Figure 4.3, where the low recall of class 1 indicates that the classifiers

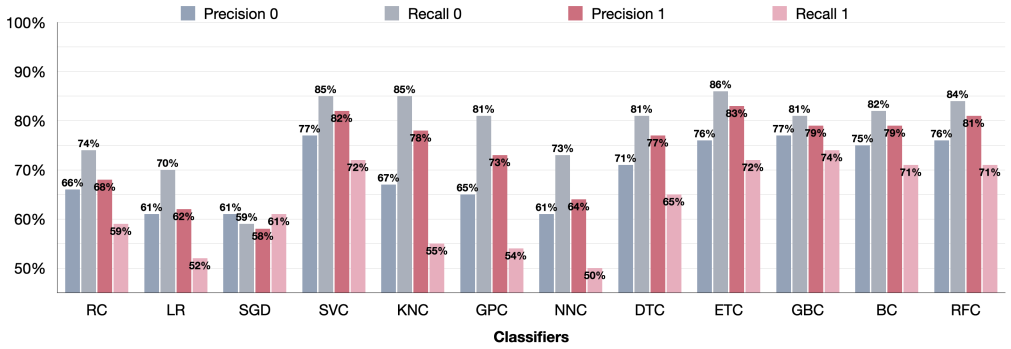

Figure 4.3 Evaluation metrics of the classifiers in metallicity classification.

have some difficulty in finding the semiconductors. Hence, such accuracy can be barely used in practice. On the other hand, the SC classifier that combines multiple classifiers is an ideal solution as shown in Figure 4.2. Different combinations are tested and the combination that gives best performance is the three highest accuracy in training, i.e., SVC, ETC and RFC plus DTC. With this combination, the training accuracy reaches 100% regardless of the choice of input features. However, the performance on the test set differs in that the SC trained by GSRM+Databse features results in 91% accuracy compared to the 83% of the SC trained by GSRM features. This observation further emphasizes that the addition of features from EC-MOF Database improves the stability and transferability of classification models. The final message in this section is that a combination of GSRM and EC-MOF Database features together with stacking of appropriate ML models creates a predictive tool for classification of EC MOFs to metal or semiconductor without resorting to expensive ab initio calculations.

4.2.2 Regression of band gaps

After classification of EC MOFs based on their metallicity, we use ML technique to predict the band gap values (E_g) of the semiconductive ones. E_g values are

significantly important for semiconductors by affecting the electrical conductivity of the materials. Therefore, the 359 semiconductors form another 359 $\times$ 102 training matrix for regression models. Same scaling process as classification task is implemented to avoid bias. Linear regression models including, Ridge regressor (RR), Linear Regression (LIR), Stochastic Gradient Descent regressor (SGD) and Passive Aggressive regressor (PAR), Support Vector Machine regressor (SVR), k-Nearest Neighbors regressor (KNR), Neural Network regressor (NNR) and ensemble methods including Extra Tree regressor (ETR), Gradient Boosted Decision Trees regressor (GBR), Bagging regressor (BR), AdaBoost regressor (ABR) and Random Forest regressor (RFR) are tested in regression of E_g values. We enforce positive E_g values by assigning "Positive" keyword as "True", if applicable. The optimal setting, listed in the Table 4.4, is chosen according to the mean absolute error (MAE) in eV for training and test sets. The range of band gap distribution among mono-layer structures

Table 4.4 Optimal Settings for Tested Regressors of E_g in Scikit-learn Package

Regressors	Settings
RR	$Defaults$
LIR	$Defaults$
SGD	$Defaults$
PAR	$Defaults$
SVR	$Defaults$
KNR	$Defaults$
NNR	$activation =' tanh',\ alpha = 1.0$
ETR	$n\_estimators = 110,\ criterion =' friedman\_mse'$
GBR	$n\_estimators = 120$
BR	$Defaults$
ABR	$n\_estimators = 60,\ learning\_rate = 2.0$
RFR	$n\_estimators = 110,\ criterion =' poisson',\ min\_samples\_split = 4$

from the EC-MOF Database is pretty narrow (less than 1 eV). Considering chemical

accuracy, the regression should not be expected to be perfectly linear between these DFT calculated values and the ML predicted values. Instead, the MAE is a more appropriate evaluator of this regression task. Table 4.5 demonstrates the MAE of training and test sets among all tested regressors.

Table 4.5 Mean Absolute Errors (MAE) of Band Gap Values among Tested Regressors for Training Set and Test Set

Regressor	Trainig set (eV)	Test set (eV)
RR	0.100	0.101
LIR	0.100	0.099
SGD	0.098	0.086
PAR	0.103	0.092
SVR	0.052	0.075
KNR	0.066	0.088
NNR	0.064	0.078
ETR	0.000	0.064
GBR	0.030	0.078
BR	0.029	0.075
ABR	0.075	0.083
RFR	0.029	0.068

The highest MAE is found from the linear regression models, RR, LIR, SGD and PAR, indicating a slightly worse performance than other types of models. However, the accuracy of DFT calculations with common functionals is $2 \sim 3\ kcal/mol$ ($0.087 \sim 0.130\ eV/particle$) [14]. As a result, all the tested regression models predict the E_g within the accuracy of DFT calculations in both training and test sets. The lowest MAE is reached by ETR where the regressor is perfectly fitted to the training data even with 10-fold CV. The MAE of ETR for test set is also the lowest, i.e., 0.06 eV, which refers to an accurate predictions in band gap values compared to previous reports [160, 122, 152, 2].

4.2.3 Regression of formation energies

Absolute formation energy, E_f (eV/atom), is an important factor of the thermodynamic stability of the materials, which can be computed by the following equation:

$$E_f = E_{tot} - \frac{1}{N}\sum_{i=1}^{N} x_i \mu_i. \tag{4.7}$$

Here E_{tot} is the calculated energy for the whole system, N is the total number of atoms with x_i and μ_i being the number and chemical potential of element i in the structure.Using the DFT calculated E_f of stable mono-layers ($E_f < 0$) in EC-MOF Database as target properties, we train the same regressors as used in E_g predictions. Scaling process is also implemented as usual. The optimal settings are listed in the Table 4.6

Table 4.6 Optimal Settings for Tested Regressors of E_f in Scikit-learn package

Regressors	Settings
RR	$Defaults$
LIR	$Defaults$
SGD	$penalty =' elasticnet'$
PAR	$Defaults$
SVR	$kernel =' rbf',\ C = 0.5,\ nu = 0.5$
KNR	$leaf\_size = 20$
NNR	$activation =' logistic',\ solver =' lbfgs'$
ETR	$criterion =' friedman\_mse',\ n\_estimators = 120,\ min\_samples\_split = 3$
GBR	$n\_estimators = 120$
BR	$n\_estimators = 20$
ABR	$loss =' square',\ n\_estimators = 70$
RFR	$criterion =' absolute\_error',\ n\_estimators = 80$

Notably, a "dimensional reduction" process is carried out to reduce the feature dimensions and improve the performance. In such a process, factor analysis method,

which is a simple linear generative model with Gaussian latent variables [8], is adopted. A series of tests between number of components and the coefficient of determination (R^2) during training is implemented. As a result, 102 features columns are reduced to 40 components to maximize the training score, R^2. Then, the 40 components are kept consistent in all regressors for E_f. The plots of calculated E_f values (DFT) vs. ML predicted values of all regressors are depicted in Figure 4.4. Predicted points during training are represented by a cumulative hexa bins. The

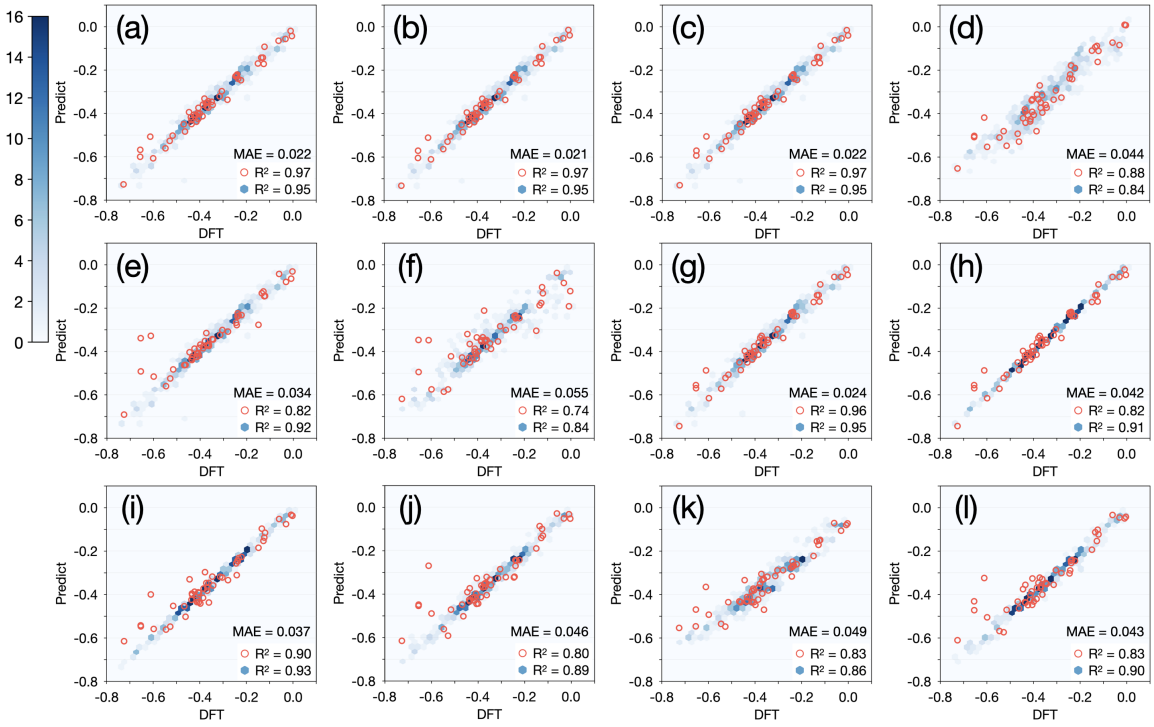

Figure 4.4 Coefficient of determination (R^2) of ML models in training (blue hexagonal bins) and test sets (red circles). The models tested are (a) RR, (b) LIR, (c) SGD, (d) PAR, (e) SVR, (f) KNR, (g) NNR, (h) ETR, (i) GBR, (j) BR, (k) ABR and (l) RFR. The unit for E_f is $eV/atom$.

number of points located in one bin is indicated by the blue color bar on the left. The predicted points in test set is shown as red circles. All plots show a nearly linear tendency between DFT calculated values and predicted E_fs. Three linear regression models, RR, LIR and SGD, perform the best in this task. The R^2 values are highest and above 0.95 for both training and test sets. Due to the random distribution of

instances in the test set, the R^2s are even 0.01 higher than training set, which is a negligible difference. The rest of regressors show average 0.90 and 0.84 R^2 values for the training and test sets, respectively, which should be credited to the created features that properly describe the systems and distinguish the difference among them. The lowest MAE of E_f predictions is given by LIR, 0.021 eV/atom, which indicates that LIR makes accurate predictions comparing to the DFT values. In line with other reports[160, 2, 6], we believe such MAE values indicate LIR is a good model for this task.

4.2.4 Validation by the hypothetical MOF

Implementing ML within the database is the first step in validating the effectiveness of created features and ML models for future investigations on EC MOFs. The final purpose is to get the important target properties for the future systems without running DFT calculations. For such a purpose, we have to make sure that the ML models used here are transferable to the MOFs not included in the EC-MOFs Database neither in the training process. To test this idea, we create a class of hypothetical MOFs based on the organic linker depicted in Figure 4.5. This organic

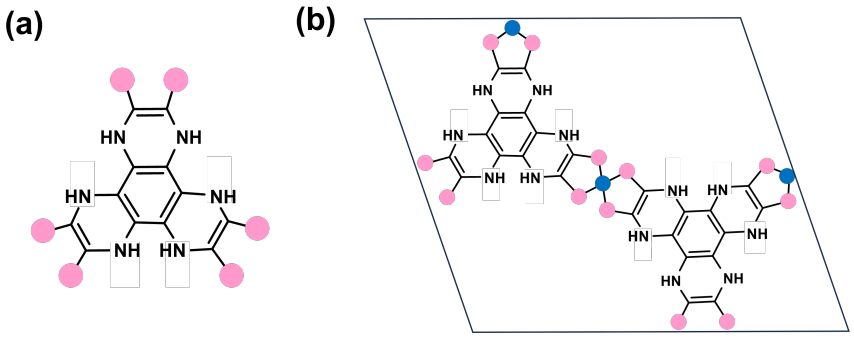

Figure 4.5 The structures of the hypothetical organic linker (a) and the hypothetical MOFs (b).

linker is a modified form of 1,4,5,8,9,12-hexaazatriphenylene (HAT) linker that has been already used to synthesize EC MOFs in 2022 [96]. The pink circles are the

replaceable points for functional groups, hydroxyl, imino, and thiol, and the blue circles are metal nodes which can be replaced with Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Pd^{2+} and Pt^{2+} cations. We utilize Crystal Structure Producer (CrySP) developed in our previous study[170] to survey all possible combination among the building blocks. As a result, a total number of 24 hypothetical EC MOFs are built. A High-throughput (HT) computing process that comprises of two steps of DFT calculations is employed on the built structures. DFT calculations are implemented in the Vienna ab initio simulation package (VASP) version 5.4.4 [80, 81, 79, 78]. Projector-augmented wave (PAW) potentials [13, 83] with a cut-off energy of 500 eV are used to describe the ion-electron interactions in the VASP calculations. Perdew-Burke-Ernzenhof (PBE) functional with Grimme's damped D3 dispersion correction within the generalized gradient approximation (GGA) formalism [115, 45] is chosen because it has been proved to be a ideal exchange-correlation functional for such MOF systems [169]. Spin-polarized effect is considered in all DFT calculations. The first step is geometric optimization of the build structures where the convergence thresholds of optimization is 0.02 eV/Åand electronic configuration minimization is $10^{-4}eV$ in each ionic step. During optimization, only lattice parameter c is set to 20 Åto avoid interaction between neighbor cells. The second step is single-point (SP) energy calculations that generate the target properties we need to validate the ML models. The k-points in the Brillouin zones is set to 3×1 and convergence criteria is increased to $10^{-4}eV$. To get an accurate E_g values, Hubbard U approach, a semi-empirically Coulomb interaction potential (U), is adopted in the SP energy calculations and U parameters are listed in Table 3.1 from our previous study [170].

For the input features of the hypothetical MOFs, 102 columns of features are generated in the same manner as mentioned above based on the optimized structures. Metallicity, E_g and E_f are again target properties. The best trained models, Stacking Classifier (SC), Extra Tree regressor (ETR) and Linear regressor

(LIR), are chosen to make predictions on metallicity, E_g and E_f, respectively. The SC classifier predicts that among the 24 hypothetical MOFs, six are metallic and 18 are semiconductive systems. To have a reference, we also carry out DFT calculations via a high-throughput (HT) approach where there is one metallic and 23 semiconductive MOFs. As a result, the accuracy of metallicity classification is 79%. Figure 4.6 shows the E_g values of the hypothetical MOFs obtained from the ETR model as well as E_f values predicted by LIR models, both trained using the GSRM+Database features. As can be seen from the color bar, the E_g values range from 0 to 0.230 eV. For comparison, the DFT calculated results are also shown alongside the predicted results by the ML models. The MAE of these predictions is 0.102 eV that is still within the accurcay of DFT calculations indicating acceptable transferability of our ML models. Regarding the formation energies, E_f values of all hypothetical structures are predicted by LIR models using the same dimensional reduction procedure as explained before. The MAE of E_f prediction is 0.141 eV/atom. The pattern of formation energies among different functional groups deduced from the ML predicted E_f values is that imino group creates the most stable MOFs while thiol group creates the least stable ones. Such pattern is also confirmed by the DFT calculations here. It is worth mentioning that this is the same pattern we have already reported for the EC-MOF Database. According to the comparison in Figure 4.6, the distribution of predicted values shows agreement with DFT results, which provide a decent possibilty that we can predict the properties using our created features and ML models to the future structures without extra DFT calculations.

4.3 Concluding Remarks

In this section, we explore feature creation for π-stacked layered electrically conductive MOFs. Using the created features, three machine learning (ML) tasks, i.e., classification of metallicity as well as prediction of E_g and E_f, are carried out

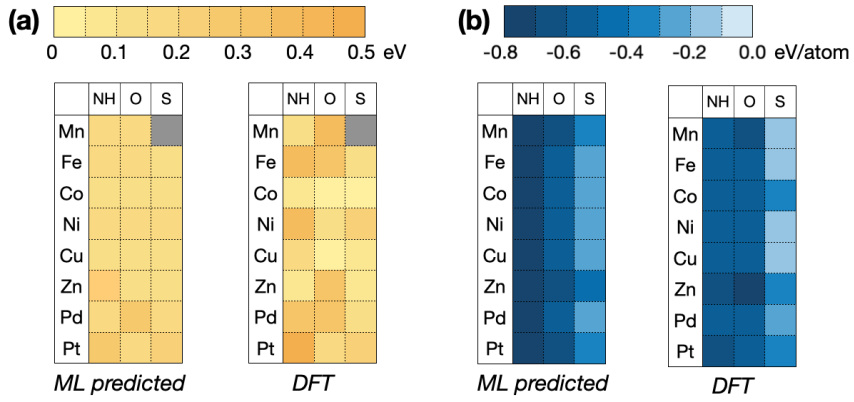

Figure 4.6 Comparison of band gap values (a) and formation energies (b) of hypothetical MOFs calculated by DFT and ML models.

with multiple ML models. The features computed by GSRM do not sufficiently describe the systems. Here, we show that we can remedy this shortcoming by complimenting the features with columns of data from our recently developed EC-MOF/Phase-I Database. Overall, the GSRM+Database features effectively improved the performance of all tested ML models. The final results show high accuracy for metallicity classification with the stacking classifier producing an accuracy of 100% for the training set and 91% for the test set. The mean absolute errors (MAE) of ML predicted E_g values (0.06 eV) are even smaller than the accuracy limit of DFT calculations. The R^2 values of E_f predicting are higher than 0.95 in the both training and test sets. Additionally, 24 hypothetical MOFs are built and are subjected to the best ML models to oredict target properties. The trained ML models show promising transferability to the MOFs that are not introduced in training process at all. We can reasonably deduct that if DFT results of a few other classes of MOFs are added into the training process, the accuracy of ML models will be greatly improved. In conclusion, the performance of ML models shows great potentials in using created features to correctly describe the EC MOF structures and also indicates that ML can be a useful tool to skip expensive DFT calculations.

CHAPTER 5

CONCLUSIONS

5.1 Concluding Remarks

In this dissertation, I develop a data-driven material discovery strategy with a special focus on two-dimensional (2D) or electrically-conductive (EC) metal organic frameworks (MOFs). Chapter 2 provides some insights into the flexible and dynamical structure of 2D MOFs, $Ni_3(HITP)_2$ specifically, and its effect on their intrinsic electrical conductive behavior. Due to contradictory reports from different research groups, it is not easy to determine the nature of conductive behavior of $Ni_3(HITP)_2$. To create reliable structures of $Ni_3(HITP)_2$, I take experimentally reported structure of another similar MOF, $Co_3(HHTP)_2$, as prototype. By removing unnecessary solvent molecules and replacing the metal nodes and functional groups, $Ni_3(HITP)_2$ is assembled. For the structural optimization, I implement a benchmark of different exchange-correlation (XC) functionals used in density functional theory (DFT). Comparing our optimized structure to the structural data from powder X-ray diffraction (PXRD) and crystal images from electron microscopy, the XC functional PBE-D3 is the method of choice. To determine charge transport pathways in 2D MOFs, the optimized structure is further modified by changing its inter-layer distance from 3.3 Å to 5 Å and replacing Ni^{2+} metal node with Zn^{2+}. Then, high accuracy electronic structure calculations at DFT level are implemented to get the density of states (DOS) and band structures (BS). Results indicated that the increase of inter-layer distance suppress the dispersion of bands in vertical direction of the cell, which confirms the through-bond in-plane charge transfer. The presence of Zn^{2+} suppresses the in-plane $d-\pi$ conjugation and shows the strong interaction between layers. Therefore, through-space charge transfer pathway is also proved to be a great contribution in such materials. Overall, these two exemplary cases confirm

the existence of a strong interconnection/competition between the two CT pathways where weakening of one may result in strengthening of the other. To mimic the dynamic motions of flexible $Ni_3(HITP)_2$, *ab initio* molecular dynamics (AIMD) is implemented at 1 atm pressure and 293 K temperature based on the optimized structures. The simulation lasts 8 ps and reaches equilibrium after 1 ps. The DOS and BS of all equilibrated structures indicate a band gap opening compared to the metallic $Ni_3(HITP)_2$ optimized at 0K. Therefore, I concluded that the discrepancy about the conductive behavior of 2D MOFs should be addressed with regard to the flexible nature of the material as well as different settings/conditions of the reported experimental studies. The intrinsic electrical conductive behavior of 2D MOFs is proved to be highly determined by its flexible nature and dynamic motions. Such highly sensitive structures need to be carefully prepared and measured in the future applications.

Modern material design procedure requires a more precise, comprehensive and efficient strategy than traditional trial-and-error mode. In Chapter 3, I introduce a strategy that utilizes the power of data and computer science to accelerate and assist material discovery process. First and foremost, this line of research needs a data pool that contains all possible information about one class of material, i.e., a material database. After browsing currently available databases, it was apparent that there is none that fully includes 2D EC MOFs. Based on my understanding of these materials, I carried out a thorough literature survey and collected a list of experimentally synthesized EC MOFs. By recognizing the building blocks of EC MOFs, my self-developed software, Crystal Structure Producer (CrySP), takes all possible permutations among the subsets of building blocks to create structures that fit periodic boundary conditions. Then, the high-throughput (HT) workflow, using my self-developed central control software, carries out multiple stages of DFT calculations on all the produced structures. Optimized cell structures, geometric

data and property data are collected from geometry optimization and single-point (SP) energy calculations to form the EC-MOF/Phase-I Database which is an online database and available to the public. From the optimized structures, geometric data such as surface area, pore diameters and void fraction are computed and show good agreement with the experimentally measured data. From SP calculations, the formation energy E_f as one important factor for the thermal stability of materials is computed. Most structures in the EC-MOF Database have negative E_f values, which means they should be possibly synthesized. The most crucial advantage, electrical conductive behavior, is shown in the form of metallicity. All created structures are conductive as metals or semiconductors with low band gaps ($E_g < 1$ eV). As an instance of material discovery for one specific purpose, i.e., electrode material in lithium sulfide batteries, the adsorption ability for Li_2S is tested using some representative EC MOFs. The correlation between calculated adsorption energies and the d-band center data indicates great potentials that we can filter materials using columns of data to select the best candidates for our interested application. In conclusion, I created a novel material design strategy that takes a few experimentally synthesized material as the starting point and reasonably expands upon them by creating hypothetical structures. This will end in building a comprehensive material database. Then, implementation of HT workflow will generate geometric and property data that can be used to filter structures for specific purpose. Such strategy can be further applied not only in MOF materials but all crystalline inorganic and organometallic compounds, or even organic compounds.

In Chapter 4, I argued that considering the expense of DFT or similar calculations at quantum mechanic level, HT computing becomes unaffordable for large numbers of structures. Machine learning technique seems a promising tool to learn the pattern of existing data and make predictions for the unknown. I created features for the EC MOFs using existing data, optimized structures from the EC-MOF Database

and generic statistical reduction methods (GSRM). The properties such as metallicity, E_g and E_f that require DFT calculations are recognized as target property. Three tasks of ML are initiated as (1) classification of metallicity, (2) prediction of E_g and (3) prediction of E_f using the created features. The ML models in classification task show great performance where accuracy reach 100% in training set and 91% in test set. For the prediction of E_g, the mean absolute errors (MAEs) of predictions compared to DFT results are as low as 0.06 eV which is lower than the common accuracy of DFT calculations, indicating an accurate predicting ability of our ML models based on the given features. Predictions of E_f have a coefficient of determination (R^2) above 0.95 for training and test sets. The MAE of predicted E_f is 0.021 eV/atom with the best ML model. In addition, a hypothetical MOF is created to validate the transferability of my features and ML models. The results are promising that the models are able to make close predictions even for the structures that are never introduced. As a result, ML will have a great potential to skip quantum-mechanical level calculations by making accurate predictions about our target properties, if sufficient training data is generated.

5.2 Future Work

The dissertation covers a complete workflow starting from the study of individual materials, to creating design strategy and applying ML technique to accelerate the discovery process. Nevertheless, there are still more interesting directions that can be further investigated. The first focus of research should be expanding the EC-MOF Database by introducing more building blocks on the one hand and creating more types of data on the other hand. As I mentioned, there are databases that possess near millions of structures. Therefore, to get a more comprehensive picture of EC MOF materials, more structures are needed. At the same time, the CrySP software should be improved by including more building logic circuits or finding a

more automatic logical process that can deal with different types of crystal lattices. The User Interface of our database should be modified so that users can download multiple structures and data at one time. Second research focus is related to the implementation of ML technique. The features I created till now show effectiveness for current target properties, but with the increase of structures and more columns of data, more features must be created to describe a diverse range of systems and their properties. A general feature engineering process should be established that is able to create and modify features regularly based on the newly added structures.

REFERENCES

[1] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, 1959.

[2] Alice EA Allen and Alexandre Tkatchenko. Machine learning of material properties: Predictive and interpretable multilinear models. *Science Advances*, 8(18):eabm7185, 2022.

[3] M. D. Allendorf, Z. Hulvey, T. Gennett, A. Ahmed, T. Autrey, J. Camp, Seon C. E., H. Furukawa, M. Haranczyk, M. Head-Gordon, S. Jeong, A. Karkamkar, D. Liu, J. R. Long, K. R. Meihaus, I. H. Nayyar, R. Nazarov, D. J. Siegel, V. Stavila, J. J. Urban, S. P. Veccham, and B. C. Wood. An assessment of strategies for the development of solid-state adsorbents for vehicular hydrogen storage. *Energy & Environmental Science*, 11(10):2784–2812, 2018.

[4] C. Altintas, O. F. Altundal, S. Keskin, and R. Yildirim. Machine learning meets with metal organic frameworks for gas storage and separation. *Journal of Chemical Information and Modeling*, 61(5):2131–2146, 2021.

[5] Y. Andersson, D. C. Langreth, and B. I. Lundqvist. van der waals interactions in density-functional theory. *Physical Review Letters*, 76:102–105, 1996.

[6] Luis M Antunes, Ricardo Grau-Crespo, and Keith T Butler. Distributed representations of atoms and materials for machine learning. *npj Computational Materials*, 8(1):44, 2022.

[7] H. Arora, R. Dong, T. Venanzi, J. Zscharschuch, H. Schneider, M. Helm, X. Feng, E. Cánovas, and A. Erbe. Demonstration of a broadband photodetector based on a two-dimensional metal–organic framework. *Advanced Materials*, 32(9):1907063, 2020.

[8] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[9] A. Benayad, D. Diddens, A. Heuer, A. N. Krishnamoorthy, M. Maiti, F. L. Cras, M. Legallais, F. Rahmanian, Y. Shin, H. Stein, et al. High-throughput experimentation and computational freeway lanes for accelerated battery electrolyte and interface development research. *Advanced Energy Materials*, 12(17):2102678, 2022.

[10] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[11] T. Björkman, A. Gulans, A. V. Krasheninnikov, and R. M. Nieminen. van der waals bonding in layered compounds from advanced density-functional first-principles calculations. *Physical Review Letters*, 108:235502, 2012.

[12] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50:17953–17979, Dec 1994.

[13] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953, 1994.

[14] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K. Müller, and K. Burke. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications*, 11(1):5223, 2020.

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

[16] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[17] D. Cai, M. Lu, L. Li, J. Cao, D. Chen, H. Tu, J. Li, and W. Han. Lithium–sulfur batteries: A highly conductive mof of graphene analogue $ni_3(hitp)_2$ as a sulfur host for high-performance lithium–sulfur batteries. *Small*, 15(44):1970240, 2019.

[18] M. G. Campbell, D. Sheberla, S. F. Liu, T. M. Swager, and M. Dincă. Cu_3(hexaiminotriphenylene)$_2$: An electrically conductive 2d metal–organic framework for chemiresistive sensing. *Angewandte Chemie International Edition*, 54:4349–4352, 2015.

[19] A. K. Cheetham, C. N. R. Rao, and R.K. Feller. Structural diversity and chemical trends in hybrid inorganic–organic framework materials. *Chemical Communications*, pages 4780–4795, 2006.

[20] S. Chen, J. Dai, and X.C. Zeng. Metal-organic kagome lattices m_3(2,3,6,7,10,11-hexaiminotriphenylene)$_2$ (m = ni and cu): from semiconducting to metallic by metal substitution. *Physical Chemistry Chemical Physics*, 17:5954–5958, 2015.

[21] T. Chen, J. H. Dou, L. Yang, C. Sun, N. J. Libretto, G Skorupskii, J. T. Miller, and M. Dincă. Continuous electrical conductivity variation in m3(hexaiminotriphenylene)2 (m = co, ni, cu) mof alloys. *Journal of the American Chemical Society*, 142:12367–12373, 2020.

[22] S. Chong, S. Lee, B. Kim, and J. Kim. Applications of machine learning in metal-organic frameworks. *Coordination Chemistry Reviews*, 423:213487, 2020.

[23] I. Choudhuri and D. G. Truhlar. Hle17: An efficient way to predict band gaps of complex materials. *The Journal of Physical Chemistry C*, 123(28):17416–17424, 2019.

[24] S. S. Y Chui, S. M. F Lo, J. P. H. Charmant, A. G. Orpen, and I. D. Williams. A chemically functionalizable nanoporous material [cu3 (tma) 2 (h2o) 3] n. *Science*, 283(5405):1148–1150, 1999.

[25] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K Farha, D. S Sholl, and R. Q. Snurr. Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials*, 26(21):6185–6192, 2014.

[26] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: Core mof 2019. *Journal of Chemical & Engineering Data*, 64(12):5985–5998, 2019.

[27] Y. J. Colón, D. A. Gomez-Gualdron, and R. Q. Snurr. Topologically guided, automated construction of metal–organic frameworks and their evaluation for energy-related applications. *Crystal Growth & Design*, 17(11):5801–5810, 2017.

[28] A. Corma and F. X. Garcís i Xamena. Engineering metal organic frameworks for heterogeneous catalysis. *Chemical Reviews*, 110:4606–4655, 2010.

[29] L. E. Darago, M. L. Aubrey, C. J. Yu, M. I. Gonzalez, and J. R. Long. Electronic conductivity, ferrimagnetic ordering, and reductive insertion mediated by organic mixed-valence in a ferric semiquinoid metal–organic framework. *Journal of the American Chemical Society*, 137(50):15703–15711, 2015.

[30] R. W. Day, D. K. Bediako, M. Rezaee, L. R. Parent, G. Skorupakii, M. Q. Arguilla, C. H. Hendon, I. Stassen, N. C. Gianneschi, P. Kim, and M. Dincă. Single crystals of electrically conductive two-dimensional metal-organic frameworks: Structural and electrical transport properties. *ACS Central Science*, 5:1959–1964, 2019.

[31] X. Di, Y. Wang, and N. Qian. Valley-contrasting physics in graphene: Magnetic moment and topological transport. *Physical Review Letters*, 99:236809, 2007.

[32] R. Dong, P. Han, H. Arora, M. Ballabio, M Karakus, Z Zhang, C. Shekhar, P. Adler, P. St. Petkov, A. Erbe, S. C. B. Mannsfeld, C. Felser, T. Heine, M. Bonn, X. Feng, and E. Cánovas. High–mobility band–like charge transport in a semiconducting two–dimensional metal–organic framework. *Nature Materials*, 17:1027–1032, 2018.

[33] R. Dong, Z. Zhang, D. C. Tranca, S. Zhou, M. Wang, P. Adler, Z. Liao, F. Liu, Y. Sun, W. Shi, Z. Zhang, E. Zschech, S. C. B. Mannsfeld, C. Felser, and X. Feng. A coronene-based semiconducting two-dimensional metal-organic framework with ferromagnetic behavior. *Nature Communications*, 9(1):1–9, 2018.

[34] J. Dou, M. Q. Arguilla, Y. Luo, J. Li, W. Zhang, L. Sun, J. L. Mancuso, L. Yang, T. Chen, L. R. Parent, G. Skorupskii, N. J. Libretto, C. Sun, M. C. Yang, P. V. Dip, E. J. Brignole, J.y T. Miller, J. Kong, C. H. Hendon, J. Sun, and M. Dincă. Atomically precise single-crystal structures of electrically conducting 2d metal–organic frameworks. *Nature Materials*, 20(2):222–228, 2021.

[35] J. H. Dou, L. Sun, Y. Ge, W. Li, C.H. Hendon, J. Li, S. Gul, J. Yano, E.A. Stach, and M. Dincă. Signature of metallic behavior in the metal-organic frameworks m_3(hexaiminobenzene)$_2$ (m = ni, cu). *Journal of the American Chemical Society*, 139:13608–13611, 2017.

[36] P. O. Dral. Quantum chemistry in the age of machine learning. *The Journal of Physical Chemistry Letters*, 11(6):2336–2347, 2020.

[37] M. E. Foster, K. Sohlberg, M.D. Allendorf, and A.A. Talin. Unraveling the semiconducting/metallic discrepancy in ni_3(hitp)$_2$. *The Journal of Physical Chemistry Letters*, 9:481–486, 2018.

[38] M. E. Foster, K. Sohlberg, C.D. Spataru, and M.D. Allendorf. Proposed modification of the graphene analogue ni_3(hitp)$_2$ to yield a semiconducting material. *The Journal of Physical Chemistry C*, 120:15001–15008, 2016.

[39] G. Gao, E. R. Waclawik, and A. Du. Computational screening of two-dimensional coordination polymers as efficient catalysts for oxygen evolution and reduction reaction. *Journal of Catalysis*, 352:579–585, 2017.

[40] F. Gharagheizi, Z. Yu, and D. S. Sholl. Curated collection of more than 20,000 experimentally reported one-dimensional metal–organic frameworks. *ACS Applied Materials & Interfaces*, 14(37):42258–42266, 2022.

[41] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Physical Review B*, 54:1703–1710, Jul 1996.

[42] D. A. Gómez-Gualdrón, T. C. Wang, P. García-Holley, R. M. Sawelewa, E. Argueta, R. Q. Snurr, J. T. Hupp, T. Yildirim, and O. K. Farha. Understanding volumetric and gravimetric hydrogen adsorption trade-off in metal–organic frameworks. *ACS Applied Materials & Interfaces*, 9(39):33419–33428, 2017.

[43] S. Gong, S. Wang, T. Xie, W. H. Chae, R. Liu, and J. C. Grossman. Calibrating dft formation enthalpy calculations by multifidelity machine learning. *JACS Au*, 2(9):1964–1977, 2022.

[44] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132:154104, 2010.

[45] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *Chemical Physics*, 132(15):154104, 2010.

[46] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward. The cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(2):171–179, 2016.

[47] L. Guillaume, N. Fernando, and K. A. Christos. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[48] B. Guo, R. Xu, J. Liang, L. Zou, A. Terfort, Z. Tian, P. Liu, T. Wang, and J. Liu. Dialytic synthesis of two-dimensional cu-based metal–organic frameworks for gas separation: Designable mof–polymer interface. *Inorganic Chemistry*, 61(40):16197–16202, 2022.

[49] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical Review B*, 59:7413–7421, Mar 1999.

[50] B. Hammer and J. K. Norskov. Why gold is the noblest of all the metals. *Nature*, 376(6537):238–240, 1995.

[51] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from h to rn. *Physical Review B*, 58:3641–3662, Aug 1998.

[52] H. He, Y. Bai, E. A. Garcia, and S. Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1322–1328. IEEE, 2008.

[53] X. He, M. Zhao, X. Tian, Y. Lu, S. Yang, Q. Peng, M. Yang, and W. Jiang. Redox-responsive nano-micelles containing trisulfide bonds to enhance photodynamic efficacy of zinc naphthalocyanine. *Chemical Physics Letters*, 803:139785, 2022.

[54] Y. He, E. D. Cubuk, M. D. Allendorf, and E. J. Reed. Metallic metal–organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *The Journal of Physical Chemistry Letters*, 9(16):4562–4569, 2018.

[55] W. Heisenberg. *Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen*. Springer, Berlin, Heidelberg, 1985.

[56] Jochen Heyd, Gustavo E Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of chemical physics*, 118(18):8207–8215, 2003.

[57] A. C. Hinckley, J. Park, J. Gomes, E. Carlson, and Z. Bao. Air-stability and carrier type in conductive m_3(hexaaminobenzene)$_2$, (m = co, ni, cu). *Journal of the American Chemical Society*, 142(25):11123–11130, 2020.

[58] M. Hmadeh, Z. Lu, Z. Liu, F. Gándara, H. Furukawa, S. Wan, V. Augustyn, R. Chang, L. Liao, F. Zhou, E. Perre, V. Ozolins, K. Suenaga, X. Duan, B. Dunn, Y. Yamamto, O. Terasaki, and O. M. Yaghi. New porous crystals of extended metal-catecholates. *Chemistry of Materials*, 24(18):3511–3513, 2012.

[59] B. F. Hoskins and R. Robson. Infinite polymeric frameworks consisting of three dimensionally linked rod-like segments. *Journal of the American Chemical Society*, 111(15):5962–5964, 1989.

[60] B. F. Hoskins and R. Robson. Design and construction of a new class of scaffolding-like materials comprising infinite polymeric frameworks of 3D-linked molecular rods. a reappraisal of the zinc cyanide and cadmium cyanide structures and the synthesis and structure of the diamond-related frameworks $[N(CH_3)_4][CuIZ_nII(CN)_4]$ and CuI[4,4',4",4"'-tetracyanotetraphenylmethane]$BF_4.xC_6H_5NO_2$. *Journal of the American Chemical Society*, 112(4):1546–1554, 1990.

[61] J. Huang, M. J. Golomb, S. R. Kavanagh, K. Tolborg, A. M. Ganose, and A. Walsh. Band gap opening from displacive instabilities in layered covalent-organic frameworks. *Journal of Materials Chemistry A*, 10:13500–13507, 2022.

[62] X. Huang, P. Sheng, Z. Tu, F. Zhang, J. Wang, H. Geng, Y. Zou, C. A. Di, Y. Yi, Y. Sun, and D. Zhu. Two–dimensional πd conjugated coordination polymer with extremely high electrical conductivity and ambipolar transport behavior. *Nature Communications*, 6:7408, 2015.

[63] X. Huang, S. Zhang, L. Liu, L. Yu, G. Chen, W. Xu, and D. Zhu. Superconductivity in a copper(ii)-based coordination polymer with perfect kagome structure. *Angewandte Chemie International Edition*, 57(1):146–150, 2018.

[64] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. cp2k: atomistic simulations of condensed matter systems. *WIREs Computational Molecular Science*, 4(1):15–25, 2014.

[65] R. Iftimie, P. Minary, and M. E. Tuckerman. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences*, 102(19):6654–6659, 2005.

[66] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science*, 50(8):2295–2310, 2011.

[67] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[68] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. Müller, and A. Tkatchenko. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical Reviews*, 121(16):9816–9872, 2021.

[69] U. Khan, A. Nairan, J. Gao, and Q. Zhang. Current progress in 2d metal–organic frameworks for electrocatalysis. *Small Structures*, page 2200109, 2022.

[70] M. Kim, J. Im, A. J. Freeman, J. Ihm, and H. Jin. Switchable s = 1/2 and j = 1/2 rashba bands in ferroelectric halide perovskites. *Proceedings of the National Academy of Sciences of the United States of America*, 111:6900–6904, 2014.

[71] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton. The open quantum materials database (oqmd): Assessing the accuracy of dft formation energies. *npj Computational Materials*, 1(1):1–15, 2015.

[72] K. Kitahara, Y. Takagiwa, and K. Kimura. Semimetallic band structure and cluster-based description of a cubic quasicrystalline approximant in the al–cu–ir system. *Journal of the Physical Society of Japan*, 84(1):014703, 2015.

[73] J. Klimeš and A. Michaelides. Perspective: Advances and challenges in treating van der waals dispersion forces in density functional theory. *The Journal of Chemical Physics*, 137(12):120901, 2012.

[74] M. Ko, L. Mendecki, A. M. Eagleton, C. G. Durbin, R. M. Stolz, Z. Meng, and K. A. Mirica. Employing conductive metal–organic frameworks for voltammetric detection of neurochemicals. *Journal of the American Chemical Society*, 142(27):11717–11733, 2020.

[75] M. Ko, L. Mendecki, and K. A. Mirica. Conductive two-dimensional metal–organic frameworks as multifunctional materials. *Chemical Communications*, 54(57):7873–7891, 2018.

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

[77] M. Krack. Pseudopotentials for h to kr optimized for gradient-corrected exchange-correlation functionals. *Theoretical Chemistry Accounts*, 114(1):145–152, 2005.

[78] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54:11169–11186, Oct 1996.

[79] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15 – 50, 1996.

[80] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47:558–561, Jan 1993.

[81] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 49:14251–14269, May 1994.

[82] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59:1758–1775, Jan 1999.

[83] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758, 1999.

[84] I. Kulu, V. Mantareva, V. Kussovski, I. Angelov, and M. Durmuş. Effects of metal ion in cationic pd (ii) and ni (ii) phthalocyanines on physicochemical and photodynamic inactivation properties. *Journal of Molecular Structure*, 1247:131288, 2022.

[85] C. Kung, P. Han, C. Chuang, and K. C-W Wu. Electronically conductive metal–organic framework-based materials. *APL Materials*, 7(11):110902, 2019.

[86] D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard, and B. L. Lundqvist. Van der waals density functional theory with applications. *International Journal of Quantum Chemistry*, 101:599 –610, 2005.

[87] F. Li, X. Zhang, X. Liu, and M. Zhao. Novel conductive metal–organic framework for a high-performance lithium–sulfur battery host: 2d cu-benzenehexathial (bht). *ACS Applied Materials & Interfaces*, 10(17):15012–15020, 2018.

[88] H. Li, M. Eddaoudi, M O'Keeffe, and O. M. Yaghi. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature*, 402:276–279, 1999.

[89] T. Li, C. He, and W. Zhang. Two-dimensional porous transition metal organic framework materials with strongly anchoring ability as lithium-sulfur cathode. *Energy Storage Materials*, 25:866–875, 2020.

[90] W. Li, K. Ding, H. Tian, M. Yao, B. Nath, W. Deng, Y. Wang, and G. Xu. Conductive metal–organic framework nanowire array electrodes for high-performance solid-state supercapacitors. *Advanced Functional Materials*, 27(27):1702067, 2017.

[91] W. M. Liao, J. H. Zhang, S. Y. Yin, H. Lin, X. Zhang, J. Wang, K. Wang, H. P.and Wu, Z. Wang, Y. N. Fan, M. Pan, and C. Y. Su. Tailoring exciton and excimer emission in an exfoliated ultrathin 2d metal-organic framework. *Nature Communications*, 9:2401, 2018.

[92] J. Liu, Y. Chen, X. Feng, and R. Dong. Conductive 2d conjugated metal–organic framework thin films: Synthesis and functions for (opto-) electronics. *Small Structures*, 3(5):2100210, 2022.

[93] J. Liu, Y. Zhou, Z. Xie, Y. Li, Y. Liu, J. Sun, Y. Ma, O. Terasaki, and L. Chen. Conjugated copper-catecholate framework electrodes for efficient energy storage. *Angewandte Chemie International Edition*, 59:1081–1086, 2020.

[94] L. Liu, J. Chen, Y. Zhang, C. Yu, W. Du, X. Sun, J. Zhang, F. Hu, Y. Mi, and L. Ma. Fabrication of ultrathin single-layer 2d metal–organic framework nanosheets with excellent adsorption performance via a facile exfoliation approach. *Journal of Materials Chemistry A*, 9(1):546–555, 2021.

[95] C. Lv, X. Zhou, L. Zhong, C. Yan, M. Srinivasan, Z. W. Seh, C. Liu, H. Pan, S. Li, Y. Wen, et al. Machine learning: an advanced platform for materials development and state prediction in lithium-ion batteries. *Advanced Materials*, 34(25):2101474, 2022.

[96] C. Lyu, Y. Gao, Z. Gao, S. Mo, M. Hua, E. Li, S. Fu, J. Chen, P. Liu, L. Huang, and N. Lin. Synthesis of single-layer two-dimensional metal–organic frameworks $m_3(hat)_2$ (m= ni, fe, co, hat= 1, 4, 5, 8, 9, 12-hexaazatriphenylene) using an on-surface reaction. *Angewandte Chemie International Edition*, 134(27):e202204528, 2022.

[97] H. Mai, T. C. Le, D. Chen, D. A. Winkler, and R. A. Caruso. Machine learning for electrocatalyst and photocatalyst design and discovery. *Chemical Reviews*, 122(16):13478–13515, 2022.

[98] G. W. Mann, K. Lee, M. Cococcioni, B. Smit, and J. B. Neaton. First-principles hubbard u approach for small molecule binding in metal-organic frameworks. *The Journal of Chemical Physics*, 144(17):174104, 2016.

[99] Y. Mao, H. Yang, Y. Sheng, J. Wang, R. Ouyang, C. Ye, J. Yang, and W. Zhang. Prediction and classification of formation energies of binary compounds by machine learning: an approach without crystal structure information. *ACS Omega*, 6(22):14533–14541, 2021.

[100] Z. Meng, A. Aykanat, and K. A. Mirica. Welding metallophthalocyanines into bimetallic molecular meshes for ultrasensitive, low-power chemiresistive detection of gases. *Journal of the American Chemical Society*, 141(5):2046–2053, 2018.

[101] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. GP Maloney, P. A. Wood, S. C. Ward, and D. Fairen-Jimenez. Development of a cambridge structural database subset: a collection of metal–organic frameworks for past, present, and future. *Chemistry of Materials*, 29(7):2618–2625, 2017.

[102] M. R. Momeni and C. J. Cramer. Structural characterization of pristine and defective $[zr_{12}(\mu_3{-}o)_8(\mu_3{-}oh)_8(\mu_2{-}oh)_6]^{18+}$ double-node metal-organic framework and predicted applications for single-site catalytic hydrolysis of sarin. *Chemistry of Materials*, 30(13):4432–4439, 2018.

[103] M. R. Momeni and C. J. Cramer. Computational screening of roles of defects and metal substitution on reactivity of different single– vs double–node metal–organic frameworks for sarin decomposition. *The Journal of Physical Chemistry C*, 123:15157–15165, 2019.

[104] M. R. Momeni, Z. Zhang, and F. A. Shakib. Deterministic role of structural flexibility on catalytic activity of conductive 2d layered metal–organic frameworks. *Chemical Communications*, page DOI:10.1039/D0CC07430F, 2021.

[105] J. E. Mondloch, M. J. Katz, W. C. Isley, P. Ghosh, P. Liao, W. Bury, Wagner G. W., M. G. Hall, J. B. DeCoste, G. W. Peterson, R. Q. Snurr, C. J. Cramer, J. T. Hupp, and O. K. Farha. Destruction of chemical warfare agents using metal–organic frameworks. *Nature Materials*, 14:512–516, 2015.

[106] K. W. Nam, S. S. Park, Dos R. R., V. P. Dravid, H. Kim, C. A. Mirkin, and J. F. Stoddart. Conductive 2d metal-organic framework for high-performance cathodes in aqueous rechargeable zinc batteries. *Nature Communications*, 10(1):1–10, 2019.

[107] D. Nazarian, J. S. Camp, Y. G. Chung, R. Q. Snurr, and D. S. Sholl. Large-scale refinement of metal- organic framework structures using density functional theory. *Chemistry of Materials*, 29(6):2521–2528, 2017.

[108] D. Nazarian, J. S. Camp, and D. S. Sholl. A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. *Chemistry of Materials*, 28(3):785–793, 2016.

[109] N. Nnaji, P. Sen, and T. Nyokong. Symmetry effect of cobalt phthalocyanines on the aluminium corrosion inhibition in hydrochloric acid. *Materials Letters*, 306:130892, 2022.

[110] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen. Towards the computational design of solid catalysts. *Nature Chemistry*, 1(1):37–46, 2009.

[111] G. Panapitiya, G. Avendaño-Franco, P. Ren, X. Wen, Y. Li, and J. P. Lewis. Machine-learning prediction of co adsorption in thiolated, ag-alloyed au nanoclusters. *Journal of the American Chemical Society*, 140(50):17508–17514, 2018.

[112] J. Park, M. Lee, D. Feng, Z. Huang, A. C. Hinckley, A. Yakovenko, X. Zou, Y. Cui, and Z. Bao. Stabilization of hexaaminobenzene in a 2d conductive metal–organic framework for high power sodium storage. *Journal of the American Chemical Society*, 140(32):10315–10323, 2018.

[113] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[114] J. P. Perdew, k. Burke, and M. Ernzerhof. Erratum: Generalized gradient approximation made simple. *Physical Review Letters*, 78:1396–1396, 1996.

[115] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.

[116] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. a. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Physical Review Letters*, 100:136406, 2008.

[117] H. TB. Pham, J. Y. Choi, S. Huang, X. Wang, A. Claman, M. Stodolka, S. Yazdi, S. Sharma, W. Zhang, and J. Park. Imparting functionality and enhanced surface area to a 2d electrically conductive mof via macrocyclic linker. *Journal of the American Chemical Society*, 144(23):10615–10621, 2022.

[118] L. Principe. In retrospect: the sceptical chymist. *Nature*, 469:30–31, 2011.

[119] F. A. Rasmussen and K. S. Thygesen. Computational 2d materials database: Electronic structure of transition-metal dichalcogenides and oxides. *The Journal of Physical Chemistry C*, 119(23):13169–13183, 2015.

[120] S. Rojas and P. Horcajada. Metal–organic frameworks for the removal of emerging organic contaminants in water. *Chemical Reviews*, 120(16):8378–8415, 2020.

[121] A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein, and R. Q. Snurr. High-throughput predictions of metal–organic framework electronic properties: Theoretical challenges, graph neural networks, and data exploration. *npj Computational Materials*, 8(1):1–10, 2022.

[122] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, 2021.

[123] A. S. Rosen, J. M. Notestein, and R. Q. Snurr. Identifying promising metal–organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *Journal of Computational Chemistry*, 40(12):1305–1318, 2019.

[124] N. L. Rosi, J. Kim, M. Eddaoudi, B. Chen, M. O'Keeffe, and O. M. Yaghi. Rod packings and metal- organic frameworks constructed from rod-shaped

secondary building units. *Journal of the American Chemical Society*, 127(5):1504–1518, 2005.

[125] A. Rycerz, J. Tworzydlo, and C. W. J. Beenakker. Valley filter and valley valve in graphene. *Nature Physics*, 3(3):172–175, 2007.

[126] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *The Journal of The Minerals*, 65:1501–1509, 2013.

[127] S. Sakaida, T. Haraguchi, K. Otsubo, O. Sakata, A. Fujiwara, M. Takata, and H. Kitagawa. Fabrication and structural characterization of an ultrathin film of a two-dimensional-layered metal-organic framework, $Fe(py)_2[Ni(CN)_4]$ (py=pyridine). *Inorganic Chemistry*, 56:7606–7609, 2017.

[128] S. Sakaida, K. Otsubo, O. Sakata, C. Song, A. Fujiwara, M. Takata, and H. Kitagawa. Crystalline coordination framework endowed with dynamic gate-opening behaviour by being downsized to a thin film. *Nature Chemistry*, 8:377–383, 2016.

[129] T. Sato, T. Tsuneda, and K. Hirao. Van der waals interactions studied by density functional theory. *Molecular Physics*, 103:1151–1164, 2005.

[130] E. Schrödinger. The schrödinger equation. *Physical Review*, 28(6):1049–1070, 1926.

[131] D. Sheberla, L. Sun, M.A. Blood-Forsythe, S. Er, C.R. Wade, C.K. Brozek, A. Aspuru-Guzik, and M. Dincă. High electrical conductivity in ni_3(2,3,6,7,10,11- hexaiminotriphenylene)$_2$, a semiconducting metal-organic graphene analogue. *Journal of the American Chemical Society*, 136:8859–8862, 2014.

[132] B. Shen, R. Gnanasambandam, R. Wang, and Z. J. Kong. Multi-task gaussian process upper confidence bound for hyperparameter tuning and its application for simulation studies of additive manufacturing. *IISE Transactions*, 55(5):496–508, 2023.

[133] Y. Shi, M. R. Momeni, Y. J. Chen, Z. Zhang, and F. A. Shakib. Water–induced structural transformations in flexible two–dimensional layered conductive metal–organic frameworks. *Chemistry of Materials*, 32:9664–9674, 2020.

[134] D. S. Sholl and J. A. Steckel. *Density Functional Theory: a Practical Introduction*. John Wiley & Sons, Hoboken, New Jersey, 2011.

[135] G. Skorupskii, B.A. Trump, T.W. Kasel, C.M. Brown, C. H. Hendon, and M Dincă. Efficient and tunable one–dimensional charge transport in layered lanthanide metal–organic frameworks. *Nature Chemistry*, 12:131–136, 2020.

[136] M. P. Suh, H. J. Park, T. K. Prasad, and D. W. Lim. Hydrogen storage in metal organic frameworks. *Chemical Reviews*, 112:782–835, 2012.

[137] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T. H. Bae, and J. R. Long. Carrbon dioxide capture in metal-organic frameworks. *Chemical Reviews*, 112:724–781, 2012.

[138] L. Sun, C. H. Hendon, S. S. Park, Y. Tulchinsky, R. Wan, F. Wang, A. Walsh, and M. Dincă. Is iron unique in promoting electrical conductivity in mofs? *Chemical Science*, 8(6):4450–4457, 2017.

[139] L. Sun, T. Miyakai, S. Seki, and M. Dincă. Mn2 (2, 5-disulfhydrylbenzene-1, 4-dicarboxylate): a microporous metal–organic framework with infinite (-mn-s-) ∞ chains and high intrinsic charge mobility. *Journal of the American Chemical Society*, 135(22):8185–8188, 2013.

[140] F. W. Thomas, H. R. Chris, K. Michaeel, C. M. Juan, and H. Maciej. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials*, 149(1):134–141, 2012.

[141] R. Vasudevan, G. Pilania, and P. V. Balachandran. Machine learning for materials design and discovery. *Journal of Applied Physics*, 129(7), 2021.

[142] P. Verma and D. G. Truhlar. Hle17: An improved local exchange–correlation functional for computing semiconductor band gaps and molecular excitation energies. *The Journal of Physical Chemistry C*, 121(13):7144–7154, 2017.

[143] K. Wada, K. Sakaushi, S. Sasaki, and H. Nishihara. Multielectron-transfer-based rechargeable energy storage of two-dimensional coordination frameworks with non-innocent ligands. *Angewandte Chemie International Edition*, 130(29):9024–9028, 2018.

[144] M. Wang, R. Dong, and X. Feng. Two-dimensional conjugated metal–organic frameworks (2d c-mofs): Chemistry and function for moftronics. *Chemical Society Reviews*, 50(4):2764–2793, 2021.

[145] M. Wang, H. Shi, P. Zhang, Z. Liao, M. Wang, H. Zhong, F. Schwotzer, A. S. Nia, E. Zschech, S. Zhou, S. Kaskel, R. Dong, and X. Feng. Phthalocyanine-based 2d conjugated metal-organic framework nanosheets for high-performance micro-supercapacitors. *Advanced Functional Materials*, 30(30):2002664, 2020.

[146] P. Wang, J. Xing, X. Jiang, and J. Zhao. Transition-metal interlink neural network: Machine learning of 2d metal–organic frameworks with high magnetic anisotropy. *ACS Applied Materials & Interfaces*, 14(29):33726–33733, 2022.

[147] Ruihan Wang, Yeshuang Zhong, Leming Bi, Mingli Yang, and Dingguo Xu. Accelerating discovery of metal–organic frameworks for methane adsorption with hierarchical screening and deep learning. *ACS Applied Materials & Interfaces*, 12(47):52797–52807, 2020.

[148] X. Wang, S. Ye, W. Hu, E. Sharman, R. Liu, Y. Liu, Y. Luo, and J. Jiang. Electric dipole descriptor for machine learning prediction of catalyst surface–molecular adsorbate interactions. *Journal of the American Chemical Society*, 142(17):7737–7743, 2020.

[149] X. Wang, L. Zhai, Wang. Y., R. Li, X. Gu, Y. D. Yuan, Y. Qian, Z. Hu, and D. Zhao. Improving water-treatment performance of zirconium metal-organic framework membranes by postsynthetic defect healing. *ACS Applied Materials & Interfaces*, 9:37848–37855, 2017.

[150] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(1):1–7, 2016.

[151] J. Q. Wen, Y. S. Han, X. Yang, and J. M. Zhang. Computational research of electronic, optical and magnetic properties of ce and nd co-doped zno. *Journal of Physics and Chemistry of Solids*, 125:90 – 95, 2019.

[152] Daniel Willhelm, Nathan Wilson, Raymundo Arroyave, Xiaoning Qian, Tahir Cagin, Ruth Pachter, and Xiaofeng Qian. Predicting van der waals heterostructures by a combined machine learning and density functional theory approach. *ACS Applied Materials & Interfaces*, 14(22):25907–25919, 2022.

[153] C. E Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature Chemistry*, 4(2):83–89, 2012.

[154] D. H Wolpert. Stacked generalization. *Neural Network*, 5(2):241–259, 1992.

[155] G. Wu, J. Huang, Y. Zang, J. He, and G. Xu. Porous field-effect transistors based on a semiconductive metal–organic framework. *Journal of the American Chemical Society*, 139(4):1360–1363, 2017.

[156] Xiaoyu Wu, Yu Che, Linjiang Chen, Eric Jean Amigues, Ruiyao Wang, Jinghui He, Huilong Dong, and Lifeng Ding. Mapping the porous and chemical structure–function relationships of trace ch3i capture by metal–organic frameworks using machine learning. *ACS Applied Materials & Interfaces*, 14(41):47209–47221, 2022.

[157] L. S Xie, E. V. Alexandrov, G. Skorupskii, D. M. Proserpio, and M. Dincă. Diverse π–π stacking motifs modulate electrical conductivity in tetrathiafulvalene-based metal–organic frameworks. *Chemical Science*, 10(37):8558–8565, 2019.

[158] L. S. Xie, G. Skorupskii, and M. Dincă. Electrically conductive metal–organic frameworks. *Chemical Reviews*, 120:8536–8580, 2020.

[159] L. S. Xie, G. Skorupskii, and M. Dincă. Electrically conductive metal–organic frameworks. *Chemical Reviews*, 120(16):8536–8580, 2020.

[160] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters*, 120(14):145301, 2018.

[161] O. M. Yaghi. Reticular chemistry in all dimensions. *ACS Central Science*, 5(8):1295–1300, 2019.

[162] O. M. Yaghi, G. Li, and H. Li. Selective binding and removal of guests in a microporous metal–organic framework. *Nature*, 378(6558):703–706, 1995.

[163] O. M. Yaghi and H. Li. Hydrothermal synthesis of a metal-organic framework containing large rectangular channels. *Journal of the American Chemical Society*, 117(41):10401–10402, 1995.

[164] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, and J. Kim. Reticular synthesis and the design of new materials. *Nature*, 423(6941):705–714, 2003.

[165] C. Yang, R. Dong, M. Wang, P. S. Petkov, Z. Zhang, M. Wang, P. Han, M. Ballabio, S. A. Bräuninger, Z. Liao, J. Zhang, F. Schwotzer, E. Zschech, H. Klauss, E. Cánovas, S. Kaskel, M. Bonn, S. Zhou, T. Heine, and X. Feng. A semiconducting layered metal-organic framework magnet. *Nature Communications*, 10(1):1–9, 2019.

[166] M. Yoon, R. Srirambalaji, and K. Kim. Homochiral metal-organic frameworks for asymmetric heterogeneous catalysis. *Chemical Reviews*, 112:1196–1231, 2012.

[167] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme. Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features. *Journal of Applied Crystallography*, 52(5):918–925, 2019.

[168] P. Zeng, C. Liu, C. Cheng, C. Yuan, K. Dai, J. Mao, L. Zheng, J. Zhang, L. Chang, S. Haw, T. Chan, H. Lin, and L. Zhang. Propelling polysulfide redox conversion by d-band modulation for high sulfur loading and low temperature lithium–sulfur batteries. *Journal of Materials Chemistry A*, 9(34):18526–18536, 2021.

[169] Z. Zhang, D. Dell'Angelo, M. R. Momeni, Y. Shi, and F. A. Shakib. Metal-to-semiconductor transition in two-dimensional metal–organic frameworks: An ab initio dynamics perspective. *ACS Applied Materials & Interfaces*, 13(21):25270–25279, 2021.

[170] Z. Zhang, D. S. Valente, Y. Shi, D. K. Limbu, M. R. Momeni, and F. A. Shakib. In silico high-throughput design and prediction of structural and electronic properties of low-dimensional metal–organic frameworks. *ACS Applied Materials & Interfaces*, 15(7):9494–9507, 2023.

[171] B. Zhao, J. Zhang, W. Feng, Y. Yao, and Z. Yang. Quantum spin hall and z2 metallic states in an organic material. *Physical Review B*, 90(20):201403, 2014.

[172] Q. Zhao, S. Li, R. Chai, X. Ren, and C. Zhang. Two-dimensional conductive metal–organic frameworks based on truxene. *ACS Applied Materials & Interfaces*, 12(6):7504–7509, 2020.

[173] F. Zheng, L. Z. Tan, S. Liu, and A. M. Rappe. Rashba spin–orbit coupling enhanced carrier lifetime in $ch_3nh_3pbi_3$. *Nano Letters*, 15(12):7794–7800, 2015.

[174] Y. Zheng, F. Sun, X. Han, J. Xu, and X. Bu. Recent progress in 2d metal-organic frameworks for optical applications. *Advanced Optical Materials*, 8(13):2000110, 2020.

[175] J. Zhou, Z. Han, X. Wang, H. Gai, Z. Chen, T. Guo, X. Hou, L. Xu, X. Hu, M. Huang, S. V. Levchenko, and H. Jiang. Discovery of quantitative electronic structure-oer activity relationship in metal-organic framework electrocatalysts using an integrated theoretical-experimental approach. *Advanced Functional Materials*, 31(33):2102066, 2021.